UNIVERSITY OF LEEDS

Estimation and Simulation in Directional and Statistical Shape Models

Asaad Mohammed Ganeiber

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

School of Mathematics

March 2012

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

Dedicated to my wife NASRIN and my son ANAS

Acknowledgements

First of all, praise be to Allah for the glory of Islam who told us in the Holy Koran "Verily, the noblest of you, in the sight of Allah, is the best in conduct and Allah has full knowledge and is well acquainted" [Koran XLIX:13].

Next, I offer my gratitude and sincere thanks to my supervisor Professor John T. Kent. He has taught me, both consciously and un-consciously, how good research study is done. I appreciate all his contributions of time and ideas to make my PhD experience productive and stimulating. The joy and enthusiasm he has for his research were contagious and motivational for me, even during tough times in the PhD pursuit. I am also thankful for the excellent example he has provided as a successful statistician and professor.

I am also very grateful to the Senior Research Professor Kanti V. Mardia, Professor Ian L. Dryden, Professor Charles C. Taylor, Sir David R. Cox, Professor David Firth, Professor Simon Wood, Professor Fred Bookstein, Professor Wilfrid Kendall, Professor Christian Klingenberg and Professor Andrew Wood for all the advanced background materials in Statistics I obtained during my attending some academic classes in the School of Mathematics, departmental seminars and conferences in UK.

I would like also to take this opportunity to thank all those who have contributed in any way, shape or form to the completion of this thesis. I am grateful to Dr Robert G. Aykroyd and Dr Alfred Kume for comments which helped to improve the clarity of the thesis, Dr Leonid V. Bogachev and the postgraduate secretary Mrs Jeanne S. Shuttleworth in School of Mathematics for their administrative advices and assistants, friends for their ideas and criticisms. My time in Leeds was made enjoyable in large part due to the many friends and groups that became a part of my life.

Lastly, I would like to thank my family for all their love and encouragement. For my parents and my brother Mr Ahmed M. Ganeiber who raised me with a love of science and supported me in all my pursuits. Without their assistance this work would not have been possible. Thank you.

Abstract

This thesis is concerned with problems in two related areas of statistical shape analysis in two dimensional landmarks data and directional statistics in various sample spaces.

Directional observations can be regarded as points on the circumference of a circle of unit radius in two dimensions or on the surface of a sphere in three dimensions. Special directional methods and models are required which take into account the structure of these sample spaces. Shape analysis involves methods for the study of the shape of objects where location, scale and orientation are removed. Specifically, we consider the situation where the objects are summarized by points on the object called landmarks. The non-Euclidean nature of the shape space causes several problems when defining a distribution on it. Any distribution which could be considered needs to be tractable and a realistic model for landmark data.

One aim of this thesis is to investigate the saddlepoint approximations for the normalizing constants of some directional and shape distributions. In particular, we consider the normalizing constant of the CBQ distribution which can be expressed as a one dimensional integral of normalizing constants for Bingham distributions. Two new methods are explored to evaluate this normalizing constant based on saddlepoint approximations namely the Integrated Saddlepoint (ISP) approximation and the Saddlepoint-Integration (SPI) approximation.

Another objective of this thesis is to develop new simulation methods for some directional and shape models. We propose an efficient acceptance-rejection simulation algorithm for the Bingham distribution on unit sphere using an angular central Gaussian (ACG) density as an envelope. This envelope is justified using inequalities based on concave functions. An immediate consequence is a method to simulate 3×3 matrix Fisher rotation matrices. In addition, a new accept-reject algorithm is developed to generate samples from the complex Bingham quartic (CBQ) distribution.

The last objective of this thesis is to develop a new moment method to estimate the parameters of the wrapped normal torus distribution based on the sample sine and cosine moments.

Contents

1 Introduction **1**

1.1 Directional Statistics 1

1.1.1 Circular Models 2

1.1.2 Spherical Models 3

1.1.3 Special Orthogonal Rotation Matrices and Torus Models 4

1.2 Statistical Analysis of Shapes 4

1.2.1 Shapes and Landmarks 4

1.2.2 Configurations 6

1.2.3 Shape and Pre-Shape Spaces 6

1.2.4 Shape Models 9

1.2.5 Relationship between Directional Statistics and Shape Analysis 10

1.3 Outline of the Thesis 11

1.3.1 Part I: Saddlepoint Approximations 11

1.3.2 Part II: Rejection Simulation Techniques 13

1.3.3 Part III: Methods of Estimation for Torus Data 14

I Saddlepoint Approximations **16**

2 Saddlepoint Approximations in Circular and Spherical Models **17**

2.1 Introduction 17

2.2 Background Ideas 18

2.3 Simple Saddlepoint Approximations 19

2.4 Refined Saddlepoint Approximations and Motivation 21
2.5 Tilting and Saddlepoint Approximations 24
2.6 Noncentral Chi-square Distribution 25
2.7 von Mises (Circular Normal) Distribution 32
2.7.1 Background 32
2.7.2 Saddlepoint Approximations for the Normalizing Constant 35
2.7.3 Saddlepoint Approximations for the Mean Resultant Length 36
2.8 Fisher Distribution on the Sphere 39
2.8.1 Background 39
2.8.2 Saddlepoint Approximations for the Normalizing Constant 41
2.9 Bingham Distribution on the Sphere 44
2.9.1 Background 44
2.9.2 Saddlepoint Approximations for the Normalizing Constant 47

3 Saddlepoint Approximations for the Complex Bingham Quartic Distribution 50
3.1 Introduction 50
3.2 Quadrature Methods 50
3.2.1 Trapezoidal Rule 51
3.2.2 Recursive Adaptive Simpson Quadrature 52
3.2.3 Gauss-Legendre Quadrature 52
3.3 Saddlepoint Approximations for Finite Mixtures 53
3.3.1 Background on Finite Mixtures 53
3.3.2 Saddlepoint Approximations 54
3.3.3 Application to Gamma Mixture Distribution 55
3.4 Complex Bingham Quartic Distribution 57
3.4.1 Background 58
3.4.2 Decomposition of Quadratic Forms 59
3.4.3 Some Properties and Motivation 60
3.4.4 Representation of the Normalizing Constant 61
3.4.5 Saddlepoint Approximations for the Normalizing Constant:
Integrated Saddlepoint Approximations Approach 62

3.4.6 Saddlepoint Approximations for the Normalizing Constant: Saddlepoint of Integration Approximations Approach 63
3.4.7 Change of Variables 64
3.4.8 Performance Assessment for Mixture Saddlepoint Approximations Approaches 68

II Rejection Simulation Techniques 76

4 Simulation of the Bingham Distribution Using an Inequality for Concave Functions 77

4.1 Introduction 77
4.2 Principles of Acceptance-Rejection Simulation Scheme 78
4.3 Envelopes Based on Concave Functions 80
4.4 Simulation from the Multivariate Normal Distribution with Multivariate Cauchy Envelope 83
4.5 Bilateral Exponential Envelope for Standard Normal Distribution 84
4.6 Simulation from the Real Bingham Distribution with ACG Envelope 86
4.7 Simulation from the von Mises Distribution with a Wrapped Cauchy Envelope 90
4.8 Link Between the Best-Fisher Method and the Concave Inequality 92
4.9 Simulation from the Matrix Fisher Distribution 95
4.9.1 The Matrix Fisher Probability Density Function 95
4.9.2 Simulation Scheme 95

5 General Techniques of Simulation from Directional and Shape Models 99

5.1 Introduction 99
5.2 A Brief Historical Survey of Simulation Methods 100
5.3 Simulation from the von Mises and von Mises-Fisher Distributions 102
5.3.1 von Mises Distribution with Real Bingham Envelope 102
5.3.2 von Mises-Fisher Distribution with Beta Envelope 104
5.4 Bipartite Rejection Scheme for the Real Bingham Distribution on the Sphere 108
5.5 Simulation from the Fisher Distribution with Bingham Envelope 110

5.6 Simulation from the Fisher-Bingham (FB5) Distribution 112
5.6.1 Background 112
5.6.2 FB5 Distribution with Uniform Envelope 113
5.6.3 FB5 Distribution with Truncated Exponential Envelope 114
5.7 Simulation from the Complex Bingham Distribution 116
5.7.1 The Complex Bingham Density Function 116
5.7.2 Simulation Schemes 118
5.8 Simulation from the Complex Bingham Quartic Distribution 120

III Methods of Estimation for Torus Data 127

6 Methods of Estimation for Torus Data 128
6.1 Introduction 128
6.2 Bivariate von Mises Torus Distribution 129
6.2.1 Marginal and Conditional Models 132
6.2.2 Comparing Models using Log-Densities 133
6.3 Moments and Correlations Under High Concentration 134
6.4 Approaches to Estimation 140
6.4.1 Maximum Likelihood Method 140
6.4.2 Maximum Pseudolikelihood Method 141
6.5 Wrapped Normal Torus Distribution 142
6.5.1 Overview and Background 142
6.5.2 Moments for the Wrapped Normal Torus Distribution 144

7 Conclusions and Future Directions 147
7.1 Saddlepoint Approximation 147
7.1.1 Motivation 147
7.1.2 Critical Summary 147
7.1.3 Future Work 148
7.2 Simulation Techniques 148
7.2.1 Motivation and Limitation 148

7.2.2 Critical Summary 148
7.2.3 Future Work 149
7.3 Estimation Methods for Torus Data 149
7.3.1 Motivation 149
7.3.2 Critical Summary 149
7.3.3 Future Work 150

A Appendix A **151**
A.1 R Functions for Simulating Bingham Distribution using ACG Envelope 151
A.2 R Functions for Best-Fisher algorithm of the von Mises Distribution 153
A.3 MATLAB Functions for Simulating von Mises-Fisher Distribution 154
A.4 Evaluate $\mathbb{E}\big(\cos\theta_1 \sin\theta_2\big) = 0$ for the Bivariate Sine Distribution 163

Bibliography **164**

List of Tables

1.1 Relationship between some common directional and shape distributions. 15

2.1 Numerical unnormalizing first order saddlepoint approximations and unnormalizing second order saddlepoint approximations for the normalizing constant of the von Mises distribution with various values of the concentration parameter κ. 37

2.2 Numerical unnormalizing first order saddlepoint approximations and unnormalizing second order saddlepoint approximations for the mean resultant length of the von Mises distribution with various values of the concentration parameter κ. 39

2.3 Numerical first order saddlepoint approximations $\hat{c}_{02}$ and second order saddlepoint approximations $\hat{c}_{03}$ for the normalizing constant of the Fisher distribution under various values of κ; n is the sample size required for the difference between the true κ and $\hat{\kappa}_3$ to be significant at the level 5% level when large sample likelihood ratio test is used (see Kume and Wood [54]). 45

2.4 Numerical results for the second order saddlepoint approximations for the normalizing constant of the real Bingham distribution with varying κ. 49

3.1 Numerical results for true value of the normalizing constant of the complex Bingham quartic distribution $c_{\mathrm{CBQ}}(\Omega)$: The second order integrated saddlepoint $\hat{c}_{\mathrm{CBQ,ISP}}(\Omega)$ approximations and the second order saddlepoint of integration $\hat{c}_{\mathrm{CBQ,\ SPI}}(\Omega)$ approximations with various values of κ , $\beta = 0.4\kappa$, λ_1, λ_2 and $n = 1000$. 74

3.2 Numerical results for the second order integrated saddlepoint $\hat{c}_{\mathrm{CBQ,ISP}}(\Omega)$ and the second order saddlepoint of integration $\hat{c}_{\mathrm{CBQ,\ SPI}}(\Omega)$ approximations with/without change of variable, with λ_1, λ_2, λ_3 and λ_4 varying and with $n = 1000$. 75

4.1 Analytical efficiencies for Multivariate Normal/Multivariate Cauchy Envelope A/R simulation with various values of p and b. 85

4.2 Simulated efficiencies rates and their standard errors for Bingham/Angular central Gaussian (ACG) Envelope A/R simulation with $n = 10000$ and various values of λ_1, λ_2 and $\lambda_3 = 0$. 90

4.3 Simulated efficiencies rates and their standard errors for matrix Fisher/Angular central Gaussian (ACG) Envelope A/R simulation with $n = 1000$ and various values of ϕ_1, ϕ_2, ϕ_3, λ_1, λ_2, λ_3 and $\lambda_4 = 0$. 98

5.1 Analytical efficiencies for Von Mises/Wrapped Cauchy, Beta and Real Bingham Envelopes A/R simulation A/R simulation with various values of κ, τ, ρ and x_0. 107

5.2 Analytical efficiencies for Bingham/Bipartite Envelopes A/R simulation with various values of κ. . 110

5.3 Analytical efficiencies for the FB5 distribution based on the Kent-Hamelryck, the uniform and the real Bingham envelopes A/R simulation with various values of κ and β ($\kappa > 2\beta$). 117

5.4 Simulated efficiencies M of truncated multivariate exponential envelope and angular central Gaussian envelope needed for the simulation methods from the complex Bingham distribution with k varying and with a common concentration value λ. 121

5.5 Simulated efficiencies rates and their standard errors for the complex Bingham quartic (CBQ) distribution/Mixture Multivariate Normals and the Kent-Hamelryck envelopes in the case $k = 3$ with various values of κ, β, λ_1, λ_2 and a sample of size $n = 10000$ (fixed $2\beta/\kappa < 1$). 125

5.6 Simulated efficiencies rates and their standard errors for the complex Bingham quartic (CBQ) distribution/Mixture Multivariate Normals envelope in the case $k = 4$ with various values of λ_1, λ_2, λ_3, λ_4 and a sample of size $n = 1000$. 126

6.1 Simulated moment estimations for the parameters of the wrapped bivariate normal torus distribution with a sample of size $n = 100$. 146

List of Figures

1.1 Stereographic projection 3

1.2 The hierarchies of the various spaces (after Goodall and Mardia [26]). 8

1.3 The hierarchies of some common shape and directional distributions. 12

2.1 The cumulant generating function $\mathbf{K}_S(u)$ for noncentral Chi-square distribution versus u. 27

2.2 Saddlepoint function $\hat{u}(s)$ versus s 28

2.3 Exact $f(s)$ (solid curve) and first-order saddlepoint density approximation $\hat{f}_1(s)$ (dashed line) versus s for noncentral chi-square distribution with various values of the noncentrality parameter $\alpha = 0, 1, 2, 3, 5$ and 7. 30

2.4 Exact $f(s)$ (solid curve), unnormalized first-order saddlepoint density approximation $\hat{f}_1(s)$ (dashed curve) and the normalized first-order saddlepoint density approximation $\bar{f}(s)$ (dotted curve) versus s for noncentral chi-square distribution. 32

3.1 The cumulant generating function $\mathbf{K}(u)$ for the gamma mixture versus u. 56

3.2 Gamma mixture random variable: Exact density and saddlepoint approximation. 57

3.3 The function $h\big(\Psi(s)\big)$ versus s. 67

3.4 The cumulant generating function $\mathbf{K}_{R^2}(u)$ versus u. 69

3.5 The saddlepoint approximation for the real Bingham distribution $\hat{c}_{\text{Bing}}\big(\Psi(s)\big)$ versus s. 71

3.6 The ratio $\hat{c}_{\text{CBQ}}(\Omega)/c_{\text{CBQ}}(\Omega)$ against the concentration parameter κ. 73

4.1 The concave function $\varphi(u)$ and the linear function $\psi(u)$ versus u. In the left panel $b = 15$ and $q = 10$. In the right panel $b = 10$, $q = 30$ and $u_0 = 10$. 81

4.2 The unnormalized standard normal function $f^*(x)$ and the unnormalized bilateral envelope function $g^*(x)$ versus x. In the left panel $\alpha = 0.3$ whereas in the right panel $\alpha = 1.0$. 85

4.3 The von Mises function $f(\theta)$ and the the simulation proportional envelope of the wrapped Cauchy distribution $Mg_\Theta(\theta)$. In the left panel $\kappa = 1.0$, $\tau = 3.236$, $\rho = 0.346$ and $M = 1.152$ whereas in the right panel $\kappa = 10$, $\tau = 21.02$, $\rho = 0.727$ and $M = 1.481$. 92

5.1 Envelope rejection for the von Mises function $f(\theta)$ with Bingham target distribution $g(\theta)$ for $\kappa = 5.0$ and $M = 1.20$. 104

5.2 Spherical plot of two random samples with 200-points each from von Mises-Fisher distribution around $\mu = [1\ 0\ 0]$ and $\kappa = 2, 10$ from left to right, respectively. 108

6.1 Contour plots of log-densities (normalizing constant omitted) for the sine, cosine with positive interaction and cosine with negative interaction models. Captions indicate the vector of parameters: κ_1, κ_2, η (Positive) for the sine model, $\kappa_1, \kappa_2, \gamma_1$ (Positive) for the cosine model with positive interaction model and $\kappa_1, \kappa_2, \gamma_2$ (Positive) for cosine with negative interaction models. 135

6.2 Contour plots of log-densities (normalizing constant omitted) for the sine, cosine with positive interaction and cosine with negative interaction models. Captions indicate the vector of parameters: κ_1, κ_2, η (Negative) for the sine model, $\kappa_1, \kappa_2, \gamma_1$ (Negative) for the cosine model with positive interaction model and $\kappa_1, \kappa_2, \gamma_2$ (Negative) for cosine with negative interaction models. 136

6.3 3D perspective plots of the log-densities (normalizing constant omitted) for the sine, cosine with positive interaction and cosine with negative interaction models. Captions indicate the vector of parameters: κ_1, κ_2, η (Positive) for the sine model, $\kappa_1, \kappa_2, \gamma_1$ (Positive) for the cosine model with positive interaction model and $\kappa_1, \kappa_2, \gamma_2$ (Positive) for cosine with negative interaction models. . . 137

6.4 3D perspective plots of the log-densities (normalizing constant omitted) for the sine, cosine with positive interaction and cosine with negative interaction models. Captions indicate the vector of parameters: κ_1, κ_2, η (Positive) for the sine model, $\kappa_1, \kappa_2, \gamma_1$ (Positive) for the cosine model with positive interaction model and $\kappa_1, \kappa_2, \gamma_2$ (Positive) for cosine with negative interaction models. . . 138

6.5 Simulated bivariate normal torus data with small variance $\sigma^2 = 1.0$ and strong correlation $\rho = 0.90$ for a sample of size $n = 100$. (a) before wrapping; (b) after wrapping. 143

6.6 Simulated bivariate normal torus data with large variance $\sigma^2 = 50$ and strong correlation $\rho = 0.90$ for a sample of size $n = 100$. (a) before wrapping; (b) after wrapping. 143

6.7 Simulated bivariate normal torus data with large variance $\sigma^2 = 50$ and weak correlation $\rho = 0.50$ for a sample of size $n = 100$. (a) before wrapping; (b) after wrapping. 144

Chapter 1

Introduction

The work contained in this thesis falls into the two related areas of statistical shape analysis and directional statistics. Let us start by introducing the fields involved in this thesis. Basic definitions referred to through out the text are given.

1.1 Directional Statistics

There are various statistical problems which arise in the analysis of data when the observations are directions. Directional data are often met in astronomy, biology, geology, medicine and meteorology, such as in investigating the origins of comets, solving bird navigational problems, interpreting palaeomagnetic currents, assessing variation in the onset of leukaemia, analysing wind directions, etc.

The directions are regarded as points on the circumference of a circle in two dimensions or on the surface of a sphere in three dimensions. In general, directions may be imagined as points on the surface of a hypersphere but observed directions are obviously angular measurements.

The difficulty in the statistical analysis of directional data stems from the disparate topology of the circle and the straight line: if angles are recorded in radians in the range $[-\pi, \pi)$, then directions close to the opposite end-points are near neighbours in a metric which respects the topology of the circle, but maximally distant in a linear metric. Thus, many standard statistical procedures are inappropriate for modelling directional data (Coles [10]).

1.1.1 Circular Models

A large number of circular probability models exists; like linear probability models, they may be either discrete or continuous. Several of the more important are discussed in this thesis.

Many useful and interesting circular models may be generated from probability distributions on the real line or on the plane, by a variety of mechanisms. We describe a few such general methods (Jammalamadaka and SenGupta [31], p. 30):

(1) By wrapping a linear distribution around the unit circle. Any linear random variable X on the real line may be transformed to a circular random variable by reducing it modulo 2π i.e. using $\theta = X(\mathrm{mod}\, 2\pi)$. The wrapped normal and the wrapped Cauchy distributions are of interest in this thesis.

(2) Through characterizing properties such as maximum entropy, etc. It is often instructive to ask if there are distributions on the circle which enjoy certain desirable properties. For instance, one may ask which distribution has the maximum entropy subject to having non-zero trigonometric moments. The uniform and von Mises (Circular Normal) distributions have the maximum entropy (Mardia and Jupp [70]. p. 42) where the entropy of a distribution on the circle with probability density function f is defined as

$$H(f) = -\int_0^{2\pi} f(\theta) \log f(\theta) d\,\theta.$$

This is one way of measuring the closeness of a distribution to the uniform distribution. von Mises distribution is of interest in this thesis. Moreover, if we ask which distribution on the circle has the property that the sample mean direction and the length of the resultant vector are independent, then the uniform or isotropic distribution is the answer (see Kent et. al. [47] and Jammalamadaka and SenGupta [31], p. 32). This characterization of the uniform distribution is similar to, and as important as, that of the normal distribution on the line as the only one in which the sample mean and sample variance are independent.

(3) By transforming a bivariate linear random variable to just its directional component, the so-called offset distributions. This is done by accumulating probabilities over all different lengths for a given direction. We transform the bivariate random vector (X, Y) into polar co-ordinates (r, θ) and integrate over r for a given θ. If $f(x, y)$ denotes the joint distribution of a bivariate

distribution on the plane, then the resulting circular offset distribution, say $g(\theta)$, is given by

$$g(\theta) = \int_0^\infty f(r\cos\theta, r\sin\theta) r\, d\, r.$$

(4) One may start with a distribution on the real line, and apply a stereographic projection that identifies points x with those on the circumference of the circle, say θ. This correspondence is one-to-one except for the fact that the mass if any, at both $+\infty$ and $-\infty$, are identified with π (Jammalamadaka and SenGupta [31], p. 31). Such a correspondence is shown in Figure 1.1.

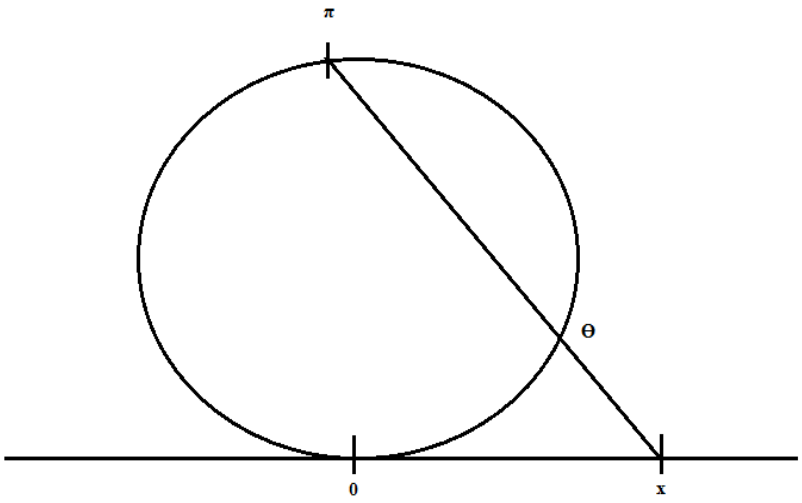

Figure 1.1: Stereographic projection

1.1.2 Spherical Models

Much of the theory of spherical statistics is analogous to that for circular statistics. Further, one can consider directions in p dimensions, i.e. unit vectors in p-dimensional Euclidean space $\mathbb{R}^p$. Directions in p dimensions can be represented as unit vectors $\mathbf{x}$, i.e. as points on $S^{p-1} = \left(\mathbf{x} : \mathbf{x}^T\mathbf{x} = 1\right)$, the $(p-1)$-dimensional sphere with unit radius and centre at the origin. Some spherical distributions are of interest in this thesis namely, von Mises-Fisher, Fisher, Fisher-Bingham (Kent) distributions for the spherical data and the Bingham and the angular central Gaussian (ACG) distributions for the axial data.

1.1.3 Special Orthogonal Rotation Matrices and Torus Models

In the previous subsections we have considered mainly observations which are unit vectors (directions) or axes. However, other types of observations occur in directional statistics, the most important of these from the practical point of view being rotations, orthonormal frames and torus.

An orthonormal r-frame in $\mathbb{R}^p$ is a set $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p)$ of orthonormal vectors in $\mathbb{R}^p$. The space of orthonormal r-frame in $\mathbb{R}^p$ is called the Stiefel manifold $V_r(\mathbb{R}^p)$. In terms of $p \times r$ matrices $\mathbf{X}$, $V_r(\mathbb{R}^p) = \left\{\mathbf{X} : \mathbf{X}^T\mathbf{X} = \mathbf{I}_r\right\}$. An orthonormal p-frame is equivalent to an orthogonal matrix, so $V_r(\mathbb{R}^p) = O(p)$, the orthogonal group consisting of all orthogonal $p \times p$ matrices. Moreover, an orthogonal $(p-1)$-frame $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{p-1})$ can be extended uniquely to an orthogonal p-frame $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p)$ with matrix of determinant 1, so $V_r(\mathbb{R}^p) = SO(p)$, the special orthogonal group consisting of all $p \times p$ rotation matrices (Mardia and Jupp [70], p. 285). The matrix Fisher distribution on a group $SO(3)$ of all rotations of $\mathbb{R}^3$ is of interest in the simulation chapters in this thesis.

In geometry, a torus is a surface of revolution generated by revolving a circle in three dimensional space about an axis coplanar (all the points lie in the same geometric plane) with the circle. In topology, a torus is homeomorphic to the Cartesian product of two circles: $\mathbb{T}^2 = S^1 \times S^1$ (Nikulin and Shafarevich [78], p. 110). Sometimes it is necessary to consider the joint distribution of two circular random variable θ_1 and θ_2. Then (θ_1, θ_2) take values on the unit torus. In the uniform distribution on the torus, θ_1 and θ_2 are independent and uniformly distributed. Some interested distributions on the torus are the bivariate von Mises (sine and cosine) distribution and the wrapped multivariate normal distribution.

1.2 Statistical Analysis of Shapes

1.2.1 Shapes and Landmarks

Shape is all the geometrical information that remains when location, scale and rotational effects are filtered out from an object.

According to this, definition of shape is invariant under Euclidean similarity transformations of translation, scaling and rotation as follows. The simplest type of object which can be studied consists of a labelled set of k points in $\mathbb{R}^m$ (where $k \geq m = 1$), represented as a $k \times m$ matrix, $\mathbf{X}$,

say. Then for any location vector $\gamma \in \mathbb{R}^m$, orthogonal $m \times m$ rotation matrix Γ satisfying $\det(\Gamma) = 1$ and $\Gamma^T\Gamma = \Gamma\Gamma^T = I_m$, and a scale $\beta > 0$, X has the same shape as $\beta\mathbf{X}\Gamma + 1_k\gamma^T$ (Euclidean similarity transformation of $\mathbf{X}$). Similarity transformations in $\mathbb{R}^2$ can also use complex notation. Consider $k \geq 3$ landmarks in $\mathbb{C}$, $\mathbf{z}^o = (z_1^o, z_2^o, \ldots, z_k^o)^T$ which are not all coincident. The Euclidean similarity transformations of $\mathbf{z}^o$ are $\beta e^{i\theta}\mathbf{z}^o + 1_k\xi$ where $\beta \in \mathbb{R}^+$ is the scale, $0 \leq \theta < 2\pi$ is the rotation angle and $\xi \in \mathbb{C}$ is the translation.

A related concept to shape is form. It is the geometrical information that remains when location and rotational effects are filtered out from an object. In other words two objects have the same size-and-shape (form) if they can be translated and rotated to each other so that they match exactly, i.e if the objects are rigid body transformations of each other.

The next question that naturally arises is: How should one describe a shape? One way to describe a shape is by locating a finite number of points on the outline. Consequently, the concept of a landmark is adopted by Dryden and Mardia [18].

A landmark is a point in two or three-dimensional space that corresponds to the position of a particular feature on an object of interest. For example, in the study of osteological remains, a landmark might be defined as the point that marks the scar of a muscle insertion on a bone, the intersection of two or more bones at a cranial suture, or the foramen that marks the path of a neurovascular bundle. We choose to focus on landmark data because we want to analyze data for which points on one object have an unambiguous correspondence to points on another object (Lele and Richtsmeier [56], p. 14).

Dryden and Mardia [18] split landmarks into three subgroups:

1. Anatomical landmarks: Points assigned by an expert that correspond between organisms in some biologically meaningful way.

2. Mathematical landmarks: Points located on an object according to some mathematical or geometrical property, i.e. high curvature or an extremum point.

3. Pseudo-landmarks: Constructed points on an object either on the outline or between landmarks.

1.2.2 Configurations

A mathematical representation of a k-point shape in m dimensions could be created by concatenate each dimension into a km-vector. The vector representation $\mathbf{x}$ for planar shapes ($m = 2$) would then be:

$$\mathbf{x} = [x_1, x_2, \ldots, x_k, y_1, y_2, \ldots, y_k]^T.$$

Alternatively, we may recast a km-vector as a configuration matrix with k rows and m columns. Thus, the configuration matrix X for planar shapes ($m = 2$) would then be:

$$\mathbf{X} = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_k & y_k \end{bmatrix}.$$

To perform a shape analysis, a biologist traditionally selects ratios of the distances between landmarks or angles, and then submits these to a multivariate analysis. This approach has been called multivariate morphometrics or the traditional method. Another approach is to consider a shape space obtained directly from the landmark coordinates, which retains the geometry of a point configuration, this has been called geometric shape analysis or the geometrical method.

Shape variables are features constructed from the configuration $\mathbf{X}$ that are unchanged under similarity transformations (translation, scaling and rotation). Similarly size-and-shape or form variables are unchanged under rigid body motions (translation and rotation). Size variables are form variables that are invariant under scaling changes; that is, if $\beta > 0$ is some constant, the size variable for $\beta\mathbf{X}$ must be β times the size variable for $\mathbf{X}$.

1.2.3 Shape and Pre-Shape Spaces

Shape space is the set of all possible shapes. Formally, the shape space Σ_m^k is the orbit space of the non-coincident k point set configurations in $\mathbb{R}^m$ under the action of the Euclidean similarity transformations.

The pre-shape, $\mathbf{Z}$, of a configuration matrix $\mathbf{X}$ has all the information about location and scale removed. it is usually constructed by centring the configuration and then dividing by size. The

pre-shape is given by

$$\mathbf{Z} = \frac{\mathbf{H}\ \mathbf{X}}{\|\mathbf{H}\ \mathbf{X}\|}$$

where $\mathbf{H}$ is a $(k-1)\times k$ Helmert matrix without the first row and it is called the Helmert sub-matrix. The centred pre-shape, $\mathbf{Z}_C = \mathbf{C}\ \mathbf{X}/\|\mathbf{C}\ \mathbf{X}\|$ is another pre-shape representation where $\mathbf{C} = \mathbf{I}_k - \frac{1}{k}1_k 1_k^T$ is centred matrix and also an idempotent ($\mathbf{C}^T\mathbf{C} = \mathbf{C},\ \mathbf{C}^2 = \mathbf{C}$) (Dryden and Mardia [18], p.55). Note that $\mathbf{Z}$ is a $(k-1)\times m$ matrix whereas $\mathbf{Z}_C$ is a $k \times m$ matrix and the relationship between the pre-shape and centred pre-shape is $\mathbf{Z}_C = \mathbf{H}^T\mathbf{Z}$. Both pre-shape representations are equally suitable for the pre-shape space which has real dimension $km-1$. The advantage in using $\mathbf{Z}$ is that it is of full rank and its dimension is less than that of $\mathbf{Z}_C$. On the other hand, the advantage of working with the centred pre-shape $\mathbf{Z}_C$ is that a plot of the Cartesian coordinates gives a correct geometrical view of the shape of the original configuration (Dryden and Mardia [18], p.55).

Pre-shape space is the space of all possible pre-shapes. Formally, the pre-shape space S_m^k is the orbit space of the non-coincident k point set configurations in $\mathbb{R}^m$ under the action of translation and isotropic scaling.

If we remove translation from the original configuration then the resulting landmarks are called Helmertized. Filtering scale from those Helmertized landmarks yields pre-shape whereas eliminating rotation from them should create size-and-shape. Again, removing rotation from pre-shape or removing scale from size-and-shape should result shape and after removing reflection for these shape landmarks the result is reflection shape (Dryden and Mardia [18], p.55). Figure 1.2 gives a diagram indicating the hierarchies of the different spaces.

Complex arithmetic when $m = 2$ enables us to deal with shape analysis very effectively. The advantage of using complex notation is that rescaling and rotation of an object in two dimensions can be obtained by complex multiplication by a complex number; for example, $\lambda\mathbf{z} = r\exp(i\,\theta)\mathbf{z}$ has the same shape as $\mathbf{z}$, although being rescaled by r and rotated anticlockwise by θ radians about the origin (Mardia [63]).

Consider $k \geq 3$ landmarks in $\mathbb{C}$, $\mathbf{z}^o = (z_1^o, z_2^o, \ldots, z_k^o)$ which are not all coincident. Location is removed by pre-multiplying by the Helmert sub-matrix H giving the complex Helmertized landmarks $z_H = Hz^o$. The centroid size is

$$S(\mathbf{z}^o) = \{(\mathbf{z}^o)^*\mathbf{C}\,\mathbf{z}^o\}^{1/2} = \|\mathbf{z}_H\| = \sqrt{(\mathbf{z}_H)^*\mathbf{z}_H},$$

where $(\mathbf{z}^o)^*$ denotes the complex conjugate of the transpose of $\mathbf{z}^o$. Hence the complex pre-shape $\mathbf{z}$

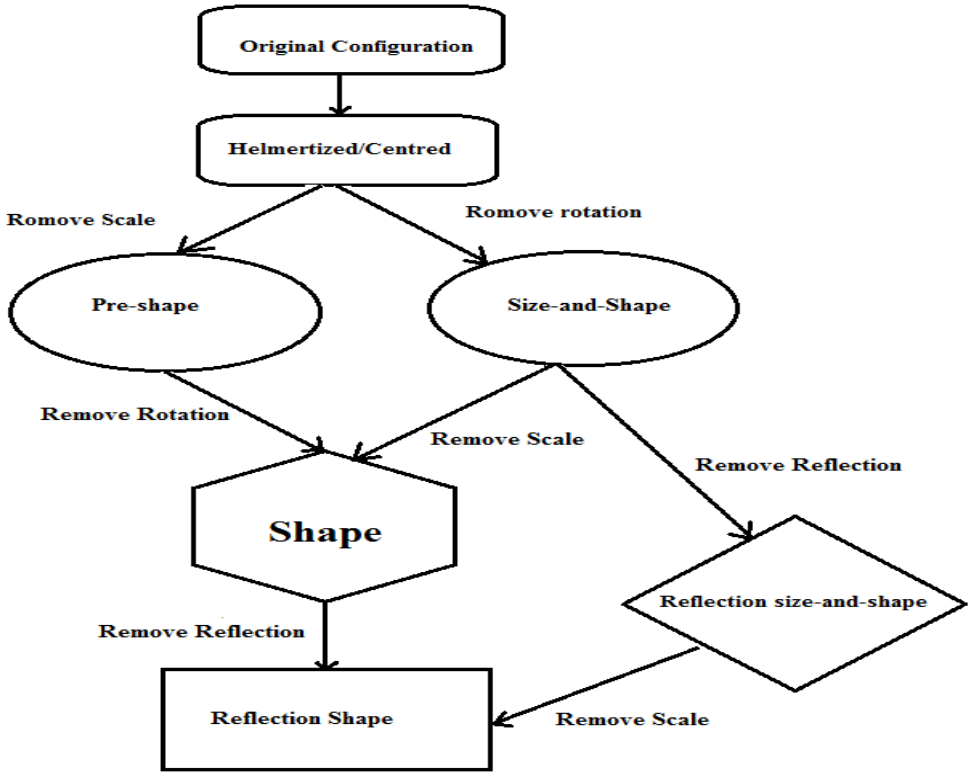

Figure 1.2: The hierarchies of the various spaces (after Goodall and Mardia [26]).

is obtained by dividing the Helmertized landmarks by the centroid size,

$$\mathbf{z} = \mathbf{z}_H / S(\mathbf{z}^o), \quad \mathbf{z} \in S_2^k$$

We see that the pre-shape space S_2^k is the complex sphere in $k-1$ complex dimensions

$$\mathbb{C}S^{k-2} = \{\mathbf{z} : \mathbf{z}^*\mathbf{z} = 1, \quad \mathbf{z} \in \mathbb{C}^{k-1}\},$$

which is the same as the real sphere of unit radius in $2k-2$ real dimensions, S^{2k-2}. In order to remove rotation we identify all rotated versions of $\mathbf{z}$ with each other, i.e. the shape of $\mathbf{z}^o$ is

$$[\mathbf{z}^o] = \{\mathbf{z}e^{i\theta} : 0 \leq \theta < 2\pi\}$$

The complex sphere $\mathbb{C}S^{k-2}$ which has points $\mathbf{z}$ identified with $\mathbf{z}e^{i\theta}$ $(0 \leq \theta < 2\pi)$ is the complex projective space $\mathbb{C}P^{k-2}$. Hence, the shape space for k points (Dryden and Mardia [18], pp. 58-59) in two dimensions is

$$\Sigma_2^k = \mathbb{C}P^{k-2}$$

1.2.4 Shape Models

The current work on the shape analysis in this thesis focuses on the distributions of shape analysis in two dimensions. There are several issues to consider and there are various difficulties to overcome. Since the shape space is non-Euclidean special care is required. Our main emphasis will be on distributions on $\mathbb{C}P^{k-2}$. Any distribution which could be considered needs to be tractable and a realistic model for landmark data.

Suitable ways of obtaining shape distributions (Dryden and Mardia [18], p. 109):

(1) Consider distributions in configuration space, conditional on size. This proposal is called the conditional approach, where the non-directional variables are held constant.

(2) Consider distributions in configuration space, with the similarity transformations integrated out. This proposal is called the marginal approach, where we integrate out the non-directional variables e.g. offset normal shape distribution (Dryden and Mardia [18], p. 124) and Mardia-Dryden distributions (Mardia and Dryden [66] and Dryden and Mardia [17]).

(3) Consider distributions on the pre-shape space which are invariant under rotations.

(4) Consider distributions based on shape distances.

(5) Consider distributions in the tangent space.

Recently these approaches have produced useful shape distributions, starting with the distributions of Mardia and Dryden [67] following the marginal approach. Kent [41] adapted the conditional approach and introduced the complex Bingham (CB) distribution. The complex Watson distribution is another shape model for the landmark data (Mardia [62] and Dryden and Mardia [18], p. 118). For triangles $k = 3$ it is the same as the complex Bingham (CB) distribution (Mardia and Dryden [67], p. 119). Kent [41] suggests the complex angular central Gaussian ACG distribution for shape data. Kent et al. [46] suggest also the complex Bingham quartic (CBQ) distribution on the unit complex sphere in $\mathbb{C}^{k-1}$. The CBQ distribution is an extension to the complex Bingham distribution. Under high concentrations the complex Bingham distribution has a complex normal distribution. By adding a quartic term to the complex Bingham density the CBQ distribution is obtained, which allows a full normal distribution under high concentrations. Our major contribution in this thesis concentrates on the complex Bingham (CBQ) distribution.

1.2.5 Relationship between Directional Statistics and Shape Analysis

We can construct shape distributions directly from directional distributions themselves.

(1) For the triangle case, $k = 3$, the identification of $\mathbb{C}P^1$ to S^2 allows immediately a shape distribution using the isometric transformation

$$x = |z_1|^2 - |z_2|^2, \quad y = 2\,\mathrm{Re}(\bar{z}_1 z_2), \quad z = 2\,\mathrm{Im}(\bar{z}_1 z_2)$$

to any spherical distribution. Here $\bar{z}$ is the complex conjugate of z. In fact, this mapping is isometric ($\Sigma_2^3 = \mathbb{C}P^1 = S^2(\frac{1}{2})$), so we may call such distributions the isometric distributions (Kendall et. al. [38], pp. 4-12 and Kendall, [36]).

(2) For $k > 3$, we can use a directional distribution $\mathbf{z}$ on a preshape $\mathbb{C}S^{k-2}$ and integrate out, say, ψ in $z_p = r\exp(i\,\psi)$, $r > 0$, $0 < \psi \leq 2\pi$, to obtain a shape density. However, a simpler approach is to take a density on the preshape $\mathbf{z} \in \mathbb{C}S^{k-2}$ which satisfies the rotational symmetry, so integrating out over ψ is not necessary. In particular, complex symmetric distributions with the density of the form $f(\mathbf{z}^*\mathbf{A}\ \mathbf{z})$ are automatically shape distributions (Mardia [63], Kent [41] and Mardia and Dryden, [67]).

Table 1.1 and Figure 1.3 give the relationship between some of common directional and shape distributions. Here κ is a concentration parameter, μ is a mean direction, β is an ovalness parameter for FB5 distribution, $\mathbf{A}$ is a symmetric $p \times p$ matrix with trace $\mathbf{A} = 0$ and $\mathbf{B}$ is a $(k-2) \times (k-2)$ negative positive complex matrix for the CBQ distribution in terms of the partial Procrustes tangent co-ordinates (Kume and Wood [54], Kent [41], Dryden and Mardia [18], Mardia and Jupp [70], Fisher et. al [22], Watson [96], Mardia [62], Kent [40], Kent et al. [46] and Mardia and Dryden [66]). The diagram describes the relationship between some common shape distributions themselves, the relationship between some famous directional distributions themselves and the relationship between both the shape and directional models. It is clear from the diagram that there is a direct link between the von Mises-Fisher distribution and the uniform, the von Mises and the Fisher distributions. Another link can be observed between the Fisher-Bingham, the Bingham, the von Mises-Fisher, the Fisher, the Kent and the 2-Wrapped distributions. On the other hand, there is a third link between some famous shape models. For the triangle case and presence just two distinct eigenvalues in the parameter matrix $\mathbf{A}$ (a single distinct largest eigenvalue and all other eigenvalues being equal), the complex Watson distribution is a special case of the complex Bingham distribution. The

complex Bingham distribution is also a special case of the complex Bingham quartic distribution if the $(k-2) \times (k-2)$ negative positive complex matrix $\mathbf{B} = 0$ (in terms of the partial Procrustes tangent co-ordinates). In the triangle case ($k = 3$), there is a fourth link between some common shape and directional distributions. In particular, the complex Bingham distribution tends to the Fisher distribution and the the complex Bingham quartic distribution becomes the Kent (FB5) distribution.

1.3 Outline of the Thesis

The title of this thesis is Estimation and Simulation in Directional and Statistical Shape Models. The material discussed divides naturally into three major parts namely, saddlepoint approximations as statistical tools of estimation, rejection simulation techniques and method of estimation for torus data.

1.3.1 Part I: Saddlepoint Approximations

In Chapter 2 we begin by looking at some basic principles of approximation using the familiar tool of Taylor expansion. The underlying strategy of the approximation carries through to more sophisticated saddlepoint approximation. Although the theory of saddlepoint approximations is quite complex, use of the approximations is fairly straightforward. The saddlepoint method provides an accurate approximation to the density or the distribution of a statistic, even for small tail probabilities and with very small sample sizes. This accuracy is seen not only in numerical work, but also in theoretical calculations. We apply this technique to the normalizing constants of some circular directional distributions such as von Mises distribution as well as to approximate the normalizing constants for some suitable distributions for spherical and axial data such as the Fisher and the Bingham distributions.

Chapter 3 starts with a review of some numerical integration methods. The normalizing constant of the CBQ distribution has no closed form and therefore we provide an approximation procedure based on saddlepoint approximations for finite mixtures of distributions. Calculating the normalizing constant for the CBQ distribution is based on numerical methods of quadrature (uniform nodes). Two methods are explored to evaluate this normalizing constant based on saddlepoint approximation of Bingham densities namely, the Integrated Saddlepoint (ISP) approximation and the Saddlepoint-

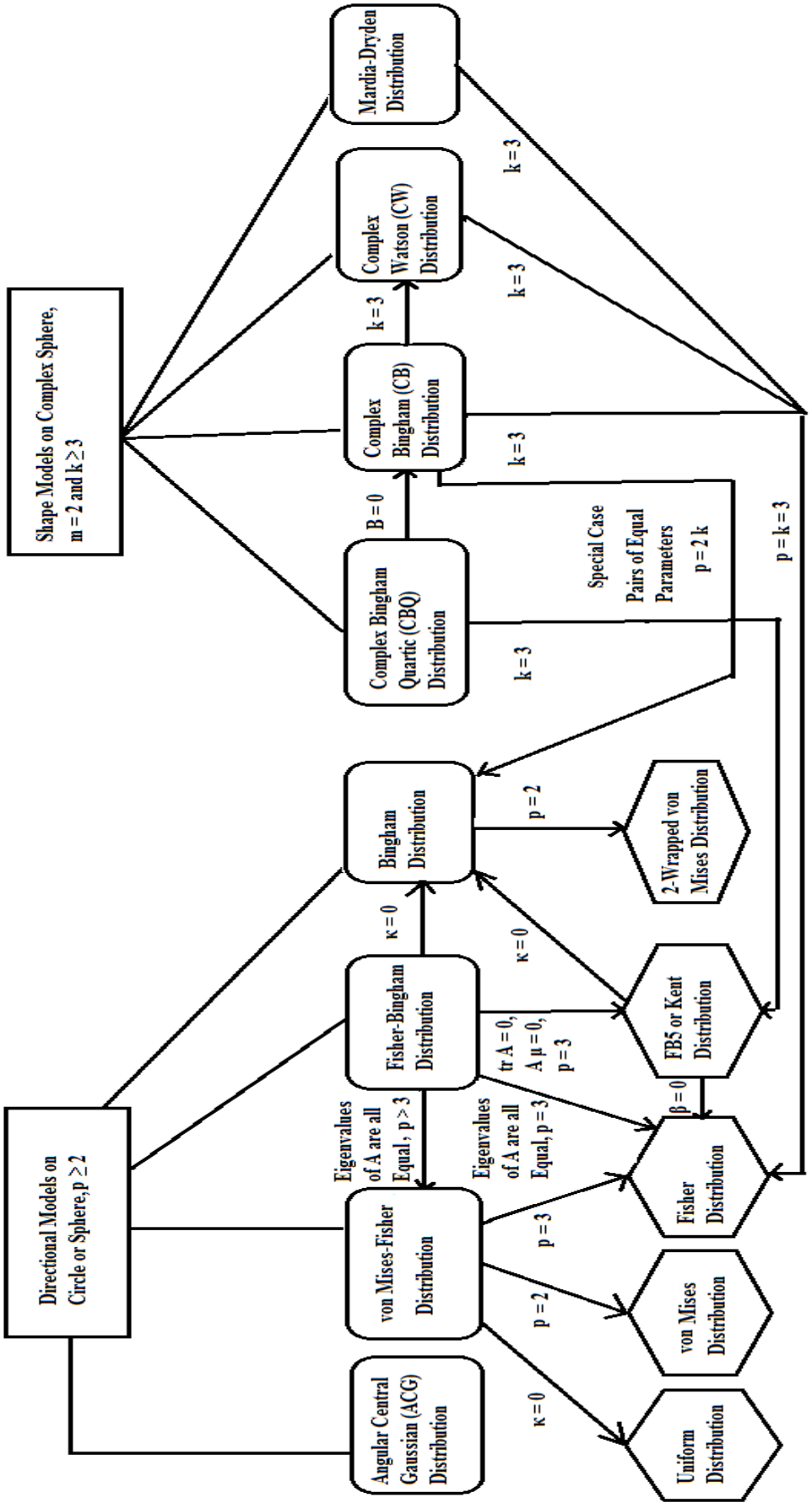

Figure 1.3: The hierarchies of some common shape and directional distributions.

Integration (SPI) approximation. One notable drawback of numerical quadrature is the need to pre-compute (or look up) the requisite weights and nodes. The uniform nodes are not a suitable choice to compute the integrand function for the normalizing constant of the CBQ distribution numerically especially under high concentration. An initial change of variable treatment is suggested instead.

1.3.2 Part II: Rejection Simulation Techniques

The second part divides into two subparts namely simulation techniques based on concave functions and general rejection schemes.

Chapter 4 discusses some new simulation methods. The main purpose of this chapter is to develop an efficient accept-reject simulation algorithm for the Bingham distribution on the unit sphere in $\mathbb{R}^p$ using an ACG envelope. The presentation proceeds in several stages. Firstly a review is given for the general A/R simulation algorithm. Secondly a general class of inequalities is given based on concave functions. These inequalities are illustrated for the multivariate normal distribution in $\mathbb{R}^p$ by finding two envelopes, viz., the multivariate Cauchy and the multivariate bilateral exponential distributions, respectively. An inequality similar to that is used to show that the ACG density can be used as an envelope for the Bingham density. The Bingham distribution on S^3 is identified to the matrix Fisher distribution on $SO(3)$. Hence the method of simulation from the Bingham distribution coincide to a method for simulating the matrix Fisher distribution.

Chapter 5 considers general simulation techniques from some directional and shape distributions. An A/R algorithm based on Bingham density is developed to generate samples from the von Mises distribution on the circle. Ulrich's simulation algorithm from the von Mises-Fisher distribution with an envelope proportional to Beta distribution is investigated. For the circular case, a comparison is given between the efficiency of the Ulrich's algorithm and that of the Best-Fisher scheme. A review is given of the Kent-Hamelryck simulation algorithm to sample from the FB5 distribution. Two other simulation methods are developed to generate samples from the 5 parameter Fisher-Bingham (FB5) using uniform and Bingham envelopes. In this chapter we also propose an acceptance-rejection simulation algorithm from the CBQ distribution. The problem of simulating from this complex shape distribution reduces to simulation from a mixture of two standard multivariate normal distributions. The efficiency rate is approximately 50% under high concentration.

1.3.3 Part III: Methods of Estimation for Torus Data

In Chapter 6 we review the sine and cosine bivariate distributions on torus. Maximum likelihood (ML) and pseudolikelihood (PL) estimators for the sine distribution are discussed. A comparison is also given between three bivariate sine and cosine models based on contours of the log-densities. For each of the three models, the parameters are chosen to match any positive definite inverse covariance matrix. For the wrapped normal torus distribution, we investigate a moment method to estimate the parameters based on the sample variance-covariances.

Chapter 7 gives a summary study and some potential work on the current fields in the future.

Shape Models	Directional Models
Complex Bingham distribution.	Real Bingham distribution: The $(k-2)$-dimensional complex Bingham (CB) distribution can be regarded as a special case of a $(2k-2)$-dimensional real Bingham distribution (Dryden and Mardia [18], p.113 and Kent [41], p. 287).
Complex Bingham distribution.	Fisher distribution: For the triangle case, the shape space is the 2-sphere of radius one-half and the complex Bingham (CB) distribution on $\mathbb{C}S^1$ is equivalent to using Fisher distribution on S^2 (Kent [41]).
Complex Bingham quartic (CBQ) distribution.	Fisher-Bingham (FB5) distribution: For the triangle case, $k=3$, the complex Bingham quartic (CBQ) distribution on $\mathbb{C}S^1$ is equivalent to using Fisher-Bingham (FB5) distribution on S^2 (Kent et al. [46]).
Complex Watson distribution: Special case of the complex Bingham (CB) distribution when $k=3$ and there are just two distinct eigenvalues in $\mathbf{A}$ (a single distinct largest eigenvalue and all other eigenvalues being equal) (Dryden and Mardia [18], p.118).	von Mises-Fisher distribution: The central role that the von Mises-Fisher distribution plays in directional data analysis is played by the complex Watson distribution for two dimensional shape analysis. For the triangle case, $k=3$, the complex Watson distribution on $\mathbb{C}S^1$ is equivalent to using Fisher distribution on S^2 (Dryden and Mardia [18], p.123).
Complex angular central Gaussian (CACG) distribution.	Angular central Gaussian (ACG) distribution: The $(k-2)$-dimensional complex angular central Gaussian (CACG) distribution can be regarded as a special case of a $(2k-2)$-dimensional angular central Gaussian (ACG) distribution.
Mardia-Dryden distribution.	Fisher distribution: For lower $\kappa \to 0$ and higher $\kappa \to \infty$ concentrations and for triangle case, $k=3$, Mardia-Dryden distribution on shape sphere behaves like the Fisher distribution (Mardia [61]).

Table 1.1: Relationship between some common directional and shape distributions.

Part I

Saddlepoint Approximations

Chapter 2

Saddlepoint Approximations in Circular and Spherical Models

2.1 Introduction

Modern statistical methods use models that require the computation of probabilities from complicated distributions, which can lead to intractable computations. Saddlepoint approximations can be the answer (Butler [9]). Although the theory of saddlepoint approximations is quite complex, use of the approximations is fairly straightforward. The saddlepoint method provides an accurate approximation to the density or the distribution of a statistic, even for small tail probabilities and with very small sample sizes. This accuracy is seen not only in numerical work, but also in theoretical calculations. The basis of this method is to overcome the inadequacy of the normal approximation in the tails by tilting the random variable of interest in such a way that the normal approximation is evaluated at a point near the mean (Paolella [79], pp. 170-171). In this chapter we apply this technique to the normalizing constants of some circular directional distributions such as the von Mises distribution as well as to approximate the normalizing constants for some suitable distributions for spherical and axial data such as Fisher and Bingham distributions. The Fisher-Bingham distribution, for instance, is obtained when a multivariate normal vector is conditioned to have unit length; its normalizing constant can be expressed as an elementary function multiplied by the density, evaluated at 1, of a linear combination of noncentral χ^2_1 random variables. Hence we may approximate the normalizing constant by applying a saddlepoint approximation to this density (Kume and Wood [54]).

We begin by looking at some basic principles of saddlepoint approximation, using the familiar tool of the Taylor expansion. As we will see, the underlying strategy of this approximation carries through to more sophisticated approximations.

2.2 Background Ideas

We recall that for a probability density function $f(x)$ on $(-\infty, \infty)$, the moment generating function (MGF), or the cumulant transform, $\mathbf{M}_X(u)$ is defined as

$$\begin{aligned} \mathbf{M}_X(u) &= \mathbf{e}^{\mathbf{K}_X(u)} \\ &= \mathbf{E}[\exp(uX)] = \int_{-\infty}^{+\infty} \exp(ux) f(x) dx, \end{aligned} \tag{2.1}$$

over values of u for which the integral converges. With real values of u, the convergence is always assured at $u = 0$. In addition, we shall presume that $\mathbf{M}(u)$ converges over an open neighbourhood of zero designated as $(-u_1, u_2)$, and that, furthermore, $(-u_1, u_2)$ is the largest such neighbourhood of convergence. This presumption is often taken as a requirement for the existence of the MGF ($\mathbf{M}_X(u) < \infty$). The function $\mathbf{K}_X(u)$ in (2.1) is called the cumulant generating function and defined as

$$\mathbf{K}_X(u) = \log\big(\mathbf{M}_X(u)\big). \tag{2.2}$$

From $\mathbf{M}_X(u)$ we can obtain $f(x)$ by using the Fourier inversion formula (Feller [21], Ch. XV and Billingsley [6], sec. 26)

$$\begin{aligned} f(x) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \mathbf{M}_X(iu) \exp(-iux) du \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \phi_X(u) \exp(-iux) du \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp\big\{\mathbf{K}_X(iu) - iux\big\} du \\ &= \frac{1}{2\pi i} \int_{-i\infty}^{+i\infty} \exp\big\{\mathbf{K}_X(z) - zx\big\} dz \\ &= \frac{1}{2\pi i} \int_{C} \exp\big\{\mathbf{K}_X(z) - zx\big\} dz, \end{aligned} \tag{2.3}$$

where $z = iu$, $du = dz/i$ and $i = \sqrt{-1}$ is the imaginary unit and we have defined $\phi_X(u) = \mathbf{M}_X(iu)$ as a characteristic function where we assume that it is integrable. The limits of integration in (2.3)

indicate a contour integral up the imaginary axis C and this formula becomes of use by the methods of complex integration which permit us to replace the path C by any other path starting and ending at the same place i.e. by any contour running up a line like $\mathrm{Re}(z) = \hat{u}$, (see Stalker [90], p.77 & Wintner [97], p.14). The value of $\hat{u}$ has to be one for which $\mathbf{M}_X(\hat{u}) < \infty$. This is called the Fourier-Mellin integral and is a standard result in the theory of Laplace transforms, (see Schiff [87], Ch. 4). Thus the integral in (2.3) can also be expressed as

$$f(x) = \frac{1}{2\pi i} \int_{\hat{u}-i\infty}^{\hat{u}+i\infty} \exp\big\{\mathbf{K}_X(z) - zx\big\} dz. \tag{2.4}$$

2.3 Simple Saddlepoint Approximations

We will provide a brief review of the saddlepoint method, which originated with Daniels [14], before specializing the results to our context of circular and spherical models. We look at the saddlepoint approximation through the inversion of a Fourier transformation and the use of the cumulant generating function.

The key to the saddlepoint method is to choose the path of integration, i.e. $\hat{u}$ in (2.4). Consider the following choice: set $\hat{u} = \hat{u}(x) \in \mathbb{R}$ that satisfies the following saddlepoint equation

$$\mathbf{K}_X^{'}(u) - x = 0. \tag{2.5}$$

In the univariate case, Daniels [14] proves that under general conditions the saddlepoint function $\mathbf{K}^{'}(u) = \xi$, say, has a unique real root $\hat{u}$ in the legitimate support $-u_1 < u < u_2$ where $0 \leqslant u_1 < \infty$ and $0 \leqslant u_2 < \infty$ for every $a < x < b$ such that the CDF has a support $0 < F(x) < 1$. We shall write

$$\mathbf{M}(u) = \mathbf{e}^{\mathbf{K}(u)} = \int_{-\infty}^{+\infty} \exp(ux) dF(x) \text{ and } \mathbf{M}(u,\xi) = \mathbf{e}^{\mathbf{K}(u)-u\xi} = \int_{-\infty}^{+\infty} \exp\{u(x-\xi)\} dF(x).$$

When $a < \xi < b$, $\mathbf{M}^{'}(-\infty, \xi) = -\infty$ and $\mathbf{M}^{'}(\infty, \xi) = \infty$, and $\mathbf{M}^{'}(u, \xi)$ is strictly increasing with u since $\mathbf{M}^{''}(u) > 0$. So for each $a < \xi < b$ there is a single root $\hat{u}$ of $\mathbf{M}^{'}(u, \xi) = 0$ and hence of $\mathbf{K}^{'}(u) = \xi$. Also $\mathbf{K}^{''}(\hat{u}) = \mathbf{M}^{''}(\hat{u}, \xi)/\mathbf{M}(\hat{u}, \xi)$ so that $0 < \mathbf{K}^{''}(\hat{u}) < \infty$ (convex), and $\hat{u}$ is a simple root and $\mathbf{K}^{'}(\hat{u})$ is a strictly increasing function of $\hat{u}$. This implies that the saddlepoint given by (2.5) must fall in the set of u where $\mathbf{K}^{'}(\hat{u})$ a strictly increases, and this is an important fact to find the appropriate boundary for u.

Expanding the function $g(z) = \mathbf{K}_X(z) - zx$ (x fixed) of the exponent in (2.4) around its minimum $\hat{u}$, say, using Taylor series expansion gives

$$\begin{aligned} g(z) = \mathbf{K}_X(z) - zx &\approx g(\hat{u}) + \frac{g^{'}(\hat{u})}{1!}(z-\hat{u}) + \frac{g^{''}(\hat{u})}{2!}(z-\hat{u})^2 \\ &= \mathbf{K}_X(\hat{u}) - \hat{u}x + \frac{1}{2}\mathbf{K}_X^{''}(\hat{u})(z-\hat{u})^2, \end{aligned} \tag{2.6}$$

where $g^{'}(\hat{u}) = \mathbf{K}_X^{'}(\hat{u}) - x = 0$. Incidentally, the name saddlepoint comes from the shape of the right-hand side of (2.6) for u in a neighborhood of $\hat{u}$. If $z - \hat{u}$ is the complex number $c = a + ib$, then the real part of the right-hand side of (2.6) is of the form $(a, b) \mapsto \alpha + \beta(a^2 - b^2)$, where α and β are real-valued constants. This function has the shape of a saddle (Sahalia and Yu [86]). Viewing as a point in the complex plane and by the convex analysis, $\mathbf{K}_X(z) - zx$ has a minimum at $\hat{u}$ for real z, the modulus of the integrand must have a maximum at $\hat{u}$ on the chosen path, (see, for example, Daniels [14] and Goutis and Casella [27]) and hence $\hat{u}$ is neither a maximum nor a minimum but a saddlepoint of $\mathbf{K}_X(z) - zx$.

On the path of integration relevant for (2.4), set $z = \hat{u} + iv$ with $v \in \mathbb{R}$, hence $z - \hat{u} = iv$ is a purely imaginary complex term and we can rewrite (2.4) in the form

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp\big\{\mathbf{K}_X(\hat{u}+iv) - (\hat{u}+iv)x\big\}dv, \tag{2.7}$$

and (2.6) becomes

$$\mathbf{K}_X(z) - zx \approx \mathbf{K}_X(\hat{u}+iv) - (\hat{u}+iv)x = \mathbf{K}_X(\hat{u}) - \hat{u}x - \frac{1}{2}\mathbf{K}_X^{''}(\hat{u})v^2. \tag{2.8}$$

Taking the exponential terms for both sides of (2.8), we get

$$\exp\big\{\mathbf{K}_X(\hat{u}+iv) - (\hat{u}+iv)x\big\} \approx \exp\big\{\mathbf{K}_X(\hat{u}) - \hat{u}x\big\}\exp\Big\{-\frac{1}{2}\mathbf{K}_X^{''}(\hat{u})v^2\Big\}. \tag{2.9}$$

We then substitute (2.9) in (2.7) to find the saddlepoint approximation for $f(x)$. Here we have that

$$f(x) \approx \frac{1}{2\pi}\exp\big\{\mathbf{K}_X(\hat{u}) - \hat{u}x\big\}\int_{-\infty}^{+\infty}\exp\Big\{-\frac{1}{2}\mathbf{K}_X^{''}(\hat{u})v^2\Big\}dv. \tag{2.10}$$

Setting $v = w/\big[\mathbf{K}_X^{''}(\hat{u})\big]^{1/2}$, the saddlepoint approximation for $f(x)$ in (2.10) becomes

$$\begin{aligned} f(x) &\approx \frac{\exp\big\{\mathbf{K}_X(\hat{u}) - \hat{u}x\big\}}{2\pi[\mathbf{K}_X^{''}(\hat{u})\big]^{1/2}}\left\{\int_{-\infty}^{+\infty}\exp\Big(-\frac{1}{2}w^2\Big)dw\right\} \\ &= \frac{\exp\big\{\mathbf{K}_X(\hat{u}) - \hat{u}x\big\}}{2\pi[\mathbf{K}_X^{''}(\hat{u})\big]^{1/2}}\Big(\sqrt{2\pi}\Big) \\ &= \frac{1}{\sqrt{2\pi\mathbf{K}_X^{''}(\hat{u})}}\exp\big\{\mathbf{K}_X(\hat{u}) - \hat{u}x\big\}. \end{aligned} \tag{2.11}$$

2.4 Refined Saddlepoint Approximations and Motivation

The saddlepoint approximation is optimal in the sense that it is based on the highly efficient numerical method of steepest descents and this efficiency can be improved using higher order expansions. Higher-order saddlepoint expansions can be obtained by expanding the function $g(z)$ in (2.6) around $\hat{u}$ to a higher order using Taylor series expansion as follows:

$$\begin{aligned} g(z) = \mathbf{K}_X(z) - zx &= g(\hat{u}) + \frac{g^{'}(\hat{u})}{1!}(z-\hat{u}) + \frac{g^{''}(\hat{u})}{2!}(z-\hat{u})^2 + \frac{g^{(3)}(\hat{u})}{3!}(z-\hat{u})^3 \\ &\quad + \frac{g^{(4)}(\hat{u})}{4!}(z-\hat{u})^4 + \frac{g^{(5)}(\hat{u})}{5!}(z-\hat{u})^5 + O\big((z-\hat{u})^6\big) \\ &= \mathbf{K}_X(\hat{u}) - \hat{u}x + \frac{1}{2}\mathbf{K}_X^{''}(\hat{u})(z-\hat{u})^2 + \frac{1}{6}\mathbf{K}_X^{(3)}(\hat{u})(z-\hat{u})^3 \\ &\quad + \frac{1}{24}\mathbf{K}_X^{(4)}(\hat{u})(z-\hat{u})^4 + \frac{1}{120}\mathbf{K}_X^{(5)}(\hat{u})(z-\hat{u})^5 + O\big((z-\hat{u})^6\big), \end{aligned} \tag{2.12}$$

where $g^{'}(\hat{u}) = \mathbf{K}_X^{'}(\hat{u}) - x = 0$. Setting $z - \hat{u} = iv$ with $v \in \mathbb{R}$, the expansion in (2.12) becomes

$$\begin{aligned} \mathbf{K}_X(\hat{u}+iv) - (\hat{u}+iv)x &= \mathbf{K}_X(\hat{u}) - \hat{u}x - \frac{1}{2}\mathbf{K}_X^{''}(\hat{u})v^2 - \frac{1}{6}\mathbf{K}_X^{(3)}(\hat{u})iv^3 \\ &\quad + \frac{1}{24}\mathbf{K}_X^{(4)}(\hat{u})v^4 + \frac{1}{120}\mathbf{K}_X^{(5)}(\hat{u})iv^5 + O\big(v^6\big). \end{aligned} \tag{2.13}$$

Next, stopping the expansion (2.13) at order 4 in v and taking the exponential terms for both sides, we get

$$\begin{aligned} \exp\big\{\mathbf{K}_X(\hat{u}+iv) - (\hat{u}+iv)x\big\} &\approx \exp\big\{\mathbf{K}_X(\hat{u}) - \hat{u}x\big\} \exp\Big\{-\frac{1}{2}\mathbf{K}_X^{''}(\hat{u})v^2\Big\} \\ &\quad \cdot \exp\Big\{-\frac{1}{6}\mathbf{K}_X^{(3)}(\hat{u})iv^3\Big\} \exp\Big\{\frac{1}{24}\mathbf{K}_X^{(4)}(\hat{u})v^4\Big\} \\ &= \exp\big\{\mathbf{K}_X(\hat{u}) - \hat{u}x\big\} \exp\Big\{-\frac{1}{2}\mathbf{K}_X^{''}(\hat{u})v^2\Big\} \\ &\quad \cdot \left\{1 - \frac{1}{6}\mathbf{K}_X^{(3)}(\hat{u})iv^3 - \frac{1}{72}\Big(\mathbf{K}_X^{(3)}(\hat{u})\Big)^2 v^6 + \ldots\right\} \\ &\quad \cdot \left\{1 + \frac{1}{24}\mathbf{K}_X^{(4)}(\hat{u})v^4 + \frac{1}{1152}\Big(\mathbf{K}_X^{(4)}(\hat{u})\Big)^2 v^8 + \ldots\right\} \\ &\approx \exp\big\{\mathbf{K}_X(\hat{u}) - \hat{u}x\big\} \exp\Big\{-\frac{1}{2}\mathbf{K}_X^{''}(\hat{u})v^2\Big\} \\ &\quad \left\{1 - \frac{1}{6}\mathbf{K}_X^{(3)}(\hat{u})iv^3 + \frac{1}{24}\mathbf{K}_X^{(4)}(\hat{u})v^4 - \frac{1}{72}\Big(\mathbf{K}_X^{(3)}(\hat{u})\Big)^2 v^6\right\}, \end{aligned} \tag{2.14}$$

where the last term in (2.14) comes from the quadratic term in expanding $e^x = 1 + x + \frac{1}{2}x^2 + O(x^3)$.

We then substitute (2.14) in (2.7) to find the saddlepoint approximation for $f(x)$. Here we have

$$\begin{aligned}
f(x) &\approx \frac{1}{2\pi}\exp\{\mathbf{K}_X(\hat{u})-\hat{u}x\}\int_{-\infty}^{+\infty}\exp\Big\{-\frac{1}{2}\mathbf{K}_X''(\hat{u})v^2\Big\}\\
&\quad\Big\{1-\frac{1}{6}\mathbf{K}_X^{(3)}(\hat{u})iv^3+\frac{1}{24}\mathbf{K}_X^{(4)}(\hat{u})v^4-\frac{1}{72}\Big(\mathbf{K}_X^{(3)}(\hat{u})\Big)^2v^6\Big\}dv\\
&=\frac{\exp\{\mathbf{K}_X(\hat{u})-\hat{u}x\}}{2\pi}\Big\{\int_{-\infty}^{+\infty}\exp\Big(-\frac{1}{2}\mathbb{K}_X''(\hat{u})v^2\Big)dv\\
&\qquad-\frac{1}{6}\mathbf{K}_X^{(3)}(\hat{u})i\int_{-\infty}^{+\infty}\exp\Big(-\frac{1}{2}\mathbf{K}_X''(\hat{u})v^2\Big)v^3dv\\
&\qquad+\frac{1}{24}\mathbf{K}_X^{(4)}(\hat{u})\int_{-\infty}^{+\infty}\exp\Big(-\frac{1}{2}\mathbf{K}_X''(\hat{u})v^2\Big)v^4dv\\
&\qquad-\frac{1}{72}\big(\mathbf{K}_X^{(3)}(\hat{u})\big)^2\int_{-\infty}^{+\infty}\exp\Big(-\frac{1}{2}\mathbf{K}_X''(\hat{u})v^2\Big)v^6dv\Big\}. \qquad (2.15)
\end{aligned}$$

Setting $v=w/\big(\mathbf{K}_X''(\hat{u})\big)^{1/2}$, the saddlepoint approximation for $f(x)$ in (2.15) becomes

$$\begin{aligned}
f(x) &= \frac{\exp\{\mathbf{K}_X(\hat{u})-\hat{u}x\}}{2\pi[\mathbf{K}_X''(\hat{u})]^{1/2}}\Big\{\int_{-\infty}^{+\infty}\exp\Big(-\frac{1}{2}w^2\Big)dw\\
&\qquad-\frac{1}{6}\Big(\mathbf{K}_X''(\hat{u})\Big)^{-3/2}\mathbf{K}_X^{(3)}(\hat{u})i\int_{-\infty}^{+\infty}\exp\Big(-\frac{1}{2}w^2\Big)w^3dw\\
&\qquad+\frac{1}{24}\Big(\mathbf{K}_X''(\hat{u})\Big)^{-2}\mathbf{K}_X^{(4)}(\hat{u})\int_{-\infty}^{+\infty}\exp\Big(-\frac{1}{2}w^2\Big)w^4dw\\
&\qquad-\frac{1}{72}\Big(\mathbf{K}_X''(\hat{u})\Big)^{-3}\big(\mathbf{K}_X^{(3)}(\hat{u})\big)^2\int_{-\infty}^{+\infty}\exp\Big(-\frac{1}{2}w^2\Big)w^6dw\Big\}\\
&=\frac{\exp\{\mathbf{K}_X(\hat{u})-\hat{u}x\}}{2\pi[\mathbf{K}_X''(\hat{u})]^{1/2}}\Big\{\sqrt{2\pi}-0+\frac{1}{24}\Big(\mathbf{K}_X''(\hat{u})\Big)^{-2}\mathbf{K}_X^{(4)}(\hat{u})3\sqrt{2\pi}\\
&\qquad-\frac{1}{72}\Big(\mathbf{K}_X''(\hat{u})\Big)^{-3}\big(\mathbf{K}_X^{(3)}(\hat{u})\big)^2 15\sqrt{2\pi}\Big\}\\
&=\frac{1}{\sqrt{2\pi\mathbf{K}_X''(\hat{u})}}\exp\{\mathbf{K}_X(\hat{u})-\hat{u}x\}\Big\{1+\frac{1}{8}\frac{\mathbf{K}_X^{(4)}(\hat{u})}{\big(\mathbf{K}_X''(\hat{u})\big)^2}-\frac{5}{24}\frac{\big(\mathbf{K}_X^{(3)}(\hat{u})\big)^2}{\big(\mathbf{K}_X''(\hat{u})\big)^3}\Big\}\\
&=\frac{1}{\sqrt{2\pi\mathbf{K}_X''(\hat{u})}}\exp\{\mathbf{K}_X(\hat{u})-\hat{u}x\}\Big\{1+\frac{1}{8}\kappa_4(\hat{u})-\frac{5}{24}\kappa_3^2(\hat{u})\Big\}, \qquad (2.16)
\end{aligned}$$

where $\kappa_j(\hat{u})=\mathbf{K}^{(j)}(\hat{u})/[\mathbf{K}''(\hat{u})]^{j/2},\ j=3,4$ and the result follows from the facts that

$$\begin{aligned}
&\int_{-\infty}^{+\infty}\exp\Big(-\frac{1}{2}w^2\Big)dw=\sqrt{2\pi}, && \int_{-\infty}^{+\infty}\exp\Big(-\frac{1}{2}w^2\Big)w^3dw=0\\
&\int_{-\infty}^{+\infty}\exp\Big(-\frac{1}{2}w^2\Big)w^4dw=3\sqrt{2\pi}, && \int_{-\infty}^{+\infty}\exp\Big(-\frac{1}{2}w^2\Big)w^6dw=15\sqrt{2\pi}. \qquad (2.17)
\end{aligned}$$

The function

$$\frac{1}{\sqrt{2\pi \mathbf{K}_X''(\hat{u})}} \exp\{\mathbf{K}_X(\hat{u}) - \hat{u}x\} = \hat{f}_1(x), \text{say}, \tag{2.18}$$

in (2.16) is often called the unnormalized first-order saddlepoint density approximation to $f(x)$. Its error of approximation is much better than the Taylor series approximation to a function. The saddlepoint is also second-order asymptotics, and can have decreased error term, which yields a big improvement in accuracy for the approximation of some function, (see, Goutis & Casella [27]). The unnormalized second-order saddlepoint density approximation to $f(x)$ is given by,

$$\hat{f}_2(x) = \hat{f}_1(x)(1 + T), \tag{2.19}$$

where

$$T = \frac{1}{8}\kappa_4(\hat{u}) - \frac{5}{24}\kappa_3^2(\hat{u}). \tag{2.20}$$

The first order saddlepoint functions in (2.18) and in (2.19) will not, in general, integrate to one, although it will usually not be far off and can be improved by renormalization, that is by computing numerically the normalization constant c where

$$c = \int_{-\infty}^{\infty} \hat{f}_1(x)dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi \mathbf{K}_X''(\hat{u})}} \exp\{\mathbf{K}_X(\hat{u}) - \hat{u}x\}dx \neq 1, \tag{2.21}$$

and the normalized first order saddlepoint density approximation $\bar{f}(x)$ to $f(x)$ is given by

$$\begin{aligned} \bar{f}(x) &= \frac{\hat{f}_1(x)}{\int \hat{f}_1(x)dx} \\ &= c^{-1}\frac{1}{\sqrt{2\pi \mathbf{K}_X''(\hat{u})}} \exp\{\mathbf{K}_X(\hat{u}) - \hat{u}x\}, \end{aligned} \tag{2.22}$$

which is a proper density and integrates to one. Note that choosing $\hat{u} = \hat{u}(x))$ in (2.5) is an application of a method called steepest ascent (Huzurbazer [30]). This method takes advantage of the fact that, since $\hat{u}(x)$ is an extreme point, the function is falling away rapidly as we move away from this point (Kolassa [50]).

2.5 Tilting and Saddlepoint Approximations

The saddlepoint approximation is used to overcome the inadequacy of the normal approximation in the tails by tilting the random variable in such away that the normal approximation is evaluated at a point near the mean (Paolella [79], pp. 170-171) i.e. by using the normal distribution to approximate the true distribution of the tilted random variable. Let T_u be a random variable having density

$$f_{T_u}(x;u) = \frac{\exp\{ux\}f(x)}{\mathbf{M}_X(u)} = \frac{\exp\{ux\}f(x)}{\exp\{\mathbf{K}_X(u)\}} = \exp\big\{ux - \mathbf{K}_X(u)\big\}f(x), \tag{2.23}$$

for some $u \in (a,b)$. This collection of densities define a tilted regular exponential family indexed by u. Density $f_{T_u}(x;u)$ is the u-tilted density and T_u is used as a tilted random variable. The mean and variance of the canonical sufficient X_u are $E(X_u) = \mathbf{K}_X^{'}(u)$ and $Var(X_u) = \mathbf{K}_X^{''}(u)$, respectively.

The Esscher [19] tilting method is an indirect Edgeworth expansion that consists of two steps: (i) First $f(x)$ is written in terms of $f_{T_u}(x;u)$ using (2.23) i.e.

$$f(x) = \exp\big\{\mathbf{K}_X(u) - ux\big\}f_{T_u}(x;u), \tag{2.24}$$

and then (ii) $f(x;u)$ is Edgeworth expanded for a judicious choice of $u \in (a,b)$. We say indirect because the Edgeworth expansion is not for $f(x)$ directly when $u = 0$, but for this specially chosen u-tilted member of exponential family. Step (ii) entails a choice for u such that the Edgeworth approximation for $f(x;u)$ is as accurate as possible. We know that, in general, the normal approximation to the distribution of a random variable X is accurate near the mean of X, but degrades in the tails. As such, we are motivated to choose an u such that x is close to the mean of the tilted distribution. In particular, we would like to find a value u so that the Edgeworth expansion of $f(x;u)$ is centred its mean. Formally, this amounts to choosing $u = \hat{u}$ to solve

$$E(X_u) = \mathbf{K}_X^{'}(\hat{u}) = x, \tag{2.25}$$

or the saddlepoint equation. The Edgeworth expansion for $f(x;u)$ at its mean, $\mathbf{K}_X^{'}(\hat{u})$, is given by

$$f_{T_u}(x;\hat{u}) \approx \frac{1}{\sqrt{2\pi\mathbf{K}_X^{''}(\hat{u})}}\Big\{1 + \frac{1}{8}\kappa_4(\hat{u}) - \frac{5}{24}\kappa_3^2(\hat{u})\Big\}, \tag{2.26}$$

(Butler [9], p. 157). Hence substitute (2.26) into (2.24) yields

$$\begin{aligned} f(x) &= \exp\big\{\mathbf{K}_X(\hat{u}) - \hat{u}x\big\}f_{T_u}(x;\hat{u}) \\ &\approx \frac{1}{\sqrt{2\pi\mathbf{K}_X^{''}(\hat{u})}}\exp\big\{\mathbf{K}_X(\hat{u}) - \hat{u}x\big\}\Big\{1 + \frac{1}{8}\kappa_4(\hat{u}) - \frac{5}{24}\kappa_3^2(\hat{u})\Big\}, \end{aligned} \tag{2.27}$$

which is the second-order saddlepoint version.

We illustrate the unnormalizing first order saddlepoint approximations and normalizing second order saddlepoint approximations for obtaining accurate expression for the noncentral chi-square density. We explicitly use the unnormalizing first order saddlepoint approximations and the unnormalizing second order saddlepoint approximations for obtaining highly accurate approximations for the normalizing constant and the mean resultant length of the von Mises distribution on the circle, the normalizing constants for the Fisher and the real Bingham distributions on the sphere in the following sections.

2.6 Noncentral Chi-square Distribution

Suppose that the function we wish to approximate via the saddlepoint technique is a noncentral chi-square density function with $k = 2$ degrees of freedom and noncentrality parameter α. The noncentral chi-square variable is derived from p normal random variable. Let X_1 and X_2 be mutually stochastically independent random variables. When

$$\mathbf{x} = (X_1, X_2)^T \sim N_2\left(\begin{bmatrix} \kappa \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \tag{2.28}$$

It is known that $s = \mathbf{x}^T\mathbf{x} = x_1^2 + x_2^2$ is distributed as a noncentral chi-square random variable with $k = 2$ degrees of freedom and noncentrality parameter $\alpha = \sum_{i=1}^{2}(\mu_i)^2 = \kappa^2 \geq 0$. To show this fact, consider the moment generating function of the random variable S, $\mathbf{M}_S(u)$, as follows:

$$\mathbf{M}_S(u) = \mathbf{E}\big(\mathbf{e}^{uS}\big) = \mathbf{E}\Big(\mathbf{e}^{u\sum_{i=1}^{2} X_i^2}\Big) = \prod_{i=1}^{2}\mathbf{E}\Big(\mathbf{e}^{uX_i^2}\Big). \tag{2.29}$$

$\mathbf{E}\Big(\mathbf{e}^{uX_i^2}\Big)$ in (2.29) is rewritten as follows:

$$\begin{aligned}
\mathbf{E}\Big(\mathbf{e}^{uX_i^2}\Big) &= \int_{-\infty}^{+\infty} \exp\big(ux_i^2\big)\frac{1}{\sqrt{2\pi}}\exp\Big\{-\frac{1}{2}\big(x_i-\kappa\big)^2\Big\}dx_i \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\Big\{-\frac{1}{2}(1-2u)x_i^2 + x_i\kappa - \frac{1}{2}\kappa^2 - \frac{u\kappa^2}{1-2u} + \frac{u\kappa^2}{1-2u}\Big\}dx_i \\
&= \exp\Big\{\frac{u\kappa^2}{1-2u}\Big\}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\Big\{-\frac{1}{2}(1-2u)\Big(x_i - \frac{\kappa}{1-2u}\Big)^2\Big\}dx_i \\
&= \frac{1}{\sqrt{1-2u}}\exp\Big\{\frac{u\kappa^2}{1-2u}\Big\}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}(1/\sqrt{1-2u})}\exp\Big\{-\frac{1}{2}\Big(\frac{x_i - \frac{\kappa}{1-2u}}{(1/\sqrt{1-2u})}\Big)^2\Big\}dx_i \quad (*) \\
&= (1-2u)^{-1/2}\exp\Big\{\frac{u\kappa^2}{1-2u}\Big\}, \qquad u < 1/2.
\end{aligned} \tag{2.30}$$

Note that the integration in $(*)$ is equal to one, because the function in the integration corresponds to the probability density function of the normal distribution with mean $\kappa/(1-2u)$ and variance $1/(1-2u)$. Accordingly, the moment generating function of S is given by:

$$\begin{aligned}
\mathbf{M}_S(u) &= \prod_{i=1}^{2} \mathbf{E}\Big(\mathbf{e}^{uX_i^2}\Big) \\
&= \mathbf{E}\Big(\mathbf{e}^{uX_1^2}\Big)\mathbf{E}\Big(\mathbf{e}^{uX_2^2}\Big) \\
&= (1-2u)^{-1/2}\exp\Big\{\frac{u\kappa^2}{1-2u}\Big\}(1-2u)^{-1/2}\exp\Big\{\frac{u\times 0}{1-2u}\Big\} \\
&= (1-2u)^{-1}\exp\Big\{\frac{\kappa^2 u}{1-2u}\Big\}, \qquad u<1/2,
\end{aligned} \tag{2.31}$$

which is equivalent to the moment generating function of a noncentral chi-square distribution with 2 degrees of freedom and noncentrality parameter equal to α where $\alpha = \sum_{i=1}^{2}(\mu_i)^2 = \kappa^2 \geq 0$. The exact noncentral chi-square density has no closed form, and is usually written (Ravishanker and Dipak [81], p.165 and Tanizaki [91], p.118)

$$\begin{aligned}
f(s) &= \frac{1}{2}\exp\{-(s+\alpha)/2\}I_0(\sqrt{\alpha s}), \qquad s>0 \\
&= \frac{1}{2}\exp\{-(s+\alpha)/2\}\sum_{j=0}^{\infty}\frac{\big(\alpha s/4\big)^j}{[\Gamma(j+1)]^2},
\end{aligned} \tag{2.32}$$

where $I_0(\cdot)$ is a modified Bassel function of the first kind and order $\nu = 0$ (see, for example, Abramowitz & Stegun [1], p.376, Mardia [59], p.57 and Mardia & Jupp [70], p.349). From (2.31) the cumulant generating function $\mathbf{K}_S(u)$ for s is given by

$$\begin{aligned}
\mathbf{K}_S(u) = \log\Big(\mathbf{M}_S(u)\Big) &= \log\frac{\exp\{\alpha u/(1-2u)\}}{1-2u} \\
&= \frac{\alpha u}{1-2u} - \log(1-2u), \qquad u\in\Big(-\infty,\frac{1}{2}\Big).
\end{aligned} \tag{2.33}$$

Figure (2.1) plots $\mathbf{K}_S(u)$ versus u for the noncentral chi-square distribution with noncentrality parameter α equal to unity. The values of the graph range from $-\infty$ as $u\downarrow -\infty$ to ∞ as $u\uparrow \frac{1}{2}$, and the function $\mathbf{K}_S(u)$ is always a strictly convex function when evaluated over $(-\infty, 1/2)$ so $\mathbf{K}_S''(u) > 0$ and the square root is well-defined.

The saddlepoint that is associated with s can be obtained as follows. The first derivative of

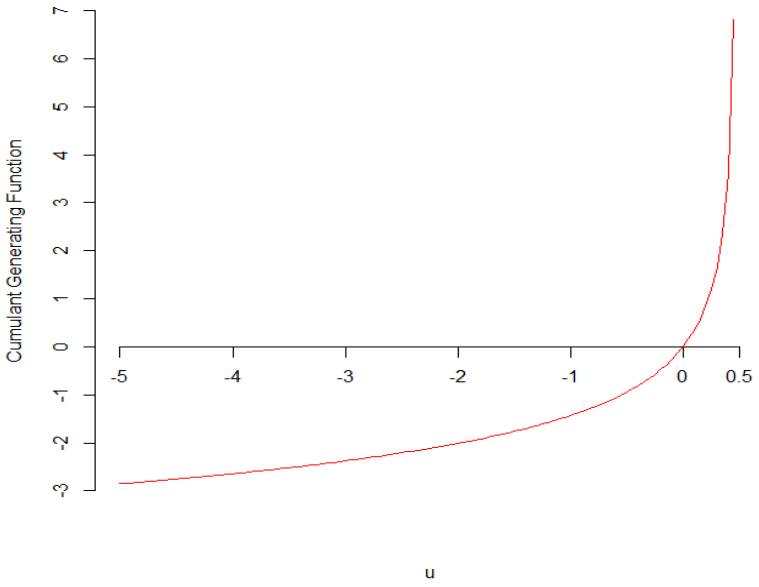

Figure 2.1: The cumulant generating function $\mathbf{K}_S(u)$ for noncentral Chi-square distribution versus u.

$\mathbf{K}_S(u)$ is given by

$$\begin{aligned}\mathbf{K}_S^{'}(u) &= \frac{(1-2u)\alpha+2\alpha u}{(1-2u)^2}+\frac{2}{(1-2u)}\\ &= \frac{\alpha-2\alpha u+2\alpha u+2(1-2u)}{(1-2u)^2}\\ &= \frac{\alpha+2-4u}{(1-2u)^2}.\end{aligned} \tag{2.34}$$

Solve $\mathbf{K}_S^{'}(u(s))=s$ for $u(s)$ in terms of s we get,

$$\begin{aligned}\mathbf{K}_S^{'}(u(s))-s &= 0\\ \frac{\alpha+2-4u(s)}{(1-2u(s))^2}-s &= 0\\ \frac{\alpha+2-4u(s)-s(1-2u(s))^2}{(1-2u(s))^2} &= 0\\ \alpha+2-4u(s)-s+4su(s)-4s\big(u(s)\big)^2 &= 0\\ 4s\big(u(s)\big)^2+4u(s)-4su(s)-2+s-\alpha &= 0\\ 4s\big(u(s)\big)^2+4s\Big(\frac{1-s}{s}\Big)u(s)+(-2+s-\alpha) &= 0\\ \big(u(s)\big)^2+\Big(\frac{1-s}{s}\Big)u(s)+\Big(\frac{-2+s-\alpha}{4s}\Big) &= 0.\end{aligned} \tag{2.35}$$

The equation (2.35) is a quadratic and we may use the completing the square technique to solve it. Move the constant to the other side, add the square of half the coefficient of $u(s)$ to both sides, factor the trinomial square and finally take the square root of both sides we get,

$$\hat{u}(s) = -\frac{1-s}{2s} \pm \frac{\sqrt{(4+4s\alpha)}}{4s}. \tag{2.36}$$

Hence

$$\hat{u}(s) = \frac{-2+2s+\sqrt{(4+4s\alpha)}}{4s} \qquad \text{or} \qquad \hat{u}(s) = \frac{-2+2s-\sqrt{(4+4s\alpha)}}{4s}. \tag{2.37}$$

The positive root in (2.37) is not feasible since $\hat{u}(s)$ would take values greater than $1/2$ as $s > 0$. The cumulant generating function $\mathbf{K}_S(u)$ is only defined for $u < 1/2$ and hence the saddlepoint $\hat{u}(s)$ is the negative root of the quadratic equation (2.35),

$$\hat{u}(s) = \hat{u}(s,\alpha) = -\frac{1}{4s}\Big\{2-2s+\sqrt{(4+4s\alpha)}\Big\}, \qquad s > 0 \quad \text{and} \quad \alpha \geqslant 0. \tag{2.38}$$

The plot of the saddlepoint $\hat{u}(s)$ versus s is shown in Figure (2.2) where the noncentrality parameter α is equal to 1.

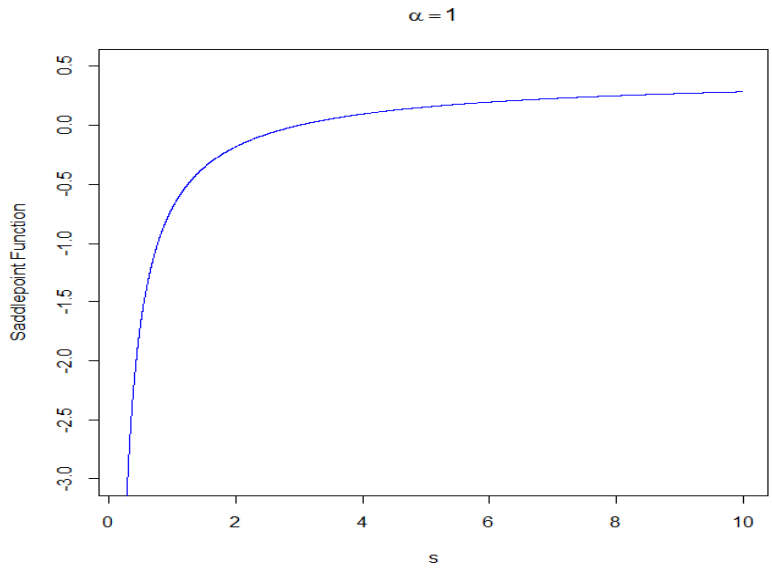

Figure 2.2: Saddlepoint function $\hat{u}(s)$ versus s

The cumulant generation function and the second derivative of $\mathbf{K}_S(u)$ about the saddlepoint

$\hat{u}(s)$ are given by

$$\begin{aligned}\mathbf{K}_S(\hat{u}(s)) &= \frac{\alpha\hat{u}(s)}{1-2\hat{u}(s)} - \log(1-2\hat{u}(s)) \\ &= \frac{\frac{-\alpha}{4s}\big[2-2s+\sqrt{(4+4s\alpha)}\big]}{1+\frac{1}{2s}\big[2-2s+\sqrt{(4+4s\alpha)}\big]} - \log\Big\{1+\frac{1}{2s}\big[2-2s+\sqrt{(4+4s\alpha)}\big]\Big\}, \quad (2.39)\end{aligned}$$

and

$$\begin{aligned}\mathbf{K}_S''(\hat{u}(s)) &= \frac{-4(1-2\hat{u}(s))^2 + 4(\alpha+2-4\hat{u}(s))(1-2\hat{u}(s))}{(1-2\hat{u}(s)^4} \\ &= \frac{-4(1-2\hat{u}(s)) + 4(\alpha+2-4\hat{u}(s))}{(1-2\hat{u}(s))^3} \\ &= \frac{-4+8\hat{t}(s)+4\alpha+8-16\hat{t}(s)}{(1-2\hat{u}(s))^3} \\ &= \frac{4\alpha+4-8\hat{u}(s)}{(1-2\hat{u}(s))^3} \\ &= \frac{4\alpha+4-8\hat{u}(s)}{\Big\{1+\frac{1}{2s}\big[2-2s+\sqrt{(4+4s\alpha)}\big]\Big\}^3}. \quad (2.40)\end{aligned}$$

The unnormalized first-order saddlepoint density approximation to $f(s)$ is given by

$$\hat{f}_1(s) = \frac{1}{\sqrt{2\pi\mathbf{K}_S''(\hat{u})}} \exp\big\{\mathbf{K}_S(\hat{u}) - \hat{u}s\big\}. \quad (2.41)$$

Figure (2.3), for instance, shows comparative plots of the true density $f(s)$ (solid line) with the unnormalized first-order saddlepoint density approximation $\hat{f}_1(s)$ (dashed line) with various values of the noncentrality parameter. Note that when increasing the values of the noncentrality parameter the relative error of each approximation stays bounded under suitable asymptotic regimes.

Note that when $\alpha = 0$, the distribution reduces to the centrality case and the exact probability density function for the noncentral distribution reduces to an exponential density with rate parameter $\frac{1}{2}$,

$$f(s) = \frac{1}{2}\exp\Big\{-\frac{1}{2}s\Big\}, \qquad s > 0. \quad (2.42)$$

For the saddlepoint approximation, we find

$$\hat{u}(s) = \hat{u}(s,\alpha) = -\frac{1}{4s}(4-2s) = -\frac{1}{2s}(2-s) = -\frac{1}{2s}\eta, \qquad s > 0, \quad (2.43)$$

where $\eta = 2-s$. The cumulant generating function $\mathbf{K}_S(t)$ and the second derivative of $\mathbf{K}_S(t)$ about the saddlepoint $\hat{u}(s)$ are given by

$$\mathbf{K}_S(\hat{u}(s)) = -\log(1-2\hat{u}(s)) = -\log\Big(1+\frac{1}{s}\eta\Big), \quad (2.44)$$

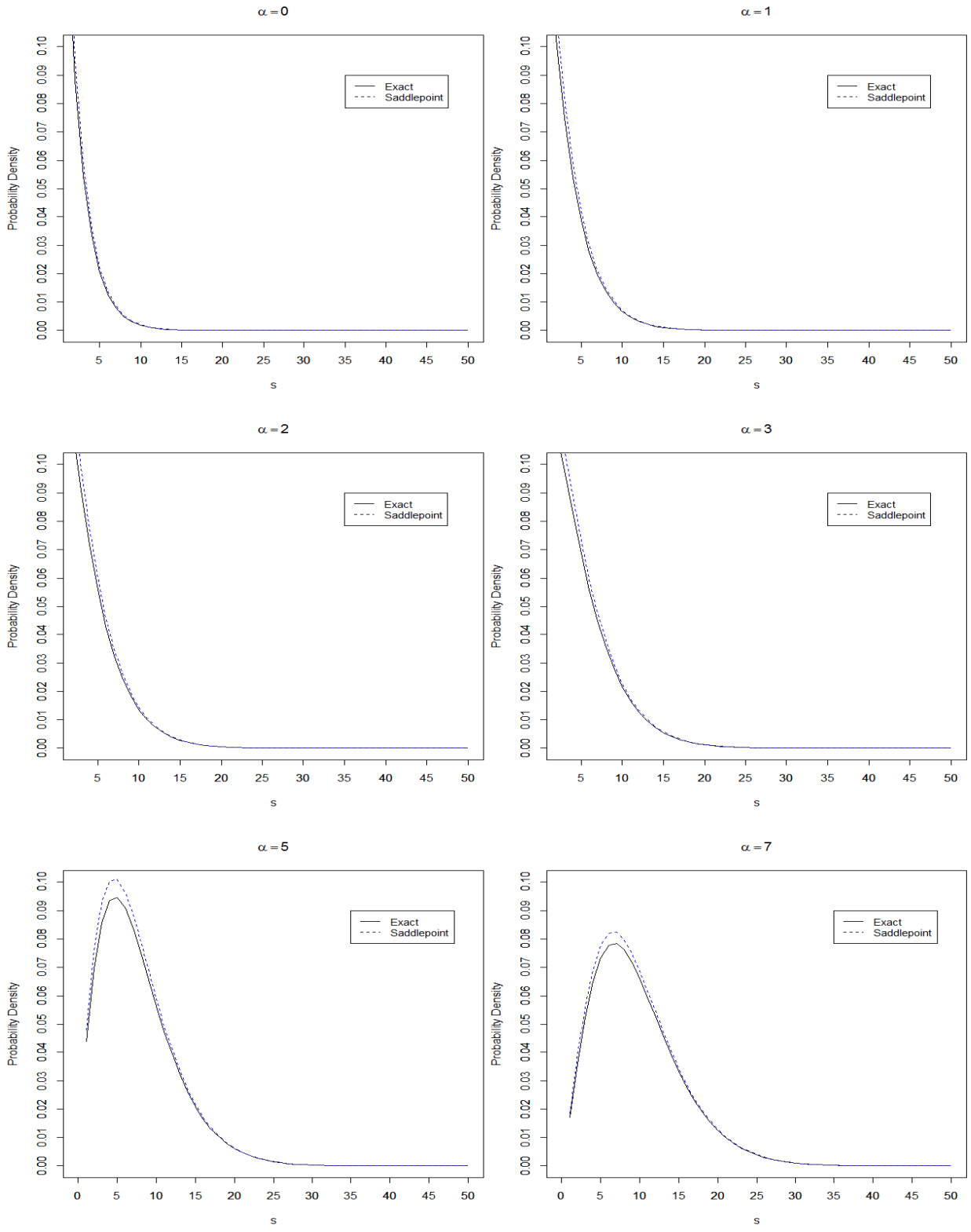

Figure 2.3: Exact $f(s)$ (solid curve) and first-order saddlepoint density approximation $\hat{f}_1(s)$ (dashed line) versus s for noncentral chi-square distribution with various values of the noncentrality parameter $\alpha = 0, 1, 2, 3, 5$ and 7.

and

$$\begin{aligned}\mathbf{K}_S''(\hat{t}(s)) &= \frac{4-8\hat{u}(s)}{\Big(1-2\hat{u}(s)\Big)^3} \\ &= \frac{4}{\Big(1-2\hat{u}(s)\Big)^2},\end{aligned} \tag{2.45}$$

so that for $s>0$ and $\alpha=0$, the unnormalized first order saddlepoint density approximation to $f(s)$ is given by

$$\begin{aligned}\hat{f}_1(s) &= \left[\frac{\big(1-2\hat{u}(s)\big)^2}{8\pi}\right]^{1/2}\exp\Big\{-\ln(1-2\hat{u}(s))-\hat{u}(s)s\Big\} \\ &= \frac{1}{2}\Big(1-2\hat{u}(s)\Big)\Big(\frac{1}{2\pi}\Big)^{1/2}\exp\{-\ln(1-2\hat{u}(s))\}\exp\{-\hat{u}(s)s\} \\ &= \frac{1}{2}\Big(1-2\hat{u}(s)\Big)\Big(\frac{1}{2\pi}\Big)^{1/2}\frac{1}{\big(1-2\hat{u}(s)\big)}\exp\{-\hat{u}(s)s\} \\ &= \Big(\frac{1}{2\pi}\Big)^{1/2}\frac{1}{2}\exp\big\{-\big(-\frac{1}{2s}(2-s)\big)s\big\} \\ &= \Big(\frac{1}{2\pi}\Big)^{1/2}\frac{1}{2}\exp\{\Big(\frac{1}{2}(2-s)\Big)\} \\ &= \frac{\Gamma(1)}{\sqrt{2\pi}1^{1-\frac{1}{2}}e^{-1}}\frac{1}{2}\exp\Big\{-\frac{1}{2}s\Big\}.\end{aligned} \tag{2.46}$$

The shape of $\hat{f}_1(s)$ in (2.46) is the same as that of $f(s)$ in (2.42) but differs from $f(s)$ in the normalization constant. Using Stirling's approximation for $\Gamma(\beta)$,

$$\hat{\Gamma}(\beta)=\sqrt{2\pi}\beta^{\beta-\frac{1}{2}}e^{-\beta}, \tag{2.47}$$

and puting $\beta=1$, we find

$$\hat{f}_1(s)=\frac{\Gamma(1)}{\hat{\Gamma}(1)}f(s)=\texttt{0.92214}f(s). \tag{2.48}$$

This first order saddlepoint approximation to the noncentral chi-square density is also accurate for large α. Assume, for example, two large values for the noncentrality parameter, $\alpha=10$ and $\alpha=14$ and use the R `integrate` function (see Crawley [11], p. 275 and Rizzo [85], p. 330) in the `stats` package, the numerical integration for $\hat{f}_1(s)$ after converting it to a one-dimensional function by fixing $\hat{u}(s)$, $\mathbf{K}_S(\hat{u}(s))$ and $\mathbf{K}_S''(\hat{u}(s))$, yields a numerical value for the normalization constant c given by `1.035951` under $\alpha=10$ with absolute error less than `4.7e-06` and `1.018552` under $\alpha=14$ with absolute error less than `3.6e-07`. Figure (2.4) shows a comparative plot of the true density $f(s)$ (solid line) with the unnormalized first order saddlepoint density approximation $\hat{f}_1(s)$ (dashed

line) and the normalized first order saddlepoint density approximation $\bar{f}(s)$ (dotted line). Here, the unnormalized first order saddlepoint and the renormalized saddlepoint saddlepoint are remarkably accurate. The graphical difference between the normalized saddlepoint approximation $\bar{f}(s)$ and the exact density $f(s)$ is slight since the unnormalized first order saddlepoint density $\hat{f}_1(s)$ mostly captures the proper shape of $f(s)$ but not the correct scaling.

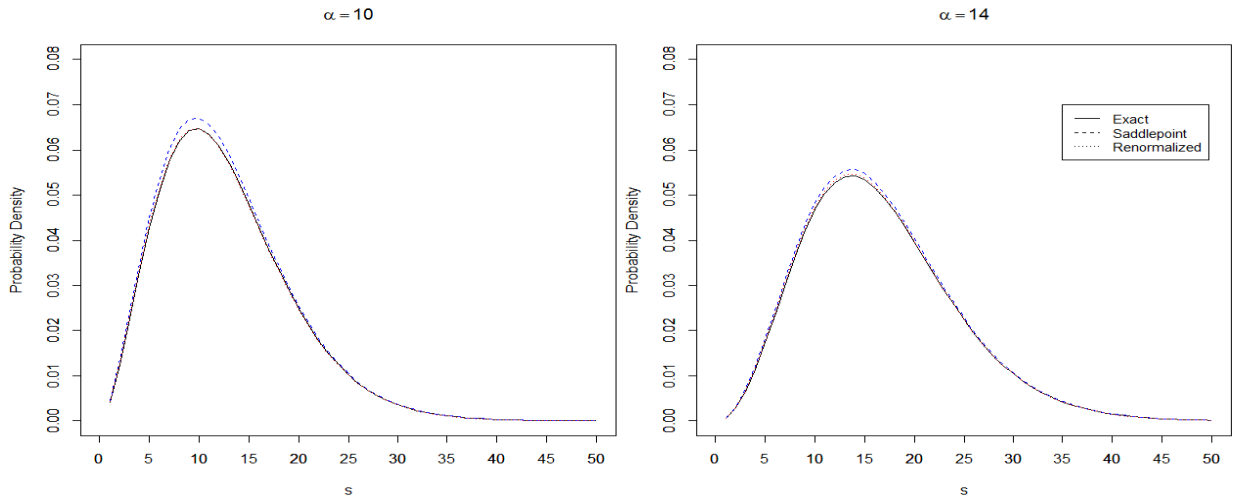

Figure 2.4: Exact $f(s)$ (solid curve), unnormalized first-order saddlepoint density approximation $\hat{f}_1(s)$ (dashed curve) and the normalized first-order saddlepoint density approximation $\bar{f}(s)$ (dotted curve) versus s for noncentral chi-square distribution.

2.7 von Mises (Circular Normal) Distribution

2.7.1 Background

A unit random vector $\mathbf{x}$ on the circle in $\mathbb{R}^2$ has the von Mises distribution, $VM(\mu, \kappa)$, with probability density function given by

$$\begin{aligned} f(\mathbf{x}; \mu, \kappa) &= \{c(\kappa)\}^{-1} \exp\{\kappa \mu^T \mathbf{x}\} \\ &= \{2\pi I_0(\kappa)\}^{-1} \exp\{\kappa \mu^T \mathbf{x}\}, \qquad \kappa > 0, \end{aligned} \tag{2.49}$$

where $\|\mathbf{x}\| = 1$, $\|\mu\| = 1$, κ is known as a concentration parameter, μ is known as a mean direction parameter and $I_0(\kappa)$ denotes the modified Bessel function of the first kind and order zero (Jammalamadaka and SenGupta [18], p. 35 and Mardia and Jupp [70], p. 36).

If we write $\mathbf{x}$ and μ in circular polar coordinates as

$$\begin{aligned} \mathbf{x} &= (\cos\theta, \sin\theta)^T \\ \mu &= (\cos\mu, \sin\mu)^T, \end{aligned} \tag{2.50}$$

then the probability density of θ is

$$\begin{aligned} g(\theta;\mu,\kappa) &= \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa[\cos\theta\cos\mu + \sin\theta\sin\mu]\} \\ &= \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa\cos(\theta-\mu)\}, \qquad \kappa > 0, \end{aligned} \tag{2.51}$$

where the Jacobian of this transformation is unity and the modified Bessel function of the first kind and order zero $I_0(\kappa)$ can be defined by

$$\begin{aligned} I_0(\kappa) &= \frac{1}{2\pi}\int_0^{2\pi} \exp\{\kappa\cos\theta\} d\theta \\ &= \sum_{j=0}^{\infty} \frac{\left(\kappa^2\right)^j}{4^j[j!]^2}. \end{aligned} \tag{2.52}$$

For simplicity let the circular random variable θ have a von Mises (Circular Normal) distribution $VM(\theta;0,\kappa)$ on $(-\pi,\pi)$ with probability density function given by

$$g(\theta;0,\kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa\cos\theta\}, \qquad \kappa > 0. \tag{2.53}$$

The von Mises distribution is related to the bivariate normal distribution as follows. Let X_1 and X_2 be independent normal variables as in (2.28). The joint probability density function of X_1 and X_2 is given by

$$\begin{aligned} f(x_1,x_2) &= f(x_1).f(x_2) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x_1-\kappa)^2\right\}.\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x_2^2\right\} \\ &= \frac{1}{2\pi} \exp\left\{-\frac{1}{2}\left[(x_1^2+x_2^2) - 2\kappa x_1 + \mu^2\right]\right\}, \ -\infty < x_i < \infty, \ \ i = 1,2. \end{aligned} \tag{2.54}$$

Switch to polar coordinates (r,θ),

$$x_1 = r\cos\theta, \qquad \text{and} \qquad x_2 = r\sin\theta. \tag{2.55}$$

Here $x_1^2 + x_2^2 = r^2$ and $\theta = \texttt{atan2}(x_2, x_1)$. The two-argument function `atan2` is a variation of the arctangent function, so that θ is the angle in $[0, 2\pi)$ satisfying $(x_1, x_2) \propto (\cos\theta, \sin\theta)$. The one-argument arctangent function `atan` does not distinguish between diametrically opposite directions

whereas the `atan2` function takes into account the signs of both vector components, and places the angle in the correct quadrant.

The Jacobian of this transformation is given by

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_1}{\partial \theta} \\ \frac{\partial x_2}{\partial r} & \frac{\partial x_2}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r\cos^2\theta + r\sin^2\theta = r, \tag{2.56}$$

and the true joint probability density function of the polar variables (r, θ) is

$$f(r,\theta) = \frac{1}{2\pi} r.\exp\Big\{-\frac{1}{2}(r^2 - 2\kappa r\cos\theta + \kappa^2)\Big\} \qquad 0 < r < \infty, \quad 0 \le \theta < 2\pi. \tag{2.57}$$

Since the range of r does not depend on θ, the conditional distribution of θ given $r = 1$ is von Mises $VM(0, \kappa)$ i.e.

$$\begin{aligned} f(r) &= \int_0^{2\pi} f(r,\theta)d\theta \\ &= r.\exp\big\{-\frac{1}{2}(r^2+\kappa^2)\big\}\Big[\frac{1}{2\pi}\int_0^{2\pi}\exp\big\{\kappa r\cos\theta\big\}d\theta\Big] \\ &= r.\exp\big\{-\frac{1}{2}(r^2+\kappa^2)\big\}I_0(\kappa r), \end{aligned} \tag{2.58}$$

and the conditional distribution of θ for a given r is given by

$$\begin{aligned} f(\theta \mid r) &= \frac{f(r,\theta)}{f(r)} \\ &= \frac{\frac{1}{2\pi} r.\exp\big\{-\frac{1}{2}\big(r^2 - 2\kappa r\cos\theta + \kappa^2\big)\big\}}{r.\exp\big\{-\frac{1}{2}(r^2+\kappa^2)\big\}I_0(\kappa r)} \\ &= \frac{\exp\big\{\kappa r\cos\theta\big\}}{2\pi I_0(\kappa r)}. \end{aligned} \tag{2.59}$$

If $r = 1$, the conditional distribution of θ in (2.59) becomes

$$f(\theta \mid r = 1) = \frac{1}{2\pi I_0(\kappa)}\exp\{\kappa\cos\theta\}, \tag{2.60}$$

which is the probability density function for $VM(0, \kappa)$. Here the exact normalizing constant for the von Mises distribution is given by

$$c(\kappa) = 2\pi I_0(\kappa), \tag{2.61}$$

(see also Mardia,[59], p. 248, Jammalamadaka, and SenGupta[31], p. 35 and Mardia,[65], p. 14).

2.7.2 Saddlepoint Approximations for the Normalizing Constant

The saddlepoint approximation for the noncentral chi-square distribution can be used to approximate the normalizing constant for von Mises distribution. Firstly we need to represent the normalizing constant as follows. The joint probability density function of the polar variables (r, θ) can be written as

$$f(r,\theta) \;=\; f(\theta \mid r)\cdot f(r), \tag{2.62}$$

with respect to $drd\theta$. Here, a convenient and accurate method for estimating $f(r)$ in (2.62) is to use a saddlepoint density approximation $\hat{f}_1(s)$ in (2.41) and put $r = \sqrt{s} = 1$, hence (2.62) becomes

$$\begin{aligned} f(1,\theta) &\approx f(\theta \mid 1)\cdot \hat{f}_1(1) \\ &= \big\{c(\kappa)\big\}^{-1}\exp\{\kappa\cos\theta\}\cdot \hat{f}_1(1). \end{aligned} \tag{2.63}$$

Then the saddlepoint approximation for the normalizing constant $c(\kappa)$ is

$$\begin{aligned} c(\kappa) &\approx \exp\{\kappa\cos\theta\}\frac{\hat{f}_1(1)}{f(1,\theta)} \\ &= \exp\{\kappa\cos\theta\}\frac{\hat{f}_1(1)}{\frac{1}{2\pi}\exp\Big\{-\frac{1}{2}(1-2\kappa\cos\theta+\kappa^2)\Big\}} \\ &= 2\pi\exp\{\kappa\cos\theta\}\frac{\hat{f}_1(1)}{\exp\Big\{-\frac{1}{2}(1+\kappa^2)\Big\}\exp\Big\{\kappa\cos\theta\Big\}} \\ &= 2\pi\hat{f}_1(1)\exp\Big\{\frac{1}{2}(1+\kappa^2)\Big\} \\ &= 2\pi\Big(2\pi\mathbf{K}_S''(\hat{u}(1))\Big)^{-1/2}\exp\big\{\mathbf{K}_S(\hat{u}(1))-\hat{u}(1)\big\}\exp\Big\{\frac{1}{2}(1+\kappa^2)\Big\} \\ &= \big(2\pi\big)^{1/2}\Big(\mathbf{K}_S''(\hat{u}(1))\Big)^{-1/2}\exp\big\{\mathbf{K}_S(\hat{u}(1))-\hat{u}(1)+\frac{1}{2}(1+\kappa^2)\big\} = \hat{c}(\kappa), \text{say}, \end{aligned} \tag{2.64}$$

where the saddlepoint function in (2.38) is

$$\hat{u}(1) = \hat{u}(1,\kappa) = -\frac{1}{4}\Big(\sqrt{4+4\kappa^2}\Big) = -\frac{1}{4}D, \tag{2.65}$$

where $D = (4{+}4\kappa^2)^{1/2}$. Moreover, the cumulant generating function $\mathbf{K}_S(u)$ and the second derivative of $\mathbf{K}_S(u)$ about the saddlepoint $\hat{u}(1,\kappa)$ in (2.39) and (2.40), respectively become

$$\mathbf{K}_S(\hat{u}(1,\kappa)) = \frac{-\frac{1}{4}\kappa^2 D}{1+\frac{1}{2}D} - \ln(1+\frac{1}{2}D), \tag{2.66}$$

and

$$\mathbf{K}_S''(\hat{u}(1,\kappa)) = \frac{4\kappa^2 + 4 + 2D}{\left(1 + \frac{1}{2}D\right)^3}. \tag{2.67}$$

Table (2.1) shows the accuracy of the unnormalizing first order saddlepoint approximations $\hat{c}_{01}(\kappa) = \{\hat{c}(\kappa)\}^{-1}$ and the unnormalizing second order saddlepoint approximations $\hat{c}_{02}(\kappa) = \{\hat{c}(\kappa)\}^{-1}(1+T)$ for the normalizing constant of the von Mises distribution versus the true normalizing constant $c_{01}(\kappa) = \{c(\kappa)\}^{-1}$ with various values of the concentration parameter κ and a second order correction term T given in (2.20). It turns out that for the von Mises distribution, the limiting relative errors of the saddlepoint approximations are zero as concentration goes to infinity, provided the distribution is unimodal. As κ tends to infinity, the limiting behaviour of the saddlepoint approximations tends to unity i.e. the saddlepoint approximations for the normalizing constants of the von Mises distribution closely resembles the true values. Numerical results show that the unnormalizing second order saddlepoint approximations do even better than the unnormalizing first order saddlepoint approximations and these results agree with that for Kume and Wood [54] i.e.

$$\lim_{\kappa\to\infty} \frac{\hat{c}_{01}(\kappa)}{c_{01}(\kappa)} = 1 \text{ and } \lim_{\kappa\to\infty} \frac{\hat{c}_{02}(\kappa)}{c_{01}(\kappa)} = 1. \tag{2.68}$$

For small values of the concentration parameter κ, the limiting behaviour of the first order saddlepoint approximation for the normalizing constant of von Mises converges to $\Gamma(1)/\hat{\Gamma}(1) =$ `0.92214` (Kume and Wood [54]). Note that when $\kappa = 0$, the von Mises distribution $VM(0,\kappa)$ is the uniform distribution and its normalizing constant reduces to $\frac{1}{2\pi} =$ `0.15916`. In this case, $D = 2$, $\hat{t}(1,0) = -\frac{1}{2}$, $K_S''(\hat{t}(1,0)) = 1$, $K_S(\hat{t}(1,0)) = -\ln(2) = -0.69315$ and the saddlepoint approximation for the normalizing constant of von Mises distribution reduces to 0.14677.

2.7.3 Saddlepoint Approximations for the Mean Resultant Length

The saddlepoint approximation is also extremely accurate for approximating the derivative of the log function of the normalizing constant for the von Mises distribution with respect to κ. The log function for $c(\kappa)$ can be written as

$$\log c(\kappa) = \log\Big(2\pi I_0(\kappa)\Big) = \log 2\pi + \log I_0(\kappa). \tag{2.69}$$

Differentiating (2.69) with respect to κ and take the absolute value gives

$$\frac{\partial \log c(\kappa)}{\partial \kappa} = \frac{1}{I_0(\kappa)} I_0'(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)} = A(\kappa), \tag{2.70}$$

κ	Exact $c_{01}(\kappa)$	SPA $\hat{c}_{01}(\kappa)$	$\hat{c}_{01}(\kappa)/c_{01}(\kappa)$	SPA $\hat{c}_{02}(\kappa)$	$\hat{c}_{02}(\kappa)/c_{01}(\kappa)$
0.0	0.15916	0.14677	0.92215	0.14891	0.93559
0.2	0.15606	0.14548	0.93221	0.14722	0.94336
0.4	0.15376	0.14311	0.93074	0.14521	0.93559
0.6	0.14908	0.13887	0.93154	0.13983	0.93795
0.8	0.14592	0.13653	0.93567	0.13703	0.93909
1.0	0.12571	0.11782	0.93722	0.11818	0.94009
5.0	0.00584	0.00549	0.94077	0.00553	0.94626
10	5.652380e-05	5.450647e-05	0.96431	5.517005e-05	0.97605
100	1.482232e-43	1.466921e-43	0.98967	1.470404e-43	0.99202

Table 2.1: Numerical unnormalizing first order saddlepoint approximations and unnormalizing second order saddlepoint approximations for the normalizing constant of the von Mises distribution with various values of the concentration parameter κ.

where $A(\kappa) = I_1(\kappa)/I_0(\kappa) = \rho$, the mean resultant length for the von Mises distribution measures the length of the centre of mass vector $\mathbf{x}$, $0 \leq \rho \leq 1$, and $I_0^{'}(\kappa) = I_1(\kappa)$ (see Mardia and Jupp [70], p. 36). The function $I_1(\kappa)$ is the modified Bessel function of the first kind and order one and can be defined by

$$\begin{aligned} I_1(\kappa) &= \frac{1}{2\pi}\int_0^{2\pi} \cos\theta \exp\{\kappa\cos\theta\} d\theta \\ &= \sum_{j=1}^{\infty} \frac{1}{\Gamma(j+2)\Gamma(j+1)}\left(\frac{\kappa}{2}\right)^{2j+1} \\ &= \frac{(\kappa/2)}{\Gamma(\frac{3}{2})\Gamma(\frac{1}{2})}\int_{-1}^{1} \exp\{\kappa t\}(1-t^2)^{1/2} dt, \end{aligned} \tag{2.71}$$

and the function $A(\kappa)$ has the power series expansion

$$A(\kappa) = \frac{\kappa}{2}\left\{1 - \frac{1}{8}\kappa^2 + \frac{1}{48}\kappa^4 - \frac{11}{3072}\kappa^6 + \ldots\right\}, \tag{2.72}$$

which is useful for small κ (Appendix 1 in Mardia and Jupp [70], p. 350). The joint probability

density function of the polar variables (r, θ) in (2.63) can be written as

$$f(1,\theta) \approx \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa\cos\theta\}.\hat{f}_1(1). \tag{2.73}$$

Then

$$\begin{aligned} I_0(\kappa) &\approx \frac{1}{2\pi}\exp\{\kappa\cos\theta\}\frac{\hat{f}_1(1)}{f(1,\theta)} \\ &= \frac{1}{2\pi}\exp\{\kappa\cos\theta\}\frac{\Big(2\pi\mathbf{K}_S''(\hat{u}(1))\Big)^{-1/2}\exp\big\{\mathbf{K}_S(\hat{u}(1))-\hat{u}(1)\big\}}{\frac{1}{2\pi}\exp\Big\{-\frac{1}{2}(1-2\mu\cos\theta+\mu^2)\Big\}} \\ &= \exp\{\kappa\cos\theta\}\frac{\Big(2\pi\mathbf{K}_S''(\hat{u}(1))\Big)^{-1/2}\exp\big\{\mathbf{K}_S(\hat{u}(1))-\hat{u}(1)\big\}}{\exp\Big\{-\frac{1}{2}(1+\kappa^2)\Big\}\exp\Big\{\kappa\cos\theta\Big\}} \\ &= \left(\frac{1}{2\pi\mathbf{K}_S''(\hat{u}(1))}\right)^{1/2}\exp\big\{\mathbf{K}_S(\hat{u}(1))-\hat{u}(1)+\frac{1}{2}(1+\kappa^2)\big\} = \hat{I}_0(\kappa), \text{say}, \end{aligned} \tag{2.74}$$

Next differentiating (2.74) with respect to κ yields

$$\begin{aligned} \hat{I}_0'(\kappa) &= \left[\left(\frac{1}{2\pi\mathbf{K}_S''(\hat{u}(1))}\right)^{1/2}\left(\mathbf{K}_S'(\hat{u}(1,\kappa))-\hat{u}'(1,\kappa)+\kappa\right)-\frac{1}{2}\left(\frac{\mathbf{K}_S^{(3)2}(\hat{u}(1,\kappa))}{2\pi\mathbf{K}_S''^3(\hat{u}(1,\kappa))}\right)^{1/2}\right] \\ &\quad .\exp\Big\{\mathbf{K}_S(\hat{u}(1,\kappa))-\hat{u}(1,\kappa)+\frac{1}{2}(1+\kappa^2)\Big\} = \hat{I}_1(\kappa), \text{say}, \end{aligned} \tag{2.75}$$

and the derivative of the saddlepoint function in (2.38) with respect to κ is given by

$$\hat{u}'(1,\kappa) = -\frac{\kappa}{\sqrt{4+4\kappa^2}} = -\frac{1}{D}\kappa, \tag{2.76}$$

where $D = (4+4\kappa^2)^{1/2}$. Moreover, the first derivative of the cumulant generating function $\mathbf{K}_S(u)$ and the third derivative of $\mathbf{K}_S(u)$ about the saddlepoint $\hat{u}(1,\kappa)$, respectively become

$$\mathbf{K}_S'(\hat{u}(1,\kappa)) = \frac{\kappa^2+2+D}{\Big(1+\frac{1}{2}D\Big)^2}, \tag{2.77}$$

and

$$\mathbf{K}_S^{(3)}(\hat{u}(1,\kappa)) = \frac{\big(1+\frac{1}{2}D\big)\big(8\kappa+\frac{8}{D}\kappa\big)-\big(4\kappa^2+4+2D\big)\frac{6}{D}\kappa}{\Big(1+\frac{1}{2}D\Big)^4}. \tag{2.78}$$

An analysis of this saddlepoint approximation method $A(\kappa)$ is given in Table 2.2. It shows numerical comparisons between the true value of the mean resultant length of the von Mises distribution $\hat{A}(\kappa) = \hat{I}_1(\kappa)/\hat{I}_0(\kappa)$ and both the unnormalizing first order saddlepoint approximations $\hat{A}_1(\kappa) =$

$\hat{I}_1(\kappa)/\hat{I}_0(\kappa)$ and the unnormalizing second order saddlepoint approximations $\hat{A}_2(\kappa) = \hat{A}_1(\kappa)(1+T)$ with various values of the concentration parameter κ and a second order correction term T given in (2.20). The table also shows that as the concentration parameter κ increases the ratio between the two version of saddlepoint approximations and the true values of $A(\kappa)$ approaches 1, that is the saddlepoint approximation becomes more accurate, but that the unnormalizing second order saddlepoint approximations $\hat{A}_2(\kappa)$ do even better.

κ	Exact $A(\kappa)$	SPA $\hat{A}_1(\kappa)$	$\hat{A}_1(\kappa)/A(\kappa)$	SPA $\hat{A}_2(\kappa)$	$\hat{A}_2(\kappa)/A(\kappa)$
0.2	0.09950	0.08665	0.87082	0.08761	0.88051
0.4	0.19610	0.17255	0.87989	0.17289	0.88164
0.6	0.28726	0.25778	0.89739	0.25825	0.89902
0.8	0.37108	0.33438	0.90111	0.33605	0.90561
1.0	0.44639	0.41069	0.92002	0.41221	0.92342
5.0	0.89338	0.84185	0.94232	0.84632	0.94733
10	0.94860	0.91038	0.95971	0.91297	0.96244
100	0.99499	0.98271	0.98766	0.98511	0.99007
200	0.99750	0.98977	0.99225	0.99402	0.99652

Table 2.2: Numerical unnormalizing first order saddlepoint approximations and unnormalizing second order saddlepoint approximations for the mean resultant length of the von Mises distribution with various values of the concentration parameter κ.

2.8 Fisher Distribution on the Sphere

2.8.1 Background

A unit random vector $\mathbf{x}$ has the $(p-1)$-dimensional von Mises-Fisher (Langevin) distribution $M_p(\mu, \kappa)$ with probability density function given by

$$f(\mathbf{x}; \mu, \kappa) = \big\{c_p(\kappa)\big\}^{-1} \exp\{\kappa \mu^T \mathbf{x}\} = \frac{\kappa^{p/2-1}}{(2\pi)^{p/2} I_{p/2-1}(\kappa)} \exp\{\kappa \mu^T \mathbf{x}\}, \tag{2.79}$$

where $\|\mathbf{x}\| = 1$, $\|\mu\| = 1$, $\kappa > 0$ is known as a concentration parameter, μ is known as a mean direction parameter and $I_v(\kappa)$ denotes the modified Bessel function of the first kind and order v defined by

$$I_v(\kappa) = \frac{1}{2\pi}\int_0^{2\pi} \cos v\theta \exp\{\kappa\cos\theta\}d\theta = \sum_{j=0}^{\infty} \frac{1}{\Gamma(v+j+1)\Gamma(j+1)}\Big(\frac{\kappa}{2}\Big)^{2j+v}, \tag{2.80}$$

(Abramowitz and Stegun [1], p. 375). If $p = 2$ the distribution reduces to the von Mises distribution on the circle and if $p = 3$ the distribution is called Fisher distribution $F(\mu, \kappa)$ on the sphere with probability density function given by

$$\begin{aligned} f(\mathbf{x};\mu,\kappa) &= \big\{c_2(\kappa)\big\}^{-1}\exp\{\kappa\mu^T\mathbf{x}\} \\ &= \left[\frac{(2\pi)^{3/2}I_{1/2}(\kappa)}{\kappa^{1/2}}\right]^{-1}\exp\{\kappa\mu^T\mathbf{x}\} \\ &= \left[\frac{4\pi\sinh\kappa}{\kappa}\right]^{-1}\exp\{\kappa\mu^T\mathbf{x}\} = \frac{\kappa}{2\pi(e^{2\kappa}-1)}\exp\{\kappa\mu^T\mathbf{x}\}, \quad \kappa > 0, \end{aligned} \tag{2.81}$$

where $I_{1/2}(\kappa) = (2\sinh\kappa)/(2\pi\kappa)^{1/2}$ (Mardia et al. [72], p. 431) and $\sinh\kappa = (e^{\kappa} - e^{-\kappa})/2 = (e^{2\kappa}-1)/2$.

If we write $\mathbf{x}$ and μ in spherical polar coordinates as

$$\begin{aligned} \mathbf{x} &= (x_1, x_2, x_3)^T = (\cos\theta, \sin\theta\cos\phi, \sin\theta\sin\phi)^T \\ \mu &= (\mu_1, \mu_2, \mu_3)^T = (\cos\alpha, \sin\alpha\cos\beta, \sin\alpha\sin\beta)^T, \end{aligned} \tag{2.82}$$

where θ is the colatitude, $0 \le \theta < \pi$, and ϕ is the longitude, $0 \le \phi < 2\pi$, (α, β) is the mean direction and κ a measure of the concentration about the mean direction. The Jacobian of this transformation is given by

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_1}{\partial\theta} & \frac{\partial x_1}{\partial\phi} \\ \frac{\partial x_2}{\partial r} & \frac{\partial x_2}{\partial\theta} & \frac{\partial x_2}{\partial\phi} \\ \frac{\partial x_3}{\partial r} & \frac{\partial x_3}{\partial\theta} & \frac{\partial x_3}{\partial\phi} \end{vmatrix} = \begin{vmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta\cos\phi & \cos\theta\cos\phi & -\sin\theta\sin\phi \\ \sin\theta\sin\phi & \cos\theta\sin\phi & -\sin\theta\cos\phi \end{vmatrix} = \sin\theta, \tag{2.83}$$

and

$$\begin{aligned} \mu^T\mathbf{x} &= \cos\theta\cos\alpha + \sin\theta\cos\phi\sin\alpha\cos\beta + \sin\theta\sin\phi\sin\alpha\sin\beta \\ &= \cos\theta\cos\alpha + \sin\theta\sin\alpha\big(\cos\phi\cos\beta + \sin\phi\sin\beta\big) \\ &= \cos\theta\cos\alpha + \sin\theta\sin\alpha\cos(\phi-\beta). \end{aligned} \tag{2.84}$$

Under $r = 1$, the probability density function of (θ, ϕ) for Fisher distribution $F\big((\alpha, \beta), \kappa\big)$ is

$$\begin{aligned} g(\theta, \phi; \alpha, \beta, \kappa) &= \left[\frac{4\pi \sinh \kappa}{\kappa}\right]^{-1} \exp\Big\{\kappa\big[\cos\theta\cos\alpha + \sin\theta\sin\alpha\cos(\phi - \beta)\big]\Big\}\sin\theta \\ &= \left[\frac{2\pi(e^{\kappa} - e^{-\kappa})}{\kappa}\right]^{-1} \exp\Big\{\kappa\big[\cos\theta\cos\alpha + \sin\theta\sin\alpha\cos(\phi - \beta)\big]\Big\}\sin\theta, \quad (2.85) \end{aligned}$$

where $0 \leq \theta,\ \alpha < \pi,\ \ 0 \leq \phi,\ \beta < 2\pi$ and $\kappa > 0$.

For simplicity let the spherical random variables (θ, ϕ) have Fisher (Langevin) distribution $F\big((0,0), \kappa\big)$ on $0 \leq \theta \leq \pi$, and $0 \leq \phi \leq 2\pi$ with probability density function given by

$$\begin{aligned} g(\omega; 0, 0, \kappa) = g(\theta, \phi; 0, 0, \kappa) &= \big\{C_3(\kappa)\big\}^{-1} \exp\big\{\kappa\cos\theta\big\}\sin\theta \\ &= \left[\frac{4\pi \sinh \kappa}{\kappa}\right]^{-1} \exp\big\{\kappa\cos\theta\big\}\sin\theta \\ &= \left[\frac{2\pi(e^{\kappa} - e^{-\kappa})}{\kappa}\right]^{-1} \exp\big\{\kappa\cos\theta\big\}\sin\theta, \quad (2.86) \end{aligned}$$

where $\omega = (\theta, \phi)$, so that with $\alpha = \beta = 0$ as the pole $\mu = (1, 0, 0)^T$, the colatitude θ and the longitude ϕ are independent. Then the probability density function of θ is

$$g(\theta, \kappa) = \frac{\kappa}{2\pi \sinh \kappa} \exp\big\{\kappa\cos\theta\big\}\sin\theta, \quad (2.87)$$

on $[0, \pi)$ and ϕ is uniform on the unit circle with probability density function

$$g(\phi) = \frac{1}{2\pi} \quad (2.88)$$

see, for example, Mardia and Jupp [70], p. 170 and Fisher et al. [23].

2.8.2 Saddlepoint Approximations for the Normalizing Constant

Let the random vector $\mathbf{x} = (x_1, x_2, \ldots, x_p)^T$ follow a multivariate normal distribution with mean vector $\mu = (\kappa, 0, 0, \ldots, 0)^T \in \mathbb{R}^p$ and variance covariance matrix $\Sigma = \mathbf{I}$ where $x_i,\ i = 1, 2, \ldots, p$ are independent, so that

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\mathbf{I}|^{1/2}} \exp\left\{-\frac{1}{2}\big(\mathbf{x} - \mu\big)^T\mathbf{I}^{-1}(\mathbf{x} - \mu)\big)\right\}. \quad (2.89)$$

Then the distribution of $s = \mathbf{x}^T\mathbf{x} = x_1^2 + x_2^2 + \ldots + x_p^2$ is noncentral chi-square with noncentrality parameter $\alpha = \mu^T\mu = \sum_{i=1}^p \mu_i^2 = \kappa^2$ and probability density function given by

$$\begin{aligned} f(s) &= \frac{1}{2}\exp\{-(s + \alpha)/2\}\left(\frac{s}{\alpha}\right)^{(p/4)-(1/2)} I_{(p/2)-1}(\sqrt{\alpha s}), \qquad s > 0 \\ &= \frac{1}{2}\exp\{-(s + \alpha)/2\}\left(\frac{s}{\alpha}\right)^{(p/4)-(1/2)} \sum_{j=0}^{\infty} \frac{\Big(\sqrt{\alpha s}/2\Big)^{\big((p-2)/2\big)+2j}}{j!\Gamma(j + \big((p-2)/2\big) + 1}, \quad (2.90) \end{aligned}$$

(see, for example, Ravishanker and Dipak [81], p.177 and Paolella [79], p. 344) where $I_\nu(.)$ is a modified Bassel function of the first kind and order ν. The moment generating function of a noncentral chi-square distribution with p degrees of freedom and noncentrality parameter equal to α is given by

$$\mathbf{M}_S(u) = \frac{\exp\{u\mu^2/(1-2u)\}}{1-2u} = (1-2u)^{-p/2}\exp\Big(\frac{\alpha u}{1-2u}\Big), \qquad u < 1/2, \tag{2.91}$$

and the cumulant generating function $\mathbf{K}_S(u)$ for s is given by

$$\begin{aligned}\mathbf{K}_S(u) = \log \mathbf{M}_S(u) &= \ln \frac{\exp\{\alpha u/(1-2u)\}}{\big(1-2u\big)^{p/2}} \\ &= \frac{\alpha u}{1-2u} - \frac{p}{2}\log(1-2u), \qquad u \in \Big(-\infty, \frac{1}{2}\Big).\end{aligned} \tag{2.92}$$

The saddlepoint associated with s can be obtained by solving $\mathbf{K}_S^{'}(u) - s = 0$, i.e.,

$$\begin{aligned}\mathbf{K}_S^{'}\big(u(s)\big) - s &= 0 \\ \frac{\alpha}{1-2u(s)} + \frac{2\alpha u(s)}{(1-2u(s))^2} + \frac{p}{(1-2u(s))} - s &= 0 \\ \frac{\alpha + p - 2pu(s)}{(1-2u(s))^2} - s &= 0 \\ 4s\big(u(s)\big)^2 - 2(2s-p)u(s) - (p+\alpha-s) &= 0.\end{aligned} \tag{2.93}$$

The equation (2.93) is a quadratic and we may use completing the square technique to solve it. Move the constant to the other side, add the square of half the coefficient of $u(s)$ to both sides, factor the trinomial square and finally take the square root of both sides. Solving this quadratic equation gives the following two roots

$$\hat{u}(s) = \frac{-p + 2s + \sqrt{(p^2 + 4s\alpha)}}{4s} \qquad \text{or} \qquad \hat{u}(s) = \frac{-p + 2s - \sqrt{p^2 + 4s\alpha)}}{4s}. \tag{2.94}$$

Rearranging and using the facts that (i) the constraint $u < 1/2$ from (2.91), (ii) $\alpha \geq 0$ and (iii) the interior of the support of s is $\mathbb{R}$ (i.e., $s > 0$), easily shows that the negative root is always the correct solution. Then the saddlepoint function is

$$\hat{u}(s) = \hat{u}(s,\alpha) = -\frac{1}{4s}\Big\{p - 2s + \sqrt{(p^2 + 4s\alpha)}\Big\}, \qquad s > 0 \ \text{ and } \ \alpha \geqslant 0. \tag{2.95}$$

In the case ($p = 3$) the joint probability density function for X_1, X_2 and X_3 is given by

$$\begin{aligned}f(x_1, x_2, x_3) &= f(x_1).f(x_2).f(x_3) \\ &= \frac{1}{\big(2\pi\big)^{3/2}}\exp\Big\{-\frac{1}{2}\Big[(x_1^2 + x_2^2 + x_3^2) - 2\kappa x_1 + 2\kappa^2\Big]\Big\}.\end{aligned} \tag{2.96}$$

Switch to spherical polar coordinates, we get

$$\begin{aligned} x_1^2 + x_2^2 + x_3^2 &= r^2\cos^2\theta + r^2\sin^2\theta\cos^2\phi + r^2\sin^2\theta\sin^2\theta \\ &= r^2\big[\cos^2\theta + \sin^2\theta(\cos^2\phi + \cos^2\phi)\big] = r^2 = s, \end{aligned} \tag{2.97}$$

where the Jacobian is equal to $r^2\sin\theta$. The true joint probability density function of the spherical polar variables (r, ω) is given by

$$\begin{aligned} f(r,\omega) = f(r,\theta,\phi) &= \frac{1}{2\pi\sqrt{2\pi}}.r^2\sin\theta\exp\Big\{-\frac{1}{2}\Big[r^2 - 2r\kappa\cos\theta + 2\kappa^2\Big]\Big\} \\ &= \frac{1}{\big(2\pi\big)^{3/2}}.r^2\sin\theta\exp\big\{r\kappa\cos\theta\big\}\cdot\exp\Big\{-\frac{1}{2}\Big[r^2 + 2\kappa^2\Big]\Big\}, \end{aligned} \tag{2.98}$$

where $\omega = (\theta, \phi)$ and it can be written as

$$f(r,\omega) = f(\omega \mid r)\cdot f(r), \tag{2.99}$$

with respect to $drd\omega$. A convenient and accurate method for estimating $f(r)$ in (2.99) is to use a saddlepoint density approximation $\hat{f}_1(s)$ in (2.41) and put $s = r^2 = 1$. The equation (2.99) then becomes

$$\begin{aligned} f(1,\omega) &\approx f(\omega \mid 1).\hat{f}_1(1) \\ &= \big\{c_2(\kappa)\big\}^{-1}\exp\big\{\kappa\cos\theta\big\}\sin\theta.\hat{f}_1(1), \end{aligned} \tag{2.100}$$

and the first order saddlepoint approximation for the normalizing constant $c_2(\kappa)$ is

$$\begin{aligned} \mathrm{c}_2(\kappa) &\approx \exp\big\{\kappa\cos\theta\big\}\sin\theta\frac{\hat{f}_1(1)}{f(1,\omega)} \\ &= \exp\big\{\kappa\cos\theta\big\}\sin\theta\frac{\hat{f}_1(1)}{\big(2\pi\big)^{-3/2}.\sin\theta\exp\big\{\kappa\cos\theta\big\}\exp\Big\{-\frac{1}{2}\big[1+2\kappa^2\big]\Big\}} \\ &= \big(2\pi\big)^{3/2}\cdot\hat{f}_1(1)\exp\Big\{\frac{1}{2}\Big(1+2\kappa^2\Big)\Big\} \\ &= \big(2\pi\big)\Big(\mathbf{K}_S''(\hat{u}(1))\Big)^{-1/2}\exp\Big\{\mathbf{K}_S(\hat{u}(1)) - \hat{u}(1) + \frac{1}{2}\big(1+2\kappa^2\big)\Big\} = \hat{c}_2(\kappa), \text{say.} \end{aligned} \tag{2.101}$$

Here the saddlepoint function is given by

$$\hat{u}(1) = \hat{u}(1,\kappa) = -\frac{1}{4}\Big\{1 + \sqrt{(9+4\kappa^2)}\Big\}, \tag{2.102}$$

and the cumulant generating function $\mathbf{K}_S(\hat{u}(1))$ and the second derivative of $\mathbf{K}_S''(\hat{u}(1))$ about the saddlepoint $\hat{u}(1)$ are given by

$$\mathbf{K}_S\big(\hat{u}(1)\big) = \frac{\kappa^2\hat{u}(1)}{1-2\hat{u}(1)} - \frac{3}{2}\log\big(1-2\hat{u}(1)\big), \tag{2.103}$$

and

$$\mathbf{K}_S''\big(\hat{u}(1)\big) = \frac{4\kappa^2 + 6 - 12\hat{u}(1)}{(1 - 2\hat{u}(1))^3}. \tag{2.104}$$

The second order saddlepoint approximation for the normalizing constant of the Fisher distribution $\hat{c}_3(\kappa)$, say, can be obtained as

$$\hat{c}_3(\kappa) = \hat{c}_2(\kappa)(1 + T),$$

where T is the second order correction term T in (2.20).

An approximate maximum likelihood estimate for the concentration parameter κ of the Fisher distribution can be obtained using the second order saddlepoint approximation (Kume and Wood [54]). The sample size indicator can also be used to show the accuracy of the second order saddlepoint approximations. Table 2.3 shows the accuracy of the first order and second order saddlepoint approximations for the normalizing constant of the Fisher distribution with various values of the concentration parameter κ where $c_{02}(\kappa) = \{c(\kappa)\}^{-1}$, $\hat{c}_{02}(\kappa) = \{\hat{c}_2(\kappa)\}^{-1}$ and $\hat{c}_{03}(\kappa) = \{\hat{c}_3(\kappa)\}^{-1}$. Note that the relative error of each saddlepoint approximation stays bounded and converges to a given limit. As κ tends to infinity, the ratio between the first order saddlepoint approximation $\hat{c}_{02}(\kappa)$ and the second order saddlepoint approximations $\hat{c}_{03}(\kappa)$, and the true values of c_{02} tends to unity. Numerical results show that the first order saddlepoint approximation is accurate, but that the second order saddlepoint approximation do even better. The final column in Table 2.3 gives the sample size required for the difference between the true κ and $\hat{\kappa}_3$. The table shows that $\hat{\kappa}_3$ provides highly accurate approximation in all cases considered.

2.9 Bingham Distribution on the Sphere

2.9.1 Background

Consider the situation when the observations are not directions but axes; that is, the unit vectors $\mathbf{x}$ and $-\mathbf{x}$ define the same axis. An important distribution for dealing with axial data is the Bingham distribution where the probability density function satisfies the antipodal symmetry property

$$f(\mathbf{x}) = f(-\mathbf{x}). \tag{2.105}$$

The real Bingham distribution can be obtained by conditioning p-variate normal distribution on $\|\mathbf{x}\| = \mathbf{x}^T\mathbf{x} = 1$. Let $\mathbf{x}$ have an $N_p\big(0, (2\mathbf{A})^{-1}\big)$ distribution. Then the conditional distribution of $\mathbf{x}$

κ	Exact c_{02}	SPA $\hat{c}_{02}$	$\hat{c}_{02}/c_{02}$	SPA $\hat{c}_{03}$	$\hat{c}_{03}/c_{03}$	$\hat{\kappa}_3$	n
1	0.0677139100	0.0626807400	0.92567	0.0638791020	0.94337	0.994	1.1×10^5
3	0.0238306400	0.0223764900	0.93898	0.0224978211	0.94401	2.897	1.2×10^6
6	0.0023670490	0.0022265880	0.94066	0.0022489743	0.95012	5.869	1.8×10^4
9	0.0001767715	0.0001685216	0.95333	0.0001698674	0.96094	8.124	1.3×10^7
12	1.173458e-05	1.120911e-05	0.95522	1.130288e-05	0.96321	10.52	1.4×10^5
15	7.302880e-07	7.010838e-07	0.96001	7.056788e-07	0.96630	14.95	1.2×10^7
18	4.363068e-08	4.193825e-08	0.96121	4.246871e-08	0.97337	18.01	1.5×10^4
25	2.114156e-11	2.043903e-11	0.96677	2.078456e-11	0.98311	15.02	1.7×10^6
50	3.512217e-15	3.442710e-15	0.98021	3.485217e-15	0.99231	50.01	2.3×10^8

Table 2.3: Numerical first order saddlepoint approximations $\hat{c}_{02}$ and second order saddlepoint approximations $\hat{c}_{03}$ for the normalizing constant of the Fisher distribution under various values of κ; n is the sample size required for the difference between the true κ and $\hat{\kappa}_3$ to be significant at the level 5% level when large sample likelihood ratio test is used (see Kume and Wood [54]).

given that $\mathbf{x}^T\mathbf{x} = 1$ is Bingham with a symmetric parameter matrix $\mathbf{A}$. The real Bingham distribution Bing($\mathbf{A}$) has the probability density function with respect to $w_p(d\mathbf{x})$, the surface measure on unit sphere

$$f(\mathbf{x};\mathbf{A}) = \big(c_{\text{Bing}}(\mathbf{A})\big)^{-1}\exp\Big(-\mathbf{x}^T\mathbf{A}\mathbf{x}\Big), \tag{2.106}$$

where the normalizing constant $c_{\text{Bing}}(\mathbf{A}) = {}_1F_1\Big(\frac{1}{2};\frac{p}{2};\mathbf{A}\Big)$ is the hypergeometric function of matrix argument (Mardia and Jupp [59], p. 289). Further, $\mathbf{A}$ is identifiable only up to the addition of a multiple of the identity matrix. That is, because of the constraint $\mathbf{x}^T\mathbf{x} = 1$, the matrix $\mathbf{A}$ and $\mathbf{A} + \alpha I_p$ for any real number α define the same real Bingham distribution with $c_{\text{Bing}}(\mathbf{A} + \alpha I_p) = \exp(\alpha)c_{\text{Bing}}(\mathbf{A})$ and for a rotation matrix Γ, $c_{\text{Bing}}(\Gamma\mathbf{A}\Gamma^T) = c_{\text{Bing}}(\mathbf{A})$. Moreover, since $\mathbf{A}$ is a positive definite and symmetric, it can be factored by spectral decomposition as $\mathbf{A} = \Gamma\Lambda\Gamma^T = \sum_{i=1}^p \lambda_i\gamma_{(i)}\gamma_{(i)}^T$, where Λ is a diagonal matrix of $\lambda_i > 0, i = 1, 2, \ldots, p$, and Γ is an orthogonal matrix ($\Gamma^T\Gamma = \Gamma\Gamma^T = I$) whose columns are standardized eigenvectors, $\gamma_{(i)}, i = 1, 2, \ldots, p$. It follows that $\Lambda = \Gamma^T\mathbf{A}\Gamma$ is symmetric and its eigenvalues are all real. Therefore, there is a nonsingular orthogonal linear transformation of $\mathbf{x}$, $\mathbf{y} = \Gamma^T\mathbf{x}$, such that

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{y}^T\Gamma^T\mathbf{A}\Gamma\mathbf{y} = \mathbf{y}^T\Gamma^T\Gamma\Lambda\Gamma^T\Gamma\mathbf{y} = \mathbf{y}^T\Lambda\mathbf{y} = \sum_{i=1}^p \lambda_i\mathbf{y}_i^2. \tag{2.107}$$

Thus the quadratic form $Q(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$ has the same distribution as $Q(\mathbf{y}) = \mathbf{y}^T\Lambda\mathbf{y} = \sum \lambda_i\mathbf{y}_i^2$, where y_i are independent standard normal variables and $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p = 0$ denote the eigenvalues of $\mathbf{A}$ (Scheffé [92], Kuonen [55] and Mardia et al. [72], p. 474-475). So, without loss of generality, we assume that $\mathbf{A} = \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$ (Mardia et al. [72], p. 181 and Kume and Walker [52]). Further, if

$$f(\mathbf{x};\Lambda) = \frac{1}{c_{\text{Bing}}(\Lambda)}\exp\Big(-\sum_{i=1}^p \lambda_i\mathbf{y}_i^2\Big),$$

and

$$\begin{aligned} f(\mathbf{x};\Lambda + \alpha I_p) &= \frac{1}{c_{\text{Bing}}(\Lambda + \alpha I_p)}\exp\Big(-\sum_{i=1}^p (\lambda_i + \alpha)\mathbf{y}_i^2\Big) \\ &= \frac{\exp(\alpha)}{c_{\text{Bing}}(\Lambda + \alpha I_p)}\exp\Big(-\sum_{i=1}^p (\lambda_i)\mathbf{y}_i^2\Big), \end{aligned}$$

If $\mathbf{x}^T\mathbf{x} = 1$, then $f(\mathbf{x};\Lambda) = f(\mathbf{x};\Lambda + \alpha I_p)$ and $c_{\text{Bing}}(\Lambda)\exp(\alpha)/c_{\text{Bing}}(\Lambda + \alpha I_p)$ equal to unity. Hence $c_{\text{Bing}}(\Lambda + \alpha I_p) = \exp(\alpha)c_{\text{Bing}}(\Lambda)$.

2.9.2 Saddlepoint Approximations for the Normalizing Constant

It is clear that $x_i^2 \sim \lambda_i \chi_1^2, i = 1, 2, \ldots, p$. Thus under the simplest case $\mathbf{A} = I$, $\mathbf{x}^T\mathbf{x} = x_1^2 + x_2^2 + \ldots + x_p^2 = r^2$, say, follows a convolution of weighted central chi-squared χ_1^2 variables i.e., $r^2 = \sum_{i=1}^p x_i^2 \sim \sum_{i=1}^p \lambda_i \chi_1^2$. The moment generating function for r^2, $\mathbf{M}_{R^2}(u)$, is given by

$$\mathbf{M}_{R^2}(u) = \prod_{i=1}^{p} \Big(1 - \frac{u}{\lambda_i}\Big)^{-1/2}, \tag{2.108}$$

and the cumulant generating function for r^2, $\mathbf{K}_{R^2}(u)$, is given by

$$\mathbf{K}_{R^2}(u) = \log\Big[\prod_{i=1}^{p} \Big(1 - \frac{u}{\lambda_i}\Big)^{-1/2}\Big] = -\frac{1}{2}\sum_{i=1}^{p} \log\Big(1 - \frac{u}{\lambda_i}\Big), \qquad u < \lambda_1, \tag{2.109}$$

(Johnson and Kotz [33], p. 152). The first derivative of $\mathbf{K}_{R^2}(u)$ is

$$\mathbf{K}_{R^2}^{'}(u) = \frac{\displaystyle\sum_{i=1}^{p} \Big[(0.5)\Big(1 - \frac{u}{\lambda_i}\Big)^{-3/2}\Big(\frac{1}{\lambda_i}\Big) \prod_{j\neq i=1}^{p} \Big(1 - \frac{u}{\lambda_j}\Big)^{-1/2}\Big]}{\displaystyle\prod_{i=1}^{p} \Big(1 - \frac{u}{\lambda_i}\Big)^{-1/2}}. \tag{2.110}$$

Once the saddlepoint equation $\mathbf{K}_{R^2}^{'}(u) - r^2 = 0$ is solved numerically, the saddlepoint approximation for the probability density function of r^2, $\hat{f}_{R^2}\big(r^2; \mathbf{A}\big)$, can be obtained after computing $\mathbf{K}_{R^2}^{''}(u)$ and using (2.41).

Switch to polar coordinates of p dimensions

$$\mathbf{x} = r\mathbf{q}(\Theta), \qquad \Theta = (\theta_1, \theta_2, \ldots, \theta_{p-1})^T, \tag{2.111}$$

where

$$q_i(\Theta) = \cos\theta_i \prod_{j=1}^{i-1} \sin\theta_j, \qquad \sin\theta_0 = \cos\theta_p = 1, \tag{2.112}$$

and

$$0 \leq \theta_j \leq \pi, \qquad j = 1, 2, \ldots, p-2, \qquad 0 \leq \theta_{p-1} < 2\pi, \qquad r > 0. \tag{2.113}$$

The Jacobian of the transformation from (r, Θ) to $\mathbf{x}$ is given by

$$\mathbf{J} = r^{p-1} \prod_{i=2}^{p-1} \sin^{p-i}\theta_{i-1}, \tag{2.114}$$

(Mardia et al. [72], p. 35-36). The normalizing constant for the Bingham distribution $c_{\text{Bing}}(\mathbf{A})$ can be written as a function proportional to the probability density function of r^2, $f_{R^2}\big(r^2;\mathbf{A}\big)$, as follows. The true joint probability density function of p independent random variables $x_1, x_2, \ldots, x_p$ follow a multivariate normal distribution with mean vector $\mu = (0, 0, \ldots, 0)^T$, say, and covariance matrix $\Sigma = (2\mathbf{A})^{-1}$, can be written in terms of marginal probability density function of r^2 and conditional probability density function of Θ given r^2 (Θ/r^2) as

$$\begin{aligned}(2\pi)^{-p/2}|2\mathbf{A}|^{1/2}\exp\{-\mathbf{x}^T\mathbf{A}\mathbf{x}\}d\mathbf{x} &= f_{R^2}\big(r^2;\mathbf{A}\big)\big\{c_{\text{Bing}}\big(r^2\mathbf{A}\big)\big\}^{-1}\exp\Big\{-r^2\big(\mathbf{q}(\Theta)\big)^T\mathbf{A}\big(\mathbf{q}(\Theta)\big)\Big\}\\ &\times\quad r^{p-1}\prod_{i=2}^{p-1}\sin^{p-i}\theta_{i-1}\,d\,\Theta\,dr. \qquad (2.115)\end{aligned}$$

When $r^2 = 1$, any θ, the normalizing constant for the real Bingham distribution is given by

$$c_{\text{Bing}}\big(\mathbf{A}\big) = \beta\big(\mathbf{A}\big)f_{R^2}\big(1;\mathbf{A}\big) \qquad (2.116)$$

where

$$\beta\big(\mathbf{A}\big) = \frac{(2\pi)^{p/2}}{|2\mathbf{A}|^{1/2}}. \qquad (2.117)$$

The second order saddlepoint approximation for the normalizing constant of real Bingham distribution is then given by

$$\begin{aligned}\hat{c}_{\text{Bing}}\big(\mathbf{A}\big) &= \beta\big(\mathbf{A}\big)\hat{f}_{R^2}\big(1;\mathbf{A}\big)\\ &= \big(2\pi^{2p-3}\big)^{1/2}\Big[\mathbf{K}^{(2)}(\hat{u}(1))\Big]^{-1/2}\Big\{\prod_{i=1}^{2p-2}(\lambda_i - \hat{u}(1))\Big\}^{-1/2}\exp(-\hat{u}(1)+T), \qquad (2.118)\end{aligned}$$

where $\hat{u}(1)$ is the saddlepoint function which is the unique solution in $(-\infty, \lambda_1)$ to the saddlepoint equation $\mathbf{K}^{(1)}(u(s)) = 1$ and

$$T = \frac{1}{8}\hat{\rho}_4 - \frac{5}{24}\hat{\rho}_3^2, \qquad \hat{\rho}_j = \frac{\mathbf{K}^{(j)}\big(\hat{u}(s)\big)}{\Big[\mathbf{K}^{(2)}\big(\hat{u}(s)\big)\Big]^{j/2}}. \qquad (2.119)$$

In the case $p = 2$ it is known that the real Bingham distribution is a 2-wrapped von Mises by doubling angles (Mardia and Jupp [70], p. 54 and p. 182). So, the symmetric parameter matrix $\mathbf{A}$ can be chosen as $\mathbf{A} = \text{diag}(\lambda, 0)$ and the normalizing constant of the von Mises distribution is a double angled version of a particular real Bingham distribution, that is

$$c_{\text{Bing}}\big(\kappa\big) = [2\pi\, I_0(\kappa)]^{-1}, \qquad (2.120)$$

where $\kappa = \lambda/2$. Hence, the second order saddlepoint approximation for the normalizing constant of the real Bingham distribution can be obtained by recalling the saddlepoint approximation for the modified Bessel function of the first kind and order zero $I_0(\kappa)$ in (2.74), that is

$$\begin{aligned} \hat{c}_{2\text{Bing}}(\kappa) &= \hat{c}_{\text{Bing}}(\kappa)(1+T) \\ &= [2\pi\, \hat{I}_0(\kappa)]^{-1}(1+T), \end{aligned} \tag{2.121}$$

where T is the second order correction term for the saddlepoint approximation in (2.119). Table 2.4 gives the second order saddlepoint approximation for the normalizing constant of the real Bingham distribution with varying the concentration parameter $\kappa = \lambda/2$. It is clear from the table that as κ tends to infinity, the limiting behaviour of the second order saddlepoint approximation tends to unity. The numerical results for the second order saddlepoint approximations show that they are more accurate than the saddlepoint approximation obtained in Table 2.1 for the normalizing constant of the von Mises distribution.

λ	κ	Exact $c_{\text{Bing}}(\kappa)$	SPA $\hat{c}_{2\text{Bing}}(\kappa)$	$\hat{c}_{2\text{Bing}}(\kappa)/c_{\text{Bing}}(\kappa)$
0.0	0.0	0.15916	0.14788	0.92912
0.4	0.2	0.15906	0.14805	0.93078
0.8	0.4	0.15876	0.14837	0.93456
1.2	0.6	0.15670	0.14835	0.94671
1.6	0.8	0.13892	0.13159	0.94723
2.0	1.0	0.12571	0.11931	0.94911
10	5.0	0.00584	0.00556	0.95111
20	10	5.652380e-05	5.470826e-05	0.96788
200	100	1.482232e-43	1.467513e-43	0.99007

Table 2.4: Numerical results for the second order saddlepoint approximations for the normalizing constant of the real Bingham distribution with varying κ.

Chapter 3

Saddlepoint Approximations for the Complex Bingham Quartic Distribution

3.1 Introduction

The complex version of the Bingham distribution is defined on the unit complex sphere in $\mathbb{C}^{k-1}$ and it is relevant for the statistical shape analysis of landmark data in two dimensions i.e. it is a suitable distribution for modeling shapes. Under high concentrations the complex Bingham has a complex normal distribution. By adding a quartic term to the complex Bingham density gives the complex Bingham quartic (CBQ) distribution, which allows a full normal distribution under high concentrations. The normalizing constant of the complex Bingham quartic (CBQ) distribution has no closed form and therefore we provide an approximation procedure based on saddlepoint approximations for finite mixtures of the normalizing constants of Bingham distribution. Two new methods are explored to evaluate this normalizing constant namely, the Integrated Saddlepoint (ISP) approximation and the Saddlepoint-Integration (SPI) approximation. Calculating the normalizing constant for the CBQ distribution is based on numerical methods of quadrature.

3.2 Quadrature Methods

Numerical integration methods can be adaptive or non-adaptive. Non-adaptive methods apply the same rules over the entire range of integration. The integrand is evaluated at a finite set of points and a weighted sum of these function values is used to obtain the estimate. The numerical estimate

of $\int_a^b f(x)dx$ is of the form $\sum_{i=1}^n w_i f(x_i)$, where x_i are points in the interval $[a, b]$ and usually called nodes (abscissas) and w_i are suitable weights.

We now describe some of the quadrature methods one of which will be used when evaluating the normalizing constant of the complex Bingham quartic (CBQ) distribution whereas others could be explored in future work. There are many methods of numerical integration such as Trapezoidal, adaptive Simpson quadrature, Gauss-Legendre quadrature and Gauss-Hermite quadrature. Trapezoidal rule is the simplest numerical integration method but it used more time as many function calls than the adaptive Simpson quadrature and it gives exact results if the function f is a constant or a linear function, otherwise there will be an error, corresponding to the extent that our trapezoidal approximation overshoots the actual graph of f (Jones, et al. [35], p. 189). When an integrand behaves well in one part of the range of integration, but not so well in another part, it helps to treat each part differently. Adaptive methods choose the subintervals based on the local behavior of the integrand (Rizzo [85], p. 331). The adaptive quadrature rule may be most efficient for non-smooth integrands or needle-shaped functions (Germund, D. and Björck [24], p. 561). The main benefit of Gaussian quadrature is very high-order accuracy with very few points (typically less than 10). However, the Gaussian quadrature methods quickly become cumbersome as the dimensions increases, especially in the complex Bingham quartic (CBQ) distribution.

3.2.1 Trapezoidal Rule

This approach divides the interval $[a, b]$ into n equal length subintervals length $h = (b - a)/n$, with endpoints $x_0, x_1, \ldots, x_n$, and uses the area of the trapezoid to estimate the integral over each subinterval. If we suppose equally spaced nodes then the i-th node is given by $x_i = a + ih$, that is

$$x_0 = a, \quad x_1 = a + h, \quad x_2 = a + 2h, \quad \ldots, \quad x_n = a + nh = b.$$

The estimate on (x_i, x_{i+1}) is $\big(f(x_i) + f(x_{i+1})\big)(h/2)$, and the numerical estimate of $\int_a^b f(x)dx$ is

$$\frac{h}{2}f(a) + h\sum_{i=1}^{n-1} f(x_i) + \frac{h}{2}f(b). \tag{3.1}$$

Note that we get a better approximation if we take more trapezoids. This method will be encountered when evaluating the normalizing constant of the complex Bingham quartic (CBQ) distribution.

3.2.2 Recursive Adaptive Simpson Quadrature

In adaptive quadrature, the approximation is also obtained by approximating the area under the function $f(x)$ over subintervals. The subinterval width h is not constant over the interval $[a, b]$, but instead adapts to the function. The key observation is that h only needs to be small where the integrand is steep. Adaptive quadrature automatically allows the interval width h to vary over the range of integration, using a recursive algorithm. The basic idea is to apply Simpson's rule using initial h and $h/2$. If the difference between the two estimates is less than some given tolerance ϵ, say, then we are done. If not we split the range of integration $[a, b]$ into two parts $[a, c]$ and $[c, b]$ and on each part we apply Simpson's rule using interval width $h/2$ and $h/4$ and a tolerance $\epsilon/2$. If the error on each subinterval is less than $\epsilon/2$, then the error of the combined estimates will be less than ϵ. By increasing h we improve the accuracy. If the desired tolerance is not met on a given subinterval then we split it further, but we only do this for the subintervals that do not achieve the desired tolerance (Jones et al, [35], p. 194).

3.2.3 Gauss-Legendre Quadrature

Gauss-Legendre quadrature approximates the integral

$$I = \int_{-1}^{1} f(x)dx. \tag{3.2}$$

The general N-point Gauss-Legendre rule is exact for polynomial functions of degree $\leq 2N - 1$, i.e. for any function of the form

$$f(x) = \sum_{i=0}^{2n-1} c_i x^i, \tag{3.3}$$

where c_i are integers. The integral in (3.2) is approximated by the summation

$$I \approx \sum_{i=1}^{n} w_i f(x_i), \tag{3.4}$$

where x_i is the i-th zero of the Legendre polynomial $P_l(x)$ (abscissas or nodes), see Abramowitz and Stegun [1], p. 775 and p. 188, defined as

$$P_l(x) = \frac{1}{2^n} \sum_{m=0}^{l/2} (-1)^m \binom{n}{m} \binom{2n-2m}{n} x^{n-2m}, \tag{3.5}$$

with weights

$$w_i = \frac{2}{(1 - x_i^2)[p_l'(x_i)]^2}. \tag{3.6}$$

Let a and b each be finite, $a < b$. By the change of variables

$$x = \frac{2}{b-a}(t-a) - 1, \tag{3.7}$$

so that

$$t = \frac{b-a}{2}(x+1) + a \quad \text{and} \quad dt = \frac{b-a}{2}dx. \tag{3.8}$$

The N-point Gauss-Legendre quadrature rule over the interval $[a, b]$ is given by

$$\begin{aligned} I = \int_a^b f(t)dt &= \frac{b-a}{2}\int_{-1}^{1} f\Big(\frac{b-a}{2}(x+1)+a\Big)dx \\ &\approx \frac{b-a}{2}\sum_{i=1}^{n} w_i f\Big(\frac{b-a}{2}(x_i+1)+a\Big). \end{aligned} \tag{3.9}$$

3.3 Saddlepoint Approximations for Finite Mixtures

3.3.1 Background on Finite Mixtures

A mixture distribution is a compounding of statistical distributions, which arises when sampling from inhomogeneous populations (or mixed populations) with a different probability density function in each component.

For a random variable X taking values in $\mathbb{R}$, finite mixture models decompose a probability density function $f(x)$ into the sum of probability density functions from l classes. A general density function $f(x)$ is considered semiparametric, since it may be decomposed into l components. Let $f_i(x)$ denote the probability density function for the $i-$th class. Then the finite mixture model with l components is

$$f_X(x) = \sum_{i=1}^{l} w_i f_i(x), \tag{3.10}$$

where

$$0 \le w_i \le 1, \quad i = 1, \ldots, l; \quad w_1 + w_2 + \ldots + w_l = 1. \tag{3.11}$$

The parameters $w_1, \ldots, w_l$ will be called the mixing weights or mixing proportions and $f_1(.), \ldots, f_l(.)$ the component densities of the mixture. Furthermore, the moment generating function of the mixture distribution associated with the random variable X is defined as

$$\mathbf{M}_X(u) = \sum_{i=1}^{l} w_i \mathbf{M}_i(u), \tag{3.12}$$

over values of u for which the integral converges. With real values of u the convergence is always assured at $u = 0$. In addition, we shall presume that $\mathbf{M}(u)$ converges over an open neighborhood of zero designated as (u_1, u_2) with $u_1 < 0$ and $u_2 > 0$, and that, furthermore, (u_1, u_2) is the smallest such neighborhood of convergence. This assumption is often taken as a requirement for the existence of the MGF ($\mathbf{M}_X(u) < \infty$) (Huzurbazer [30], Butler [9], p. 8 and Reid [82]).

3.3.2 Saddlepoint Approximations

A key element in the saddlepoint approximation is the cumulant generating functions $\mathbf{K}_i(u) = \log\big(\mathbf{M}_i(u)\big)$ for $i = 1, 2, \ldots, l$. The cumulant generating functions are convex ($0 < \mathbf{K}_i^{''}(u) < \infty$, $i = 1, 2, \ldots, l$) and the cumulant generating function $\mathbf{K}(u)$ of the mixture distribution associated with the random variable X is given by

$$\begin{aligned} \mathbf{K}(u) &= \log\Big[\sum_{i=1}^{k} w_i \mathbf{M}_i(u)\Big] \\ &= \log\Big[\sum_{i=1}^{k} w_i \exp\{\mathbf{K}_i(u)\}\Big]. \end{aligned} \tag{3.13}$$

The first derivative of $\mathbf{K}(u)$ in (3.13) is given by

$$\begin{aligned} \mathbf{K}^{'}(u) &= \frac{\sum_{i=1}^{k} w_i \mathbf{M}_i^{'}(u)}{\sum_{i=1}^{k} w_i \mathbf{M}_i(u)} \\ &= \frac{\sum_{i=1}^{k} w_i \mathbf{M}_i(u) \mathbf{K}_i^{'}(u)}{\sum_{i=1}^{k} w_i \mathbf{M}_i(u)}, \end{aligned} \tag{3.14}$$

and the second derivative is also computed as

$$\begin{aligned} \mathbf{K}^{''}(u) &= \frac{\Big[\sum_{i=1}^{k} w_i \mathbf{M}_i(u)\Big]\Big[\sum_{i=1}^{k} w_i \mathbf{M}_i^{''}(u)\Big] - \Big[\sum_{i=1}^{k} w_i \mathbf{M}_i^{'}(u)\Big]^2}{\Big[\sum_{i=1}^{k} w_i \mathbf{M}_i(u)\Big]^2} \\ &= \frac{\Big[\sum_{i=1}^{k} w_i \mathbf{M}_i(u)\Big]\Big[\sum_{i=1}^{k} w_i \mathbf{M}_i(u)\big(\mathbf{K}_i^{''}(u) + \mathbf{K}_i^{'2}(u)\big)\Big] - \Big[\sum_{i=1}^{k} w_i \mathbf{M}_i(u)\mathbf{K}_i^{'}(u)\Big]^2}{\Big[\sum_{i=1}^{k} w_i \mathbf{M}_i(u)\Big]^2}, \end{aligned} \tag{3.15}$$

where $\mathbf{M}_i^{'}(u) = \mathbf{M}_i(u)\mathbf{K}_i^{'}(u)$ and $\mathbf{M}_i^{''}(u) = \mathbf{M}_i(u)\big(\mathbf{K}_i^{''}(u) + \mathbf{K}_i^{'2}(u)\big)$ and whence $\mathbf{K}(u)$ is also convex since $\mathbf{K}^{''}(u) > 0$.

The probability density function $f_X(x)$ for the mixtures defined in (3.10) may then be approximated using the following second order saddlepoint approximation

$$\hat{f}_X(x) = \Big[2\pi \mathbf{K}^{''}(\hat{u})\Big]^{-1/2} \exp\big\{\mathbf{K}(\hat{u}) - \hat{u}x\big\}(1 + T), \tag{3.16}$$

where T is the second order correction term defined in (2.119) and the saddlepoint $\hat{u}$ are defined implicity through the saddlepoint equation $\mathbf{K}^{'}(u) = x$, which is solved numerically using a root finder and the second order saddlepoint approximation in (3.16) is subsequently computed.

3.3.3 Application to Gamma Mixture Distribution

We illustrate the foregoing saddlepoint approximation by applying to a gamma mixture distribution. Let

$$f(x;\alpha_1,\theta_1,\alpha_2,\theta_2) = wf_1(x;\alpha_1,\theta_1) + (1-w)f_2(x;\alpha_2,\theta_2), \tag{3.17}$$

be the density of a two component gamma mixture, where $0 \leq w \leq 1$ is the mixing proportion and for $i = 1, 2$,

$$f(x;\alpha_i,\theta_i) = \left[\theta_i^{\alpha_i}\Gamma(\alpha_i\right]^{-1} x^{(\alpha_i-1)}\exp(-x/\theta_i) \quad x > 0, \quad \alpha_i > 0 \quad \theta_i > 0, \tag{3.18}$$

is the probability density function of gamma random variable with shape parameter α_i and scale parameter θ_i (Everitt and Hand [20], p. 102 and Gharib [25]). The moment generating function $\mathbf{M}_i(u)$, $i = 1, 2$ is given by

$$\mathbf{M}_i(u) = (1-\theta_i u)^{-\alpha_i}, \quad u < 1/\theta_i. \tag{3.19}$$

Then with parameters $(\alpha_1 = 5, \theta_1 = 1)$, $(\alpha_2 = 10, \theta_2 = 2)$ and $w = 0.30$ computed for sample size $n = 1$, the moment generating function of the mixture distribution associated with the random variable X is given by

$$\begin{aligned} \mathbf{M}_X(u) &= w\mathbf{M}_1(u) + (1-w)\mathbf{M}_2(u) \\ &= (0.30)(1-u)^{-5} + (0.70)(1-2u)^{-10}. \end{aligned} \tag{3.20}$$

The cumulant generating function is also given by

$$\mathbf{K}(u) = \log\Big[(0.30)(1-u)^{-5} + (0.70)(1-2u)^{-10}\Big], \quad u < 1/2, \tag{3.21}$$

and Figure 3.1 plots the cumulant generating function $\mathbf{K}(u)$ in (3.21) versus u for the gamma mixture distribution under consideration. The values of the graph range from $-\infty$ as $u \downarrow -\infty$ to ∞ as $u \uparrow \frac{1}{2}$. The function $\mathbf{K}(u)$ is then always a strictly convex function when evaluated over $\big(-\infty, 1/\max(\theta_i)\big)$.

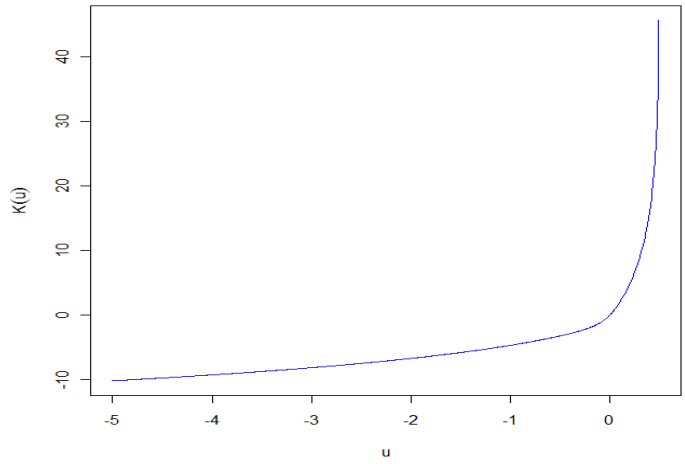

Figure 3.1: The cumulant generating function $\mathbf{K}(u)$ for the gamma mixture versus u.

The first and the second derivatives of $\mathbf{K}(u)$ can be calculated as

$$\mathbf{K}^{'}(u) = \frac{(1.50)(1-u)^{-6} + (14.00)(1-2u)^{-11}}{(0.30)(1-u)^{-5} + (0.70)(1-2u)^{-10}}, \tag{3.22}$$

and

$$\mathbf{K}^{''}(u) = \frac{(9.00)(1-u)^{-7} + (308.00)(1-2u)^{-12}}{(0.30)(1-u)^{-5} + (0.70)(1-2u)^{-10}} - \frac{\left[(1.50)(1-u)^{-6} + (14.00)(1-2u)^{-11}\right]^2}{\left[(0.30)(1-u)^{-5} + (0.70)(1-2u)^{-10}\right]^2}. \tag{3.23}$$

The saddlepoint equation $\mathbf{K}^{'}(u) - x = 0$ is then solved numerically by finding its root using any root-finding methods such as Brent's method (see for example, Brent [8] and Rizzo [85], p. 329) which is implemented in the R function `uniroot` or using Newton-Raphson method (see, for example, (Kolassa [51], p. 84 and Jones et al, [35], p. 174).

Figure 3.2 shows the curve of the exact gamma mixture density (solid line) as well as the second order saddlepoint approximation (dashed line). The exact and the second order saddlepoint approximation are indistinguishable in the tails. The second order saddlepoint density approximation does a good job of capturing the general shape in the center of the distribution and gives a sense of the bimodality of the density function.

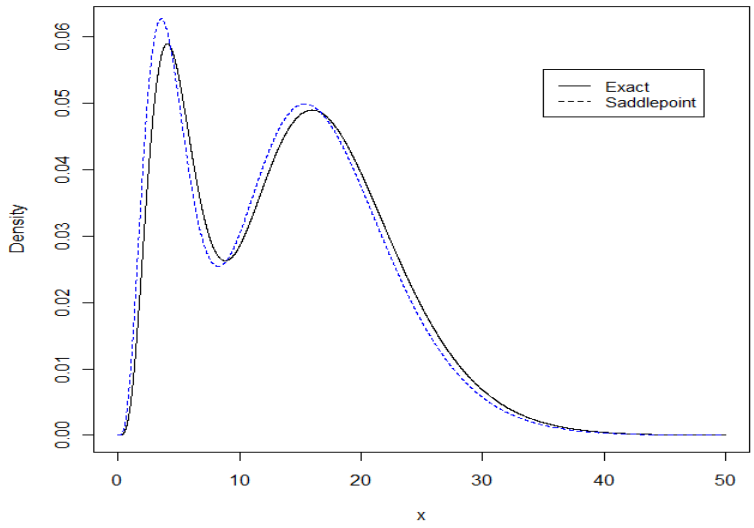

Figure 3.2: Gamma mixture random variable: Exact density and saddlepoint approximation.

3.4 Complex Bingham Quartic Distribution

The complex Bingham (CB) distribution described by Kent [41] is a suitable distribution for modelling shapes. Under high concentrations the complex Bingham has a complex normal distribution with isotropy ($\Sigma = \sigma^2 I_{2k-2}$). By adding a quartic term to the complex Bingham distribution we get the complex Bingham quartic (CBQ) distribution. The motivation behind the complex Bingham qrartic (CBQ) distribution was to develop a distribution centred at the complex Bingham (CB) distribution which includes anisotropy i.e. a full normal distribution under high concentrations. This is the same as the Fisher-Bingham (FB5) model of Kent [40]. This extended the Fisher distribution on S^2 to include anisotropy. The complex Bingham quartic (CBQ) distribution provides suitable shape parameters to include anisotropy.

We begin by looking at a background on the complex Bingham quartic distribution, some properties and motivation and a representation of the normalizing constant based on normalizing constants for the real Bingham distribution. The normalizing constant of the complex Bingham quartic (CBQ) distribution does not have a simple closed form representation and therefore we provide two new approximation procedures based on the saddlepoint approximations.

3.4.1 Background

Let $\mathbb{C}S^{k-2} \equiv S^{2k-3} = \{\mathbf{z} = (z_1, z_2, \ldots, z_{k-1})^T : \sum_{j=1}^{k-1} |z_j|^2 = 1\} \subset \mathbb{C}^{k-1}$ denote the unit complex sphere in $\mathbb{C}^{k-1}$. The complex Bingham quartic (CBQ) density on $\mathbb{C}S^{k-2}$, which is written $\mathrm{CBQ}(\mu, \Omega)$ depends on a $(k-1)$-dimensional unit complex location vector $\mu \in \mathbb{C}S^{k-2}$ and a $(2k-4)\times(2k-4)$-dimensional real symmetric reduced concentration matrix Ω. The density (with respect to the uniform measure on $\mathbb{C}S^{k-2}$) is a function of a unit complex vector $\mathbf{z} \in \mathbb{C}S^{k-2}$. Let I_{k-1} denote the $(k-1)$-dimensional identity matrix with columns $\mathbf{e}_1, \ldots, \mathbf{e}_{k-1}$. To simplify the presentation of distributions on $\mathbb{C}S^{k-2}$ (or $\mathbb{C}P^{k-2}$), it is often useful to rotate them to a standard form so that the modal direction (or axis) is given by $\mathbf{e}_{k-1}$. For analytical purposes, it is convenient to define a mapping $G : \mathbb{C}S^{k-2} \Rightarrow U(k-1)$ taking a unit complex vector $\mu \in \mathbb{C}S^{k-2}$ to a $(k-1)\times(k-1)$ unitary matrix $G(\mu)$. An example of the unitary function $G(\mu)$ is given in Kent et al. [46]. An important representation for a point $\mathbf{z}$ in $\mathbb{C}S^{k-2}$ is given by the partial Procrustes tangent projection of $\mathbf{z}$ with respect to $\mu \in \mathbb{C}S^{k-2}$.

First define the rotated version of $\mathbf{z}$,

$$G(\mu)^*\mathbf{z} = \mathbf{w} = (w_1, \ldots, w_{k-1})^T, \quad \text{say,} \tag{3.24}$$

If $\mu^*\mathbf{z} = \mathbf{e}_{k-1}^*\mathbf{w} = w_{k-1} \neq 0$, we define the $(k-2)$-dimensional complex partial Procrustes tangent vector $T_{\mathrm{PP}}(\mathbf{z};\mu) = \mathbf{v}$, say, to be the first $k-2$ coordinates of $\mathbf{w}$ after first aligning $\mathbf{w}$ with respect to $\mathbf{e}_{k-1}$, i.e. set

$$\begin{aligned} T_{\mathrm{PP}}(\mathbf{z};\mu) = \mathbf{v} &= \exp(-i\phi)(w_1, \ldots, w_{k-2})^T \\ &= \exp(-i\phi)G_{\backslash k-1}(\mu)^*\mathbf{z}, \end{aligned} \tag{3.25}$$

where $\phi = \arg(w_{k-1}) = \arg(\mu^*\mathbf{z})$ and $G_{\backslash k-1}(\mu)$ denote the $(k-1)\times(k-2)$ submatrix of $G(\mu)$ without its last column. Then $\mathbf{w}$ and $\mathbf{z}$ can be written in terms of $\mathbf{v}$ as

$$\begin{aligned} \mathbf{w} &= \exp(-i\phi)\left\{\cos(\rho)\mathbf{e}_{k-1} + \begin{pmatrix} \mathbf{v} \\ 0 \end{pmatrix}\right\} \\ \mathbf{z} &= \exp(-i\phi)\big\{\cos(\rho)\mu + G_{\backslash k-1}(\mu)\mathbf{v}\big\}, \end{aligned} \tag{3.26}$$

where

$$\cos(\rho) = |\mathbf{e}_{k-1}^*\mathbf{w}| = |\mu^*\mathbf{z}| = \sqrt{1 - \mathbf{v}^*\mathbf{v}} \in (0, 1], \tag{3.27}$$

with $\rho \in [0, \pi/2)$. Note that $(\mathbf{v}^T, 0)^T$ is orthogonal to $\mathbf{e}_{k-1}$ and $0 \leqslant \mathbf{v}^*\mathbf{v} < 1$. Also, $\mathbf{v}$ is unchanged if $\mathbf{z}$ is replaced by $\exp(i\psi)\mathbf{z}$ or μ is replaced by $\exp(i\psi)\mu$, so $\mathbf{v}$ represents a tangent projection on $\mathbb{C}P^{k-2}$ (Kent et al. [46]).

3.4.2 Decomposition of Quadratic Forms

Consider a $(2k-4) \times (2k-4)$-dimensional real symmetric matrix Ω. Such a matrix can be used (Kent et al. [46]) to construct a quadratic form involving the real and imaginary parts of a $(k-2)$-dimensional complex vector. If Ω is partitioned as

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}, \tag{3.28}$$

in terms of $(k-2)\times(k-2)$ submatrices, then a rotated version of Ω (corresponding to multiplication of the complex vector by i) can be defined by

$$\Omega^{(\mathrm{rot})} = \begin{pmatrix} \Omega_{22} & -\Omega_{21} \\ -\Omega_{12} & \Omega_{11} \end{pmatrix}. \tag{3.29}$$

The information in the parameter matrix Ω can be summarized in terms of two real matrices encoding the complex symmetric and anti-complex symmetric information,

$$\Omega^{(\mathrm{cs})} = \frac{1}{2}(\Omega + \Omega^{(rot)}) = \frac{1}{2}\begin{pmatrix} \Omega_{11}+\Omega_{22} & \Omega_{12}-\Omega_{21} \\ \Omega_{21}-\Omega_{12} & \Omega_{11}+\Omega_{22} \end{pmatrix}, \tag{3.30}$$

and

$$\Omega^{(\mathrm{as})} = \frac{1}{2}(\Omega - \Omega^{(rot)}) = \frac{1}{2}\begin{pmatrix} \Omega_{11}-\Omega_{22} & \Omega_{12}+\Omega_{21} \\ \Omega_{21}+\Omega_{12} & \Omega_{22}-\Omega_{11} \end{pmatrix}, \tag{3.31}$$

Note that $\Omega = \Omega^{(\mathrm{cs})} + \Omega^{(\mathrm{as})}$ and it is assumed that Ω is positive semidefinite ($\mathbf{d}^T\Omega\mathbf{d} \geq 0$ for all non-zero vectors $\mathbf{d}$ with real entries, $\mathbf{d} \in \mathbb{R}$), which in turn implies that $\Omega^{(\mathrm{cs})} = \frac{1}{2}(\Omega + \Omega^{(\mathrm{rot})})$ is also positive semidefinite (Kent et al. [46]).

The expression for the probability density function of the complex Bingham qrartic (CBQ) distribution is conveniently expressed in terms of the partial Procrustes tangent coordinates $\mathbf{v} = T_{\mathrm{PP}}(\mathbf{z}; \mu)$ by

$$f(\mathbf{z}) = c_{\mathrm{CBQ}}(\Omega)^{-1} \exp\big\{\mathbf{v}^*\mathbf{A}\mathbf{v} + \underbrace{(1-\mathbf{v}^*\mathbf{v})\mathrm{Re}(\mathbf{v}^T\mathbf{B}\mathbf{v})}_{\text{Quartic Term}}\big\}, \tag{3.32}$$

with respect to $d\big[\mathrm{Re}\big(\mathbf{v}\big)\big]d\big[\mathrm{Im}\big(\mathbf{v}\big)\big]$), the uniform measure on $\mathbb{C}S^{k-2}$. The information in the parameter matrix Ω can be coded by two $(k-2)\times(k-2)$ complex matrices $\mathbf{A}$ and $\mathbf{B}$,

$$\mathbf{A} = -\frac{1}{4}\big\{\Omega_{11}+\Omega_{22}+i(\Omega_{21}-\Omega_{12})\big\} \tag{3.33}$$

$$\mathbf{B} = -\frac{1}{4}\big\{\Omega_{11}-\Omega_{22}-i(\Omega_{21}+\Omega_{12})\big\}. \tag{3.34}$$

The probability density function for the CBQ distribution is also expressed in terms of $(2k-4)$-dimensional real coordinates (real tangent representation)

$$\mathbf{x} = (\mathrm{Re}(\mathbf{v})^T, \mathrm{Im}(\mathbf{v})^T)^T, \tag{3.35}$$

by

$$\begin{aligned} f(\mathbf{z}) &= c_{\mathrm{CBQ}}(\Omega)^{-1}\exp\Big\{-\frac{1}{2}\big(\mathbf{x}^T\Omega^{(cs)}\mathbf{x}+(1-\mathbf{x}^T\mathbf{x})\mathbf{x}^T\Omega^{(as)}\mathbf{x}\big)\Big\} \\ &= c_{\mathrm{CBQ}}(\Omega)^{-1}\exp\Big\{-\frac{1}{2}\big(\mathbf{x}^T\Omega\mathbf{x}-(\mathbf{x}^T\mathbf{x})\mathbf{x}^T\Omega^{(as)}\mathbf{x}\big)\Big\}. \end{aligned} \tag{3.36}$$

3.4.3 Some Properties and Motivation

(a) Simplification: If $\mathbf{B}=0$ then the complex Bingham quartic (CBQ) distribution reduces to the complex Bingham (CB) distribution with reduced concentration matrix $\mathbf{A}$.

(b) Role of the quartic parameters: In the real tangent representation (3.36), it might be thought that the essential properties of the density could be captured by the quadratic form $\mathbf{x}^T\Omega\mathbf{x}$ and the quartic terms could be dropped. In a sense this is true under high concentration because the quartic terms become negligible. However, in general the quartic terms cannot be ignored because the PDF would then become discontinuous as $|\mathbf{v}|=|\mathbf{x}|\to 1$, i.e. as $\mathbf{z}^*\mu\to 0$, and the model would be less appealing. The reason for the discontinuity is that, when $|\mathbf{v}|=1$, the tangent vectors $\mathbf{v}$ and $\exp(i\psi)\mathbf{v}$ define the same point on $\mathbb{C}P^{k-2}$, and the quadratic form $\mathrm{Re}(\mathbf{v}^T\mathbf{B}\mathbf{v}) = \mathbf{x}^T\Omega^{(\mathrm{as})}\mathbf{x}$ changes if $\mathbf{v}$ is replaced by $\exp(i\psi)\mathbf{v}$. In contrast the quadratic form $\mathbf{v}^*\mathbf{A}\mathbf{v} = \mathbf{x}^T\Omega^{(\mathrm{cs})}\mathbf{x}$ is unchanged if $\mathbf{v}$ is replaced by $\exp(i\psi)\mathbf{v}$ (Kent et al. [46]).

(c) Special case: For the triangle case, $k=3$, the complex Bingham quartic (CBQ) distribution reduces to the 5 parameter Fisher-Bingham (FB5) or Kent distribution on the sphere.

(d) Complex symmetry: Since replacing $\mathbf{z}$ by $\exp(i\theta)\mathbf{z}$ does not change the Procrustes tangent vector $\mathbf{v}$ in equation (3.25), the CBQ density clearly has complex symmetry.

Other properties and motivation are given in Kent et al. [46] and McDonnell [77].

3.4.4 Representation of the Normalizing Constant

If $\mathbf{x} = s^{1/2}\mathbf{w}$, where $\mathbf{w}$ is a real unit $(2k-4)$-dimensional vector and $0 < s < 1$, the normalizing constant for the complex Bingham quartic (CBQ) distribution can be derived as

$$\begin{aligned} c_{\mathrm{CBQ}}(\Omega) &= \int_{|\mathbf{x}|<1} \exp\{-\frac{1}{2}(\mathbf{x}^T\Omega\mathbf{x} - \mathbf{x}^T\mathbf{x}\mathbf{x}^T\Omega^{(\mathrm{as})}\mathbf{x})\}d\mathbf{x} \\ &= \pi\int_0^1 \Big[\int_{S^{2k-5}} \exp\Big\{-\frac{1}{2}\big(\mathbf{w}^T s\Omega\mathbf{w} - \mathbf{w}^T s^2\mathbf{w}\mathbf{w}^T\Omega^{(\mathrm{as})}\mathbf{w}\big)\Big\}d\mathbf{w}\Big] s^{k-3}ds \\ &= \pi\int_0^1 \Big[\int_{S^{2k-5}} \exp\big\{\mathbf{w}^T\Psi(s)\mathbf{w}\big\}d\mathbf{w}\Big] s^{k-3}ds \\ &= \pi\int_0^1 c_{\mathrm{Bing}}\big(\Psi(s)\big)s^{k-3}ds, \end{aligned} \tag{3.37}$$

where

$$\Psi(s) = -\frac{1}{2}(s\Omega - s^2\Omega^{(as)}), \qquad 0 < s < 1, \tag{3.38}$$

is also a real symmetric $(2k-4)\times(2k-4)$ matrix and $c_{\mathrm{Bing}}(\cdot)$ is the normalizing constant of the real Bingham distribution on S^{2k-5}. Thus the normalizing constant has been reduced to a one dimensional integral of normalizing constants for the Bingham distribution (Kent et al. [46]).

In practice we use the quadrature rules to approximate the definite integral in (3.37) by a summation

$$c_{\mathrm{CBQ}}(\Omega) \approx \pi\sum_{i=1}^{n} w_i c_{\mathrm{Bing}}\big(\Psi(s_i)\big)s_i^{k-3}, \tag{3.39}$$

where $s_1, s_2, \ldots, s_n$ are distinct nodes and $w_1, w_2, \ldots, w_n$ are the corresponding weights (Germund, D. and Björck [24], p. 561 and Abramowitz and Stegun [1], p. 886-887). For the moment we have used equally-spaced nodes (abscissas) with constant weights. Other choices will be explored in the future.

It is clear from the expression (2.116) that the normalizing constant for the Bingham distribution $c_{\mathrm{Bing}}(\mathbf{A})$ is proportional to the probability density function of a convolution central scaled chi-squared χ_1^2 variates, $f(r^2;\mathbf{A})$ at $r^2 = v = 1$ with cumulant generating function given in (2.109). More specifically, when $\mathbf{A} = \Psi(s)$, the normalizing constant for the real Bingham distribution is given by

$$c_{\mathrm{Bing}}\big(\Psi(s)\big) = \beta\big(\Psi(s)\big)f_{R^2}\big(1;\Psi(s)\big), \tag{3.40}$$

where

$$\beta\big(\Psi(s)\big) = \frac{(2\pi)^{(k-1)/2}}{|2\Psi(s)|^{1/2}}. \tag{3.41}$$

Thus the normalizing constant for the CBQ distribution in (3.37) becomes

$$\begin{aligned} c_{\text{CBQ}}(\Omega) &= \pi \int_0^1 \beta\big(\Psi(s)\big) f_{R^2}\big(1; \Psi(s)\big) s^{k-3} ds \\ &\approx \pi \sum_{i=1}^{n} w_i \beta\big(\Psi(s_i)\big) f_{R^2}\big(1; \Psi(s_i)\big) s_i^{k-3}. \end{aligned} \tag{3.42}$$

Notice that the normalizing constant for the CBQ distribution in (3.42) has been written as a weighed mixture of probability density functions for r^2, $f_{R^2}\big(1; \Psi(s)\big)$.

Two new methods are explored to evaluate the normalizing constant of the CBQ distribution in (3.42) based on the saddlepoint approximation of Bingham densities namely, the Integrated Saddlepoint (ISP) approximation (saddlepoint approximation first and then numerical integration) and the Saddlepoint-Integration (SPI) approximation (numerical integration first and then saddlepoint approximation). Even though both methods are based on second order saddlepoint approximations, nevertheless in practice the Integrated Saddlepoint (ISP) approximation is more accurate than the Saddlepoint-Integration (SPI) approximation but the latter could be used for all concentration parameters of the complex Bingham quartic distribution with a slight reduction in computer time.

3.4.5 Saddlepoint Approximations for the Normalizing Constant: Integrated Saddlepoint Approximations Approach

Kent et al. [46] proposed an integrated saddlepoint (ISP) approximation for the normalizing constant of the complex Bingham quartic (CBQ) distribution in which c_{Bing} in (3.37) is approximated by its saddlepoint approximation i.e. we firstly evaluate the saddlepoint approximation of c_{Bing} and then integrate the resulting function numerically. That is,

$$\begin{aligned} \hat{c}_{\text{CBQ, ISP}}(\Omega) &= \pi \int_0^1 \hat{c}_{\text{Bing}}\big(\Psi(s)\big) s^{k-3} ds \\ &\approx \pi \sum_{i=1}^{n} w_i \hat{c}_{\text{Bing}}\big(\Psi(s_i)\big) s_i^{k-3} \\ &= \pi \sum_{i=1}^{n} w_i \beta\big(\Psi(s_i)\big) \hat{f}_{R^2}\big(1; \Psi(s_i)\big) s_i^{k-3}, \end{aligned} \tag{3.43}$$

where $\hat{c}_{\text{Bing}}\big(\Psi(s)\big)$ is second-order saddlepoint approximation for the normalizing constant of the Bingham distribution and it takes the following form for a concentration matrix $\Psi(s)$:

$$\hat{c}_{\text{Bing}}\big(\Psi(s)\big) = \big(2\pi^{2k-5}\big)^{1/2}\Big[\mathbf{K}^{(2)}(\hat{u}(s))\Big]^{-1/2}\bigg\{\prod_{i=1}^{2k-4}(\lambda_i - \hat{u}(s))\bigg\}^{-1/2}\exp(-\hat{u}(s)+T), \tag{3.44}$$

where $\hat{u}(s)$ is the unique solution in $(-\infty, \lambda_1)$ to the saddlepoint equation $\mathbf{K}^{(1)}(u(s)) = 1$ and the correction term T is given in (2.119). The λ_i are the eigenvalues of $-\Psi$. The cumulant generating function $\mathbf{K}$ is given by

$$\mathbf{K}(u;\Psi) = -\frac{1}{2}\sum_{i=1}^{k-1}\log\Big(1-\frac{u}{\lambda_i}\Big),\ \ u<\lambda_1, \tag{3.45}$$

(Kent et al. [46] and Kume and Wood [54]). Moreover, the first and the higher derivatives for the cumulant generating function are given by

$$\mathbf{K}^{(j)}(u;\Psi) = \frac{(j-1)!}{2}\sum_{i=1}^{k-1}\frac{1}{(\lambda_i-u)^j},\quad j\geq 1. \tag{3.46}$$

When implementing this saddlepoint approximation, there is a simple one-dimensional search to find $\hat{u}$ for each s. This approximation works well over the whole range of choices for Ψ (Kent et al. [46]). Thus the $(2k-4)$-dimensional integral for normalizing constant of the complex Bingham quartic (CBQ) has been reduced to a simple one-dimensional integral. A numerical example will be given later.

It is also possible to evaluate the normalizing constant for the complex Bingham quartic (CBQ) distribution by using multivariate numerical integration such as Gauss-Legendre and Gauss-Hermite quadratures (McDonnell [77]). However, these multivariate numerical methods become cumbersome under high dimensions, and in general the saddlepoint approximation is both simpler and more reliable (Kent et al. [46]).

3.4.6 Saddlepoint Approximations for the Normalizing Constant: Saddlepoint of Integration Approximations Approach

An alternative to the integrated saddlepoint (ISP) approximation is to integrate first and then evaluate the saddlepoint approximation. We will call this the saddlepoint of integration (SPI) approximation. Let

$$g(1) = c_{\text{CBQ}}(\Omega) = \pi\sum_{i=1}^{n} w_i g_i\big(1;\Psi(s_i)\big), \tag{3.47}$$

where

$$g_i\big(1;\Psi(s_i)\big)=\beta\big(\Psi(s_i)\big)f_{R^2}\big(1;\Psi(s_i)\big)s_i^{k-3},\quad 0\le w_i\le 1,\quad \sum_{i=1}^{n}w_i=1,\ 0<s<1. \tag{3.48}$$

The second order saddlepoint approximation for the normalizing constant of the CBQ distribution is given by

$$\hat{c}_{\text{CBQ, SPI}}(\Omega)=\hat{g}(1)=\Big(2\pi\mathbf{K}_g^{''}(\hat{u}(s))\Big)^{-1/2}\exp\big\{\mathbf{K}_g(\hat{u}(s))-\hat{u}(s)\big\}\big(1+T\big), \tag{3.49}$$

where the second order correction term T is given in (2.119) and the saddlepoint $\hat{u}(s)$ is computed numerically by solving the saddlepoint equation $\mathbf{K}_g^{'}(u(s);\Psi)=1$ (see, for example, Kolassa [51], p. 84 and Jones et al, [35], p. 174). The moment generating function for $g(1)$ is given by

$$\mathbf{M}_g(u;\Psi)=\pi\sum_{i=1}^{n}w_i\beta\big(\Psi(s_i)\big)\mathbf{M}_i(u;\Psi(s_i))s_i^{k-3}, \tag{3.50}$$

and the cumulant generating function $\mathbf{K}_g(u)$ is computed as

$$\begin{aligned}\mathbf{K}_g(u;\Psi) &= \log\Big[\pi\sum_{i=1}^{n}w_i\beta\big(\Psi(s_i)\big)\mathbf{M}_i(u;\Psi(s_i))s_i^{k-3}\Big]\\ &= \log\Big[\pi\sum_{i=1}^{n}w_i\beta\big(\Psi(s_i)\big)\exp\big\{\mathbf{K}_i(u;\Psi(s_i))\big\}s_i^{k-3}\Big].\end{aligned} \tag{3.51}$$

3.4.7 Change of Variables

One notable drawback of numerical quadrature is the need to pre-compute (or look up) the requisite weights and nodes. Uniform nodes are not a suitable choice to compute the integrand function for the normalizing constant of the complex Bingham quartic (CBQ) distribution numerically, especially for $k>3$ and under high concentrations. Instead let us try uniform nodes with a change of variable. Notice that $r^2=\sum_{j=1}^{k-1}\lambda_j\chi_1^2=v$, say, is also approximated by $\bar{\lambda}\chi_{k-1}^2$ with probability density function given by

$$f(v)=\frac{1}{2^{(k-1)/2}\Gamma((k-1)/2)}v^{\frac{k-1}{2}-1}\exp\Big(-\frac{1}{2}v\Big), \tag{3.52}$$

and the cumulative distribution function of $\bar{\lambda}\chi_{k-1}^2$ is also given by

$$F(v)=\int_0^v\frac{w^{\frac{k-1}{2}-1}\exp(-\frac{1}{2}w)\,dw}{\Gamma((k-1)/2)2^{(k-1)/2}}=\frac{\gamma(\frac{k-1}{2},\frac{v}{2})}{\Gamma((k-1)/2)}=P\Big(\frac{k-1}{2},\frac{v}{2}\Big), \tag{3.53}$$

where $\Gamma(\cdot)$ is the gamma function, and

$$\gamma\Big(\frac{k-1}{2},\frac{v}{2}\Big)=\Big(\frac{v/2}{(k-1)/2}\Big)\Gamma\Big(\frac{k-1}{2}\Big)\exp\Big(\frac{v}{2}\Big)\sum_{j=1}^{\infty}\frac{(v/2)^j}{\Gamma\big[((k-1)/2)+j+1\big]}, \tag{3.54}$$

is the lower incomplete gamma function and $P(\cdot,\cdot)$ is the regularized Gamma function (Abramowitz and Stegun [1], p. 260).

Next the normalizing constant expression for the complex Bingham quartic (CBQ) distribution can be written as

$$\begin{aligned} c_{\mathrm{CBQ}}(\Omega) &= \pi \int_0^1 \frac{c_{\mathrm{Bing}}\big(\Psi(s)\big)s^{k-3}}{f(s)} f(s)ds \\ &= \pi \int_0^1 h(s)f(s)ds. \end{aligned} \tag{3.55}$$

Let $\tau = F(s)$ and $d\tau = f(s)\,ds$. We need to choose the probability density function $f(s)$ so that the function $h(s)$ will be nearly constant. A good test case is $\Omega^{(\mathrm{as})} = 0$ and $\Omega = c\,\mathbf{I}$, a multiple of the identity matrix with $c > 0$. The expression (3.55) can be reexpressed as

$$c_{\mathrm{CBQ}}(\Omega) = \pi \int_0^1 h\big(s(\tau)\big)d\tau. \tag{3.56}$$

In practice we use the quadrature rules to approximate the definite integral in (3.56) by a summation of the form

$$c_{\mathrm{CBQ}}(\Omega) \approx \pi \sum_{i=1}^{n} w_i h\big(\Psi(s_i)\big), \tag{3.57}$$

where

$$h\big(\Psi(s_i)\big) = \frac{c_{\mathrm{Bing}}\big(\Psi(s_i)\big)s_i^{k-3}}{f(s_i)}. \tag{3.58}$$

The normalizing constant expression for the real Bingham distribution (3.45) can be simplified as

$$\begin{aligned} c_{\mathrm{Bing}}\big(\Psi(s_i)\big) &= \beta\big(\Psi(s_i)\big) f_{R^2}\big(1;\Psi(s_i)\big) \\ &\approx \frac{(2\pi)^{(k-1)/2}}{|2\Psi(s_i)|^{1/2}} \frac{1}{2^{(k-1)/2}\Gamma(p/2)} \big(\bar{\lambda}(\Psi(s_i))\big)^{\frac{k-1}{2}-1} \exp\Big(-\frac{1}{2}\bar{\lambda}(\Psi(s_i))\Big). \end{aligned} \tag{3.59}$$

Let $\alpha = \mathrm{tr}(\Omega)$ and

$$\begin{aligned} \bar{\lambda}_i &= \frac{1}{k-1}\mathrm{trace}(\Psi(s_i)) \\ &= \frac{1}{k-1}\mathrm{trace}(-\frac{1}{2}(s_i\Omega - s_i^2\Omega^{(as)}) \\ &= -\frac{s_i}{2(k-1)}\mathrm{trace}(\Omega) = -\frac{\alpha}{2(k-1)}s_i, \end{aligned} \tag{3.60}$$

where $\mathrm{tr}(\Omega^{(as)}) = 0$. Moreover,

$$\begin{aligned}
|2\Psi(s_i)|^{1/2} &= \sqrt{2}|-\frac{1}{2}(s_i\Omega - s_i^2\Omega^{(as)})|^{1/2} \\
&\approx \sqrt{2}|-\frac{1}{2}(s_i\Omega)|^{1/2} \\
&= \sqrt{2}\Big(-\frac{s_i}{2}\Big)^{(k-1)/2}|\Omega|^{1/2} \\
&= \sqrt{2}\Big(-\frac{s_i}{2}\Big)^{(k-1)/2}\prod_{i=1}^{p}\lambda_i^{1/2} \\
&= \sqrt{2}\Big(-\frac{s_i}{2}\Big)^{(k-1)/2}\frac{1}{2}\mathrm{trace}(\Omega) = \sqrt{2}\Big(-\frac{s_i}{2}\Big)^{(k-1)/2}\frac{1}{2}\alpha, \qquad (3.61)
\end{aligned}$$

and the normalizing constant expression in (3.59) becomes

$$\begin{aligned}
c_{\mathrm{Bing}}\big(\Psi(s_i)\big) &= \frac{\pi^{(k-1)/2}}{\sqrt{2}\big(\frac{s_i}{2}\big)^{-(k-1)/2}\frac{1}{2}\bar{\alpha}\Gamma((k-1)/2)}\Big(\frac{\alpha}{2(k-1)}s_i\Big)^{\frac{(k-1)}{2}-1}\exp\Big(-\frac{1}{4(k-1)/\alpha}s_i\Big) \\
&= c_g s_i^{k-2}\exp\Big(-\frac{1}{\delta}s_i\Big), \qquad (3.62)
\end{aligned}$$

where $\delta = 4(k-1)/\alpha$ and the constant c_g is given by

$$c_g = \frac{\sqrt{2}\pi^{(k-1)/2}\big(\frac{\alpha}{2(k-1)}\big)^{\frac{(k-1)}{2}-1}}{\alpha\Gamma((k-1)/2)}. \qquad (3.63)$$

Notice that the final expression of $c_{\mathrm{Bing}}\big(\Psi(s_i)\big)$ in (3.62) is the same as the probability density function of Gamma distribution with scale and shape parameters $k-1$ and δ, respectively. Thus, the normalizing constant for the complex Bingham quartic (CBQ) distribution reduces to

$$\begin{aligned}
c_{\mathrm{CBQ}} &\approx \pi\sum_{i=1}^{n} w_i h\big(\Psi(s_i)\big) \\
&= \pi\sum_{i=1}^{n} w_i \frac{\sqrt{2}\pi^{(k-1)/2}\big(\frac{\alpha}{2k-2}\big)^{\frac{(k-1)}{2}-1}}{\alpha\Gamma((k-1)/2)}s_i^{2k-5}\exp\Big(-\frac{1}{4(k-1)/\alpha}s_i\Big). \qquad (3.64)
\end{aligned}$$

Note that the last expression for $h(s)$ in (3.64) is also the same as the probability density function of Gamma distribution with scale and shape parameters $2k-5$ and $\frac{1}{4(k-1)/\alpha}$, respectively. So, the integral of a Gamma density over the interval $(0,1)$ is less than 1. Then the probability density function needed is a truncated Gamma distribution $h^*\big(\Psi(s_i)\big)$, say, with density

$$h^*\big(\Psi(s_i)\big) = \frac{h\big(\Psi(s_i)\big)}{H(1)} = \frac{\frac{\sqrt{2}\pi^{(k-1)/2}\big(\frac{\alpha}{2k-2}\big)^{\frac{(k-1)}{2}-1}}{\alpha\Gamma((k-1)/2)}s^{(2k-4)-1}\exp\Big(-\frac{1}{4(k-1)/\alpha}s\Big)}{\int_0^1 \frac{\sqrt{2}\pi^{(k-1)/2}\big(\frac{\alpha}{2k-2}\big)^{\frac{(k-1)}{2}-1}}{\alpha\Gamma((k-1)/2)}w^{(2k-4)-1}\exp\Big(-\frac{1}{4(k-1)/\alpha}w\Big)\,dw}, \qquad (3.65)$$

and a distribution function $H^*\big(\Psi(s_i)\big) = H\big(\Psi(s_i)\big)/H(1)$. Again, for the purpose of numerical integration we have used equally weights $w_i = \frac{1}{n}$, $i = 1, 2, \ldots, n$ with the following suggested uniform nodes

$$s_i = H^{*-1}(\tau_i) = H^{*-1}\left(\frac{i - 0.5}{n}\right). \tag{3.66}$$

The cumulative distribution function $H^*(s)$ is strictly increasing. Then the equation

$$H^*(s) = \tau, \qquad 0 < \tau < 1, \tag{3.67}$$

has a unique solution, say $s = \xi_\tau$, and $H^{*-1}(\tau)$ is the unique quantile of order τ for the scaled/truncated Gamma distribution and it is increasing in $(0, 1)$ (see David and Nagaraja [15], p.159). Figure 3.3 plots the function $h(s)$ versus s with $k = 3$, $\lambda_1 = 100$, $\lambda_2 = 900$ and $\alpha = \mathrm{tr}(\Omega) = 1000$.

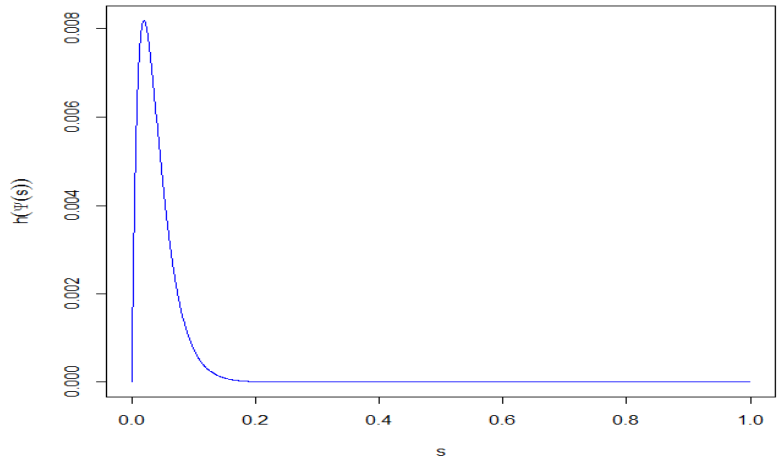

Figure 3.3: The function $h\big(\Psi(s)\big)$ versus s.

For change of variable the second order integrated saddlepoint (ISP) approximations for the normalizing constant of the complex Bingham quartic (CBQ) distribution $\hat{c}_{\mathrm{CBQ,\,CVISP}}(\Omega)$, say, can be evaluated by finding firstly the saddlepoint approximation for the truncated Gamma density in (3.65) and then evaluating the numerical integration for the saddlepoint approximation obtained.

Let

$$g(1) = c_{\mathrm{CBQ}} = \pi \sum_{i=1}^{n} w_i h_i^*\big(\Psi(s_i)\big), \tag{3.68}$$

where $h^*\big(\Psi(s_i)\big)$ is defined in (3.65). The moment generating function for $g(1)$ is given by

$$\mathbf{M}_g(u;\Psi) = \pi \sum_{i=1}^{n} w_i \mathbf{M}_i(u;\Psi(s_i)) s_i^{k-3}, \tag{3.69}$$

and the cumulant generating function $\mathbf{K}_g(u)$ is computed as

$$\begin{aligned} \mathbf{K}_g(u;\Psi) &= \log\Big[\pi \sum_{i=1}^{n} w_i \mathbf{M}_i(u;\Psi(s_i))\Big] \\ &= \log\Big[\pi \sum_{i=1}^{n} w_i \exp\big\{\mathbf{K}_i(u;\Psi(s_i))\big\}\Big]. \end{aligned} \tag{3.70}$$

The second order saddlepoint approximation of integration (SPI) approximation with change of variable for the normalizing constant of the complex Bingham quartic (CBQ) distribution $\hat{c}_{\text{CBQ, CVSPI}}(\Omega)$, say, is then given by

$$\hat{c}_{\text{CBQ,CVSPI}}(\Omega) = \hat{g}(1) = \Big(2\pi \mathbf{K}_g''(\hat{u}(s))\Big)^{-1/2} \exp\big\{\mathbf{K}_g(\hat{u}(s)) - \hat{u}(s)\big\}\big(1+T\big). \tag{3.71}$$

3.4.8 Performance Assessment for Mixture Saddlepoint Approximations Approaches

Let $\lambda_i > 0, i = 1, 2$ be the eigenvalues for the symmetry parameter matrix $\mathbf{A}$ of the real Bingham distribution satisfy $\lambda_1 \leq \lambda_2$. The moment generating function for $r^2 = \sum_{i=1}^{2} x_i^2$, $\mathbf{M}_{R^2}(u)$, is given by

$$\mathbf{M}_{R^2}(u) = \prod_{i=1}^{2} \Big(1 - \frac{u}{\lambda_i}\Big)^{-1/2}, \tag{3.72}$$

and the cumulant generating function for r^2, $\mathbf{K}_{R^2}(u)$, is given by

$$\mathbf{K}_{R^2}(u) = \log\Big[\Big(1 - \frac{u}{\lambda_1}\Big)^{-1/2}\Big(1 - \frac{u}{\lambda_2}\Big)^{-1/2}\Big], \qquad u < \lambda_1. \tag{3.73}$$

Figure 3.4 plots the cumulant generating function $\mathbf{K}_{R^2}(u)$ in (3.73) versus u with $\lambda_1 = 8$ and $\lambda_2 = 72$. The values of the graph range from $-\infty$ as $u \downarrow -\infty$ to ∞ as $u \uparrow 8$. and the function $\mathbf{K}(u)$ is then always a strictly convex function when evaluated over $(-\infty, 8)$. The corresponding u-values on the horizontal axis are the associated saddlepoints spanning $(-\infty, 8)$, the convergence neighbourhood of $\mathbf{K}_{R^2}(u)$.

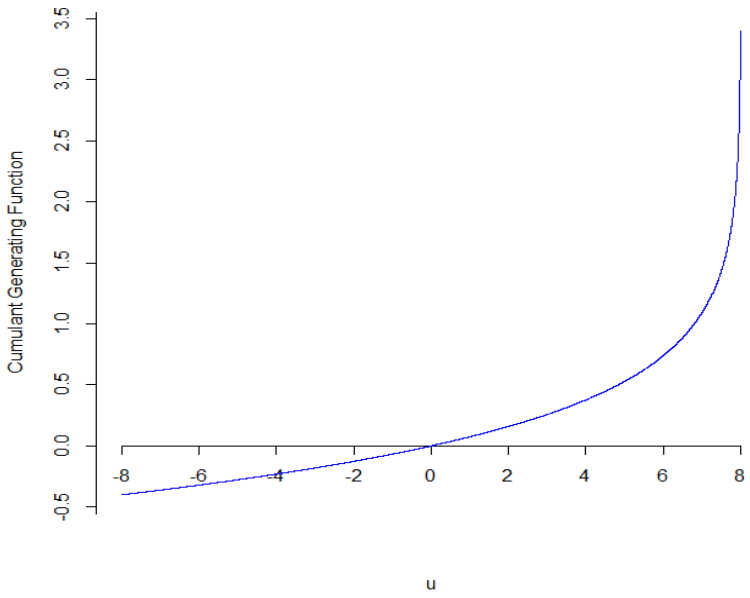

Figure 3.4: The cumulant generating function $\mathbf{K}_{R^2}(u)$ versus u.

The first and the second derivatives of $\mathbf{K}_{R^2}(u)$ are

$$\begin{aligned}\mathbf{K}'_{R^2}(u) &= \frac{\left[(0.5)\left(1-\frac{u}{\lambda_2}\right)^{-1/2}\left(1-\frac{u}{\lambda_1}\right)^{-3/2}\left(\frac{1}{\lambda_1}\right)+(0.5)\left(1-\frac{u}{\lambda_1}\right)^{-1/2}\left(1-\frac{u}{\lambda_2}\right)^{-3/2}\left(\frac{1}{\lambda_2}\right)\right]}{\left(1-\frac{u}{\lambda_1}\right)^{-1/2}\left(1-\frac{u}{\lambda_2}\right)^{-1/2}} \\ &= \frac{(0.5)(\lambda_1+\lambda_2-2u)}{(u-\lambda_1)(u-\lambda_2)}, \end{aligned} \tag{3.74}$$

and

$$\mathbf{K}''_{R^2}(u) = \frac{(0.5)\left(\lambda_1^2+\lambda_2^2-2u(\lambda_1+\lambda_2)+2u^2\right)}{(u-\lambda_1)^2(u-\lambda_2)^2}. \tag{3.75}$$

Once the saddlepoint equation $\mathbf{K}'_R(u) = r^2 = 1$ is solved numerically to compute the saddlepoint $\hat{u}(s)$, the saddlepoint approximation for the probability density function of r^2, $\hat{f}_{R^2}(r^2)$ can be computed as

$$\hat{f}_{R^2}\left(1;\Psi(s)\right) = \left(2\pi\mathbf{K}''_{R^2}(\hat{u}(s))\right)^{-1/2}\exp\left\{\mathbf{K}_{R^2}(\hat{u}(s))-\hat{u}(s)\right\}. \tag{3.76}$$

Switch to the two-dimensional polar coordinates (r,θ),

$$x_1 = r\cos\theta, \qquad \text{and} \qquad x_2 = r\sin\theta, \qquad r>0, \qquad 0\le\theta<2\pi. \tag{3.77}$$

The Jacobian of this transformation is given by

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_1}{\partial \theta} \\ \frac{\partial x_2}{\partial r} & \frac{\partial x_2}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r\cos^2\theta + r\sin^2\theta = r. \tag{3.78}$$

As we mentioned the normalizing constant for Bingham distribution $c_{\text{Bing}}(\mathbf{A})$ can be written as a proportional function to the probability density function of r^2, $f_{R^2}\big(r^2;\mathbf{A}\big)$. When $r^2 = 1$, any θ, the equation (2.114) becomes

$$(2\pi)^{-1}|2\mathbf{A}|^{1/2}\exp\big\{-\mathbf{x}^T\mathbf{A}\mathbf{x}\big\} = f_{R^2}\big(1;\mathbf{A}\big)\big\{c_{\text{Bing}}\big(\mathbf{A}\big)\big\}^{-1}\exp\left\{\begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix}^T\mathbf{A}\begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix}\right\}, \tag{3.79}$$

and when $\mathbf{A} = \Psi(s)$, the normalizing constant for the real Bingham distribution is given by

$$c_{\text{Bing}}\big(\Psi(s)\big) = \frac{2\pi}{|2\Psi(s)|^{1/2}} f_{R^2}\big(1;\Psi(s)\big), \tag{3.80}$$

and the second order saddlepoint approximation for the normalizing constant of the Bingham distribution is then given by

$$\hat{c}_{\text{Bing}}\big(\Psi(s)\big) = \frac{2\pi}{|2\Psi(s)|^{1/2}}\Big(2\pi\mathbf{K}''_{R^2}(\hat{u}(s))\Big)^{-1/2}\exp\big\{\mathbf{K}_{R^2}(\hat{u}(s)) - \hat{u}(s)\big\}(1+T), \tag{3.81}$$

where T is the second order correction term T in (2.119). Note that for each s, $0 < s < 1$, we have moment generating function $\mathbf{M}_{R^2}(u;s)$ and cumulant generating function $\mathbf{K}_{R^2}(u;s)$, say. For triangle case, $k = 3$, Figure 3.5 plots the second-order saddlepoint approximation for the normalizing constant of the Bingham distribution, $\hat{c}_{\text{Bing}}\big(\Psi(s)\big)$ versus s with $\lambda_1 = 8$ and $\lambda_2 = 72$ (low concentration) and $\lambda_1 = 400$ and $\lambda_2 = 3600$ (high concentration). Under low concentrations, the uniform nodes are suitable to evaluate the integrand in equation (3.37) numerically. On the other hand, further care is needed under high concentrations to suit the behaviour of the function $\hat{c}_{\text{Bing}}\big(\Psi(s)\big)s^{k-3}$ in (3.37). Notice that under high concentrations, $\hat{c}_{\text{Bing}}\big(\Psi(s)\big)s^{k-3}$ gets steeper/has a sharp peak and we need a smaller subinterval width h over the interval $[0,1]$ or using change of variable, to achieve acceptable accuracy.

In our case, $k = 3$, the $(2k-4)\times(2k-4)$-dimensional real symmetric matrix Ω reduces to a diagonal 2×2 matrix of the form,

$$\Omega = \begin{pmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{pmatrix}. \tag{3.82}$$

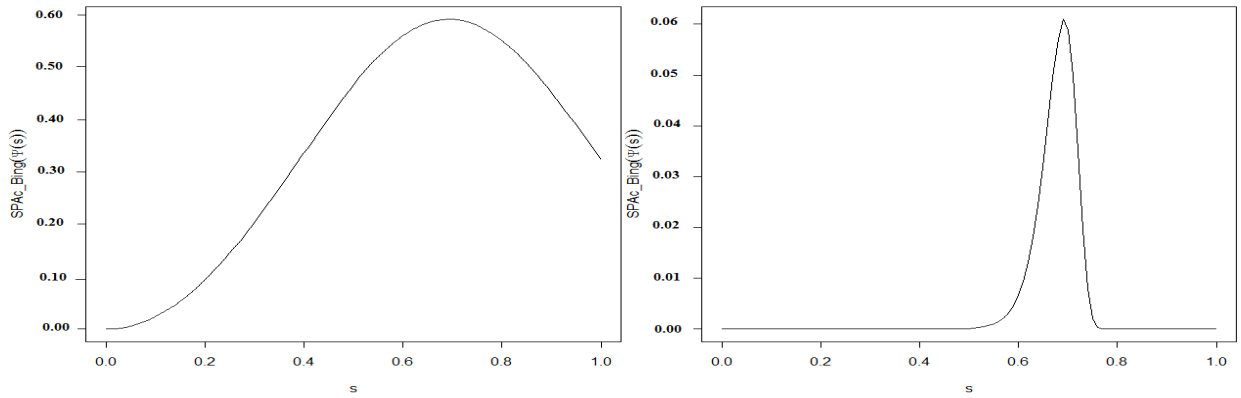

Figure 3.5: The saddlepoint approximation for the real Bingham distribution $\hat{c}_{\text{Bing}}\big(\Psi(s)\big)$ versus s.

Moreover, the anti-complex symmetric matrix $\Omega^{(\text{as})}$ also reduces to

$$\Omega^{(\text{as})} = \frac{1}{2}(\Omega - \Omega^{(rot)}) = \frac{1}{2}\begin{pmatrix} \lambda_1 - \lambda_2 & 0 \\ 0 & \lambda_2 - \lambda_1 \end{pmatrix}, \tag{3.83}$$

the complex symmetric matrix $\Omega^{(\text{cs})}$ reduces to

$$\Omega^{(\text{cs})} = \frac{1}{2}(\Omega + \Omega^{(rot)}) = \frac{1}{2}\begin{pmatrix} \lambda_1 + \lambda_2 & 0 \\ 0 & \lambda_1 + \lambda_2 \end{pmatrix}, \tag{3.84}$$

and $\Psi(s)$ becomes

$$\begin{aligned} \Psi(s) &= -\frac{1}{2}\left[s\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} - s^2\begin{pmatrix} \lambda_1 - \lambda_2 & 0 \\ 0 & \lambda_2 - \lambda_1 \end{pmatrix}\right] \\ &= -\frac{1}{2}\begin{pmatrix} \lambda_1 s\big(1-s\big) + \lambda_2 s^2 & 0 \\ 0 & \lambda_2 s\big(1-s\big) + \lambda_1 s^2 \end{pmatrix}, \qquad 0 < s < 1. \end{aligned} \tag{3.85}$$

Since there is no closed form for the exact values of the normalizing constant of the complex Bingham quartic distribution, instead we may use its relationship with that of the Kent (FB5) distribution

$$c_{\text{CBQ}}(\Omega) = \frac{\pi}{2}\exp(-\kappa)c_{\text{FB5}}(\kappa, \beta), \tag{3.86}$$

where $\kappa \geq 0$ represents the concentration parameter given by

$$\kappa = \frac{1}{8}(\lambda_1 + \lambda_2), \tag{3.87}$$

and $0 \leq \beta \leq \kappa/2$ (unimodality case) describes the ovalness of the distribution given by

$$\beta = \frac{1}{16}\big(\lambda_2 - \lambda_1\big). \tag{3.88}$$

Moreover, the normalizing constant of the Kent distribution $c_{\text{FB5}}(\kappa, \beta)$ is given by

$$\begin{aligned} c_{\text{FB5}}(\kappa, \beta) &= 2\pi \sum_{j=1}^{\infty} \frac{\Gamma(j+\frac{1}{2})}{\Gamma(j+1)} \beta^{2j} \big(\frac{1}{2}\kappa\big)^{-2j-\frac{1}{2}} I_{2j+\frac{1}{2}}(\kappa) \\ &\approx 2\pi \exp(\kappa) \Big[(\kappa - 2\beta)((\kappa + 2\beta)\Big]^{-1/2}, \qquad \text{(for large } \kappa\text{)}, \end{aligned} \tag{3.89}$$

(see, for example, Kent [40] and Kent et al. [46]). Evaluating the exact values for the normalizing constant of the Kent distribution $c_{FB5}(\kappa, \beta)$ in (3.76) is carried out using the R function `fb5.series` which accompanies Kent [40].

Table 3.1 shows a comparison between the true values of the normalizing constant of the complex Bingham quartic distribution, $c_{\text{CBQ}}(\Omega)$, the accuracy of the second order integrated saddlepoint (ISP) approximation $\hat{c}_{\text{CBQ, ISP}}(\Omega)$ and the second order saddlepoint approximation of integration (SPI) approximation $\hat{c}_{\text{CBQ, SPI}}(\Omega)$ with various values of κ , $\beta = 0.4\kappa$, λ_1, λ_2 and $n = 1000$. The table shows that both $\hat{c}_{\text{CBQ, ISP}}(\Omega)$ and $\hat{c}_{\text{CBQ, SPI}}(\Omega)$ approximations for the normalizing constant of the complex Bingham quartic distribution give less precise estimates when the concentration parameter κ tends to zero. The table also illustrates that as λ_1, λ_2 and the number of terms in the approximating integral on $(0, 1)$ increase, the ratios between both $\hat{c}_{\text{CBQ,ISP}}(\Omega)$ and $\hat{c}_{\text{CBQ,SPI}}(\Omega)$ with the true values of $c_{\text{CBQ}}(\Omega)$ approach 1. This point is further illustrated in Figure(3.6). R provides the `system.time` function, which times the evaluation of its argument. It can be used as a rough benchmark to compare the performance of the integrated saddlepoint (ISP) approximation and the saddlepoint approximation of integration (SPI) approximation. With increasing λ_1 and λ_2, the approximation $\hat{c}_{\text{CBQ,ISP}}(\Omega)$ seems to be more accurate than that of $\hat{c}_{\text{CBQ,SPI}}(\Omega)$, though the latter method is close to twice as fast under low and high concentrations.

For $k > 3$, a simple closed form representation for the true $c_{\text{CBQ}}(\Omega)$ is not available. Under higher concentrations, the integrated saddlepoint (ISP) approximation, the saddlepoint approximation of integration (SPI) and also asymptotic multivariate normal (AMN) approximations are valid to approximate $c_{\text{CBQ}}(\Omega)$. For $k = 4$, we have four concentration parameters λ_1, λ_2, λ_3 and λ_4 since the real symmetric matrix Ω has $(2k-4) \times (2k-4)$ dimensions. Table 3.2 shows also a comparison between the numerical results of the second order integrated saddlepoint (ISP) approximations $\hat{c}_{\text{CBQ, ISP}}(\Omega)$ and the second order saddlepoint approximation of integration (SPI)

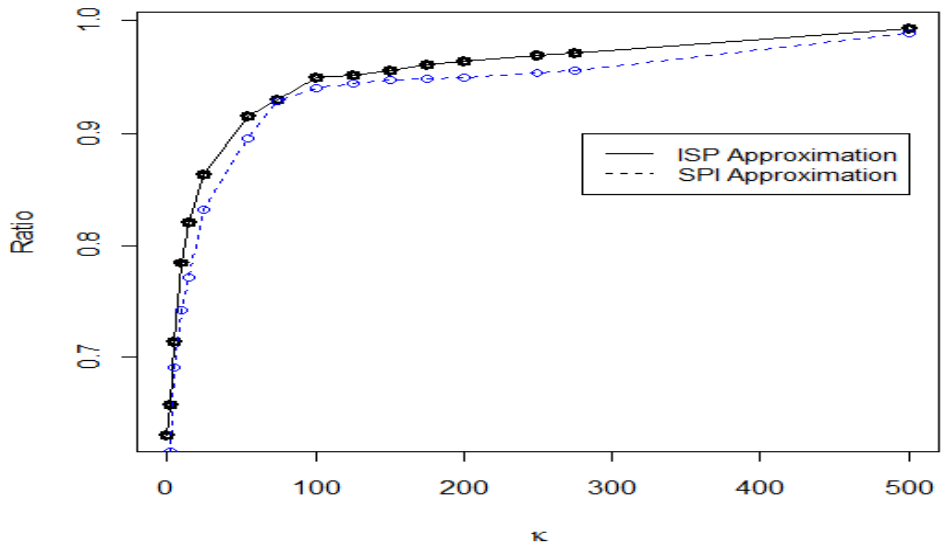

Figure 3.6: The ratio $\hat{c}_{\text{CBQ}}(\Omega)/c_{\text{CBQ}}(\Omega)$ against the concentration parameter κ.

approximation $\hat{c}_{\text{CBQ, SPI}}(\Omega)$ without change of variable and the second order integrated saddlepoint (ISP) approximation $\hat{c}_{\text{CBQ, CVISP}}(\Omega)$ and the second order saddlepoint approximation of integration (SPI) approximation $\hat{c}_{\text{CBQ, CVSPI}}(\Omega)$ subject to change of variable with λ_1, λ_2, λ_3 and λ_4 varying and with $n = 1000$. Both methods work well with/without change of variable. With increased values of the concentration parameters the difference between the numerical results of the second order saddlepoint approximation of integration (SPI) approximation and the second order integrated saddlepoint (ISP) approximation with/without change of variable tends to zero. However, the differences between the numerical results of $\hat{c}_{\text{CBQ, CVISP}}(\Omega)$ and $\hat{c}_{\text{CBQ, CVSPI}}(\Omega)$ are closer than the numerical outcomes of the second order saddlepoint approximations without change of variable.

For large κ the bulk of the probability mass is concentrated in the region $\mathbf{x} = O(\kappa^{-1/2})$, and the contribution to the logarithm of the probability density function from the quartic forms, $(\mathbf{x}^T\mathbf{x})\mathbf{x}^T\left(\kappa\Omega^{(as)}\right)\mathbf{x} = O(\kappa^{-1})$, converges to 0. Thus the distribution of $\mathbf{y} = \kappa\mathbf{x}$ converges to a multivariate normal distribution $N_{2k-4}(0, \Sigma)$, where $\Sigma = \Omega^{-1}$. Hence,

$$\hat{c}_{\text{CBQ, AMN}}(\kappa\Omega) \approx (2\pi)|2\pi\kappa^{-1}\Sigma|^{1/2} = (2\pi)^{2k-3}|\kappa^{-1}\Omega^{-1}|^{1/2}, \quad \text{as } \kappa \to \infty. \tag{3.90}$$

This asymptotic multivariate normal estimator makes it clear that the number of parameters $(k-1)$

λ_1	λ_2	κ	β	c_{CBQ} Exact	$\hat{c}_{\mathrm{CBQ}}$ ISP	$\hat{c}_{\mathrm{CBQ}}$ SPI	$\hat{c}_{\mathrm{CBQ}}/c_{\mathrm{CBQ}}$ ISP	$\hat{c}_{\mathrm{CBQ}}/c_{\mathrm{CBQ}}$ SPI
0.4	3.6	0.5	0.2	19.898	12.568	11.890	0.631	0.598
1.6	14.4	2	0.8	8.2246	5.4117	5.0678	0.658	0.616
4	36	5	2	3.2898	2.3496	2.2189	0.714	0.691
8	72	10	4	1.6449	1.2902	1.2208	0.784	0.742
12	108	15	6	1.0966	0.9010	0.8462	0.821	0.771
20	180	25	10	0.6579	0.5684	0.5467	0.863	0.832
60	540	75	30	0.2193	0.2040	0.2038	0.930	0.929
80	720	100	40	0.1644	0.1562	0.1546	0.950	0.940
100	900	125	50	0.1315	0.1252	0.1242	0.952	0.944
140	1260	175	70	0.0939	0.0903	0.0892	0.961	0.949
160	1440	200	80	0.0822	0.0793	0.0781	0.964	0.950
200	1800	250	100	0.0657	0.0637	0.0627	0.969	0.954
220	1980	275	110	0.0598	0.0581	0.0572	0.971	0.956
400	3600	500	200	0.0328	0.0326	0.0325	0.993	0.989

Table 3.1: Numerical results for true value of the normalizing constant of the complex Bingham quartic distribution $c_{\mathrm{CBQ}}(\Omega)$: The second order integrated saddlepoint $\hat{c}_{\mathrm{CBQ,ISP}}(\Omega)$ approximations and the second order saddlepoint of integration $\hat{c}_{\mathrm{CBQ,\ SPI}}(\Omega)$ approximations with various values of κ , $\beta = 0.4\kappa$, λ_1, λ_2 and $n = 1000$.

for the complex Bingham quartic distribution is the same as for $(2k-4)$-dimensional multivariate normal distribution. Indeed, the main reason that the complex Bingham quartic distribution has the number and choice of quartic terms that it does is so that it can match the general multivariate distribution under high concentrations (Kent et al. [46]).

λ_1	λ_2	λ_3	λ_4	$\hat{c}_{\mathrm{CBQ}}$ ISP	$\hat{c}_{\mathrm{CBQ}}$ SPI	Difference ISP & SPI	$\hat{c}_{\mathrm{CBQ}}(\Omega)$ CVISP	$\hat{c}_{\mathrm{CBQ}}(\Omega)$ CVSPI	Difference CVISP & CVSPI
1	0	0	0	15.249	13.221	2.028	16.044	14.911	1.133
1	1	0	0	14.912	12.911	2.001	14.986	13.885	1.101
1	1	1	0	14.548	12.571	1.977	14.684	13.626	1.058
10	0	0	0	3.1121	2.2101	0.902	3.1372	2.2852	0.852
10	10	0	0	2.8551	2.2041	0.651	2.9112	2.3102	0.601
10	10	10	0	2.5337	2.0097	0.524	2.6703	2.1613	0.509
25	0	0	0	1.0024	0.8734	0.129	1.0608	0.9368	0.124
25	25	0	0	0.9742	0.8532	0.121	1.0087	0.8937	0.115
25	25	25	0	0.9155	0.8175	0.098	0.9306	0.8386	0.092
50	0	0	0	0.5124	0.4214	0.091	0.5308	0.4448	0.086
50	50	0	0	0.5077	0.4207	0.087	0.5186	0.4356	0.083
50	50	50	0	0.5004	0.4194	0.081	0.5082	0.4322	0.076
100	0	0	0	0.1406	0.1266	0.014	0.1477	0.1367	0.011
100	100	0	0	0.1277	0.1187	0.009	0.1297	0.1227	0.007
100	100	100	0	0.1182	0.1142	0.004	0.1193	0.1163	0.003

Table 3.2: Numerical results for the second order integrated saddlepoint $\hat{c}_{\mathrm{CBQ,ISP}}(\Omega)$ and the second order saddlepoint of integration $\hat{c}_{\mathrm{CBQ,\ SPI}}(\Omega)$ approximations with/without change of variable, with λ_1, λ_2, λ_3 and λ_4 varying and with $n = 1000$.

Part II

Rejection Simulation Techniques

Chapter 4

Simulation of the Bingham Distribution Using an Inequality for Concave Functions

4.1 Introduction

Modern Markov chain Monte Carlo (MCMC) methods of simulation-based inference have renewed the need for effective simulation algorithms for wide classes of random variables. In particular, directional distributions play an important role in many geometric problems such as computer vision and protein structure.

Over the years many specialized simulation methods have been developed for specific directional distributions, but little methodology has been produced for the directional distributions on higher dimensional manifolds and shape distributions.

The main purpose of this chapter is to develop an efficient acceptance-rejection (A/R) simulation algorithm for the Bingham distribution on unit sphere in $\mathbb{R}^q$ using an angular central Gaussian (ACG) envelope. Three special cases are then discussed.

The presentation proceeds in several stages. Firstly a review is given for the general A/R simulation algorithm. Secondly a general class of inequalities is given based on concave functions. These inequalities are illustrated for the multivariate normal distribution by finding two envelopes, viz., the multivariate Cauchy and the multivariate Bilateral exponential distributions, respectively. An inequality similar to that is used to show that the angular central Gaussian (ACG) density can

be used as an envelope for the Bingham density.

There are three special cases of interest in which the unit sphere S^{q-1} in $\mathbb{R}^q$ can be mapped to another space. In the first case, the circle on S^1 is identified with itself via a 2-to-1 mapping based on angle doubling (see Mardia and Jupp [70], pp 54 and pp 182). Under this mapping the Bingham distribution becomes the von Mises and the angular central Gaussian becomes the wrapped Cauchy (Tyler [93], Kent and Tyler [44] and Auderset et. al. [3]). It turns out that the simulation strategy developed with an ACG envelope for the Bingham distribution is identical to the simulation algorithm of Best-Fisher for the von Mises density (Best and Fisher [5]).

In the second special case, S^3 is identified with $SO(3)$, the space of special orthogonal group of all 3×3 rotation matrices, via a 2-to-1 mapping (see Prentice [80] and Mardia and Jupp [70], p. 285). Further, the Bingham distribution on the unit sphere S^3 becomes the matrix Fisher distribution on $SO(3)$. Hence our simulated Bingham random vector via the ACG envelope can be transformed into simulated matrix Fisher rotation matrices.

In the third special case, there is an isometric mapping between the shape space for labelled triangles in a plane, Σ_2^3 and the complex projective space $\mathbb{C}P^1$ which is identical to the sphere in three dimensions with radius $\frac{1}{2}$ i.e. $\Sigma_2^3 = \mathbb{C}P^1(4) = S^2(1/2)$ (Kent [41], Kendall [37] and Dryden and Mardia [18], p.69). Further, the Fisher and FB5 distributions defined on S^2 map to the complex Bingham (CB) and the complex Bingham quartic (CBQ) distributions defined on $\mathbb{C}P^1$, respectively. This mapping is discussed in more details in the next chapter, where a method is given to simulate from the complex Bingham quartic distribution.

The motivation in the current chapter is to develop simulation techniques which are (a) simple to program and (b) efficient for a wide range of the parameters of the Bingham distribution.

4.2 Principles of Acceptance-Rejection Simulation Scheme

The principal simulation approach in this chapter is based on the acceptance-rejection algorithm to simulate observations from a density f when a method is available to simulate from another density g on some space $\mathfrak{X}$ with respect to some base measure $\mu(dx)$. If $\mathfrak{X} = \mathbb{R}$, we usually let $\mu(dx) = dx$ as a Lebesgue measure. If $\mathfrak{X} = S^{q-1}$, the unit sphere S^{q-1} in $\mathbb{R}^q$, we usually let $\mu(dx) = \omega(dx)$ as a surface area.

Let

$$f = c_f f^*, \qquad g = c_g g^*, \tag{4.1}$$

be two densities where c_f and c_g are (possibly unknown) normalizing constants and the starred version does not involve these constants. Note that the starred version is not uniquely defined: if f^* is multiplied by some factor then c_f is divided by the same factor. The key assumption is that

$$f^* \leq M^* g^*, \tag{4.2}$$

where M^* is a known bound. If it is possible to simulate from g, then we can simulate from f as follows:

(1) Simulate X from g and $U \sim U(0,1)$, independently of each other.

(2) If $U \leqslant f^*(X)/\{M^* g^*(X)\}$ accept X; otherwise reject X and repeat from step 1.

The efficiency of the algorithm is defined by M^{-1} where $M \geq 1$ is the expected number of iterations of the algorithm required until X is successfully generated from g. For the method to be useful,

(a) it must be easy to generate a realization from the envelope density,

(b) the efficiency should be as close to 1 as possible, and

(c) in particular, the efficiency should be bounded away from zero over a wide range of parameters.

Note that it is not necessary to know the normalizing constants c_f and c_g in order to apply the simulation algorithm; all that is needed is the starred bound M^*. However, the full bound $M = c_f M^*/c_g$ is needed to give a theoretical assessment of the efficiency of the algorithm. Note also that the inequality $f^* \leq M^* g^*$ is equal to the inequality $f \leq Mg$.

The A/R algorithm can be justified as follows.

$$P(\text{accept}|X \in (x, x+dx)) = P\left(U \leq \frac{f^*(X)}{M^* g^*(X)} | X \in (x, x+dx)\right) = \frac{f^*(x)}{M^* g^*(x)} = \frac{f(x)}{Mg(x)},$$

and

$$\begin{aligned} P(\text{accept}) &= \int_{\mathbb{R}} P\big(\text{accept}|x\big) g(x) \mu(dx) \\ &= \int_{\mathbb{R}} \frac{f(x)}{Mg(x)} g(x) dx \\ &= \frac{1}{M} \int_{\mathbb{R}} f(x) dx = \frac{1}{M}, \end{aligned}$$

(See e.g. Kotz [39], p. 186, Rizzo [85], p. 56 and Dagpunar [12], p. 54). To see that the accepted sample has the same density f, apply Bayes' Rule,

$$\begin{aligned} P(X \in (x, x+dx)|\text{accept}) &= \frac{P(\text{accept}|X \in (x, x+dx))g(x)dx}{P(\text{accept})} \\ &= \frac{\Big[f(x)/\big(Mg(x)\big)\Big]g(x)dx}{1/M} \\ &= f(x)dx. \end{aligned}$$

Note that the proportion of X-values in the population which lie in the elemental range between x and $x+dx$, in other words the probability $P(x \leq X \leq x+dx)$, is $f(x)dx$ and it is called the probability density element (pde) of X and $f(x)$ is the probability density function (pdf) of X (Fisher et. al [22], p. 67).

The accuracy of the simulated efficiencies is assessed using their standard errors. The standard errors for the simulated efficiencies can be evaluated as $\sqrt{\delta(1-\delta)/n}$ where δ is the reciprocal of full efficiency bound $\delta = 1/M$ of the algorithm (Walker [95]).

4.3 Envelopes Based on Concave Functions

Let $\varphi(u)$ be a twice-differentiable increasing strictly concave function on $u > 0$, satisfies the following properties

(a) $\varphi(0+) = 0$,

(b) $\varphi^{'}(u) > 0$ for all $u > 0$ and $\varphi^{'}(u) \to 0$ as $u \to \infty$,

(c) $0 < \varphi^{'}(0+) \leq \infty$,

(d) $\varphi^{''}(u) < 0$ for all $u > 0$.

Suppose we wish to majorize this concave function $\varphi(u)$ by a linear function $\psi(u) = a+u$, $u > 0$ of slope 1 where the intercept a is to be chosen as small as possible. For the majorization we need $\psi(u) = a + u \geq \varphi(u)$ or equivalently

$$-u \leq -\varphi(u) + a \quad \text{for all } u \geq 0. \tag{4.3}$$

There are two cases to consider. If $\varphi^{'}(0) \leq 1$, the smallest feasible value of a is $a = 0$ and the two curves touch at $u_0 = 0$. We will call this the Type I case and the left side in the Figure (4.1) displays this situation.

On the other hand, if $\varphi^{'}(0) > 1$, there is a unique $u_0 > 0$ for which the slope $\varphi^{'}(u_0) = 1$ since $\varphi^{'}$ is a decreasing function. In this case the linear function $\psi(u)$ is tangent to the function $\varphi(u)$ at u_0. That is, $\psi(u) = \varphi(u_0) + \varphi^{'}(u_0)(u - u_0)$. We will call this the Type II case and the right side in the Figure (4.1) displays this second case.

Example 1

Let $q > 0$, $b > 0$ and consider the logarithm function

$$\varphi(u) = \frac{q}{2}\log\Big(1 + \frac{2u}{b}\Big), \quad u \geq 0, \tag{4.4}$$

Note that $\varphi(0) = 0$, $\varphi^{'}(u) = \big[q/(b+2u)\big] \to 0$ as $u \to \infty$, $\varphi^{'}(0) = q/b$ and $\varphi^{''}(u) = \big[-2q/(b+2u)^2\big] < 0$. Suppose we wish to majorize this function by the linear function $\psi(u) = a + u$, $u \geq 0$. For $\varphi^{'}(0) = q/b \leq 1$, $b \geq q$, so we have the Type I case and the minimum value of a is $a = 0$. In this case the inequality (4.3) becomes

$$-u \quad \leq \quad -\frac{q}{2}\log\Big(1 + \frac{2u}{b}\Big), \ b \geq q, \ u \geq 0. \tag{4.5}$$

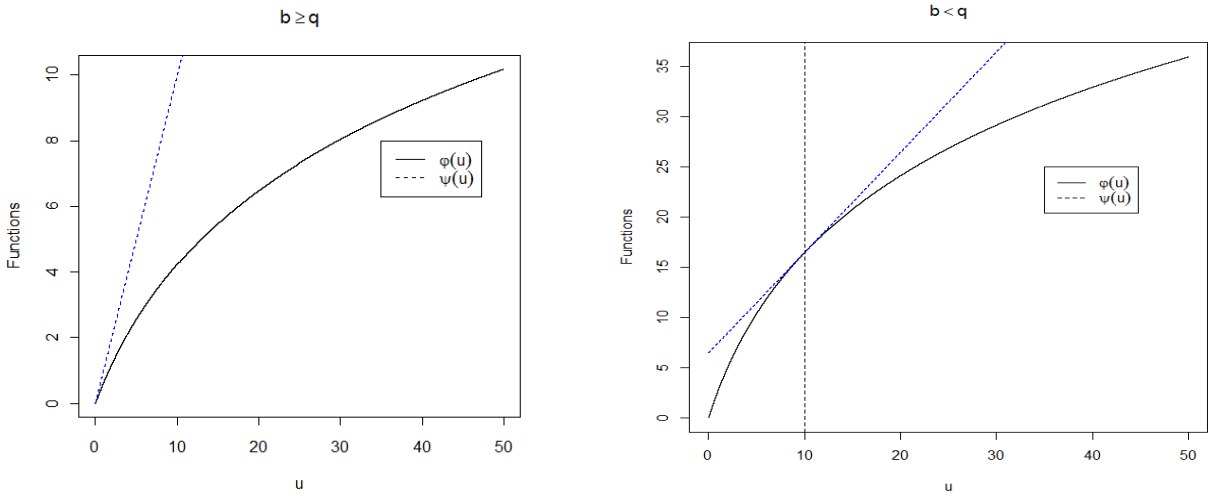

Figure 4.1: The concave function $\varphi(u)$ and the linear function $\psi(u)$ versus u. In the left panel $b = 15$ and $q = 10$. In the right panel $b = 10$, $q = 30$ and $u_0 = 10$.

If $\varphi^{'}(0) = q/b > 1$, $b < q$, so we have the Type II case and we can find a unique $u_0 > 0$ with slope $\varphi^{'}(u_0) = 1$, namely $u_0 = \frac{1}{2}(q - b)$. Choosing a so that the two curves touch at u_0, $\big[\varphi(u_0) = \psi(u_0)\big]$,

yields

$$a_0 = \frac{q}{2}\log\Big(\frac{q}{b}\Big) - \frac{1}{2}(q-b). \tag{4.6}$$

Putting these pieces together into the inequality (4.3) yields

$$-u \;\leq\; -\frac{q}{2}\log\Big(1+\frac{2u}{b}\Big) + \frac{q}{2}\log\Big(\frac{q}{b}\Big) - \frac{1}{2}\Big(q-b\Big), \quad u \geq 0,\ b < q. \tag{4.7}$$

Example 2

Let $b > 0$ and consider the squared root function

$$\varphi(u) = bu^{1/2}, \tag{4.8}$$

Note that $\varphi^{'}(u) = (1/2)bu^{-1/2}$ so that $\varphi^{'}(0+) = \infty$, $\lim \varphi^{'}(u)$ as $u \to \infty$ and $\varphi^{''}(u) = -(1/4)bu^{-3/2} < 0$. For all $b > 0$ and $\varphi^{'}(0+) > 1$, we are in the type II case. Also, there is a unique $u_0 > 0$ with slope $\varphi^{'}(u_0) = 1$, namely $u_0 = \frac{1}{4}b^2$. Choosing a so that $\varphi(u)$ touch at u_0 with $\psi(u) = a + u$, $u \geq 0$, yields $a_0 = \frac{1}{2}b^2$ and the inequality (4.3) gives a majorization of the form

$$-\frac{1}{2}u \;\leq\; \Big(\frac{1}{4}b^2 - \frac{1}{2}bu^{1/2}\Big) \text{ for all } u \geq 0. \tag{4.9}$$

The concave inequality (4.3) can be used to obtain an A/R envelope for dominating a function f on sample space $\mathfrak{X}$. Let $u = h(x)$ be a given function of x and suppose the two starred densities take the form $f^*(x) = \exp(-h(x))$ and $g^*(x) = \exp(-\varphi(h(x))$. Then the inequality (4.3) implies that

$$\underbrace{\exp(-h(x))}_{f^*(x)} \leq \exp(a)\underbrace{\exp(-\varphi(h(x))}_{g^*(x)} \quad u \geq 0. \tag{4.10}$$

and gives a starred bound $M^* = \exp(a)$. If $\varphi^{'}(0+) \leq 1$, the inequality yields an envelope of Type I with starred bound $M^* = 1$ whereas if $\varphi^{'}(0+) > 1$, the inequality yields an envelope of Type II with starred bound $M^* = \exp(a_0)$ and u_0 can be calculated using $\varphi^{'}(u_0) = 1$.

In the next sections we give more advanced examples of this strategy of simulation.

4.4 Simulation from the Multivariate Normal Distribution with Multivariate Cauchy Envelope

We can use the inequality (4.7) to dominate a multivariate normal density in $p \geq 1$ dimensions. The multivariate normal density $\text{MVN}_p(\Sigma)$ with mean 0 and variance-covariance matrix Σ has probability density function,

$$f(\mathbf{x}) = (2\pi)^{-p/2}|\Sigma|^{-1/2}\exp\Big(-\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\Big) = c_{\text{MN}}\, f^*(\mathbf{x}), \tag{4.11}$$

where $c_{\text{MVN}} = (2\pi)^{-p/2}|\Sigma|^{-1/2}$ is the constant of normalization. The multivariate Cauchy density $\text{MVC}_p(\Phi)$ centered at the origin with definite positive matrix Φ has probability density function,

$$g(\mathbf{x}) = \frac{\Gamma[(p+1)/2]}{\pi^{(p+1)/2}}|\Phi|^{-1/2}\Big(1+\mathbf{x}^T\Phi^{-1}\mathbf{x}\Big)^{-(p+1)/2} = c_{\text{MC}}\, g^*(\mathbf{x}), \tag{4.12}$$

where $c_{\text{MVC}} = \Gamma[(p+1)/2]/\pi^{(p+1)/2}|\Phi|^{-1/2}$ (Johnson and Kotz [33], p. 294).

Suppose Σ is given. Let $b > 0$ be an arbitrary constant and let $\Phi = b\Sigma$. Also if we set $q = p+1$, $u = \frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x}$, so that $\mathbf{x}^T\Phi^{-1}\mathbf{x} = 2u/b$ and $|\Phi|^{-1/2} = b^{-(p+1)/2}|\Sigma|^{-1/2}$, then for $b \geq q$ the inequality (4.5) yields

$$-\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x} \leq -\frac{(p+1)}{2}\log\Big(1+\mathbf{x}^T\Phi^{-1}\mathbf{x}\Big). \tag{4.13}$$

Exponentiating this inequality yields the Type I envelope

$$\underbrace{\exp\Big(-\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\Big)}_{f^*(\mathbf{x})} \leq \underbrace{\Big(1+\mathbf{x}^T\Phi^{-1}\mathbf{x}\Big)^{-(p+1)/2}}_{g^*(\mathbf{x})}, \tag{4.14}$$

with starred bound $M^* = \exp(0) = 1$.

For $b < q$ the inequality (4.7) yields

$$-\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x} \leq -\frac{(p+1)}{2}\log\Big(1+\mathbf{x}^T\Phi^{-1}\mathbf{x}\Big) + \frac{p+1}{2}\log\Big(\frac{p+1}{b}\Big) - \frac{1}{2}\Big[(p+1)-b\Big]. \tag{4.15}$$

Exponentiating this inequality yields the Type II envelope

$$\underbrace{\exp\Big(-\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\Big)}_{f^*(\mathbf{x})} \leq \Big(\frac{p+1}{b}\Big)^{(p+1)/2}\exp\Big(-\frac{1}{2}((p+1)-b)\Big)\underbrace{\Big(1+\mathbf{x}^T\Phi^{-1}\mathbf{x}\Big)^{-(p+1)/2}}_{g^*(\mathbf{x})}, \tag{4.16}$$

with starred bound

$$M^* \;=\; \Big(\frac{p+1}{b}\Big)^{(p+1)/2}\exp\Big(-\frac{1}{2}((p+1)-b)\Big). \tag{4.17}$$

Since the normalizing constants for both multivariate distributions are known, we can evaluate the full bound of this A/R simulation scheme,

$$\begin{aligned} M(b,p) &= \frac{c_{\mathrm{MVN}}}{c_{\mathrm{MVC}}} M^* \\ &= \frac{c_{\mathrm{MVN}}}{c_{\mathrm{MVC}}} \Big\{ \frac{(p+1)}{b} \Big\}^{(p+1)/2} \exp\Big\{ -\frac{1}{2}\Big((p+1) - b \Big) \Big\} \\ &= b^{-1/2} \exp\Big(\frac{1}{2} b\Big) (2\pi)^{-p/2} \Big(\frac{(p+1)}{b} \Big)^{\frac{(p+1)}{2}} \ \frac{\exp\Big(-\frac{1}{2}((p+1)-b) \Big)}{\Gamma[(p+1)/2]} \\ &= c_p b^{-\frac{(p+2)}{2}} \ \exp(b), \ \text{ say,} \end{aligned} \tag{4.18}$$

where

$$c_p = (2\pi)^{-p/2} \Big(p+1 \Big)^{\frac{(p+1)}{2}} \exp\Big[-\frac{1}{2}(p+1) \Big]. \tag{4.19}$$

Minimizing $\log M(b,p)$ with respect to b shows that the minimum value is attained when $b = (p+2)/2$ and

$$M(p) = (2\pi)^{-p/2} \Big(\frac{p+2}{2} \Big)^{-1/2} \Big(\frac{2(p+1)}{p+2} \Big)^{\frac{(p+1)}{2}} \ \frac{\exp\Big(-\frac{1}{2}(p+1) \Big)}{\Gamma[(p+1)/2]}. \tag{4.20}$$

Table 4.1 gives some values of the full bound M with various values of p and $b = (p+2)/2$. For the values of p of typical interest in applications (e.g. $p = 1, 2, 3$), these bounds are entirely satisfactory.

In passing, we note that this section describes a theoretical method rather than practical interest. In practice we usually use other efficient simulation algorithms to generate samples from the multivariate normal density e.g. the Box-Muller method (Jones et. al. [35], pp. 347-348, Kennedy and Gentle [39], pp. 200-202 and Rizzo [85], pp. 70-76).

4.5 Bilateral Exponential Envelope for Standard Normal Distribution

Another example for generating the MVN density is to use the multivariate Bilateral exponential (MBLE) density as an envelope. We limit attention here to the case $p = 1$. Let f denote the probability density function of the standard normal distribution $N(0,1)$,

$$f(x) = c_f \exp\Big(-\frac{1}{2} x^2 \Big) = c_f f^*(x), \qquad -\infty < x < \infty. \tag{4.21}$$

p	b	M	Efficiency Rate
1	1.5	1.52	66%
2	2.0	1.91	52%
3	2.5	2.24	45%
4	3.0	2.52	40%
5	3.5	2.78	36%
10	6.0	3.81	26%
50	26	8.30	12%
100	51	11.70	9%

Table 4.1: Analytical efficiencies for Multivariate Normal/Multivariate Cauchy Envelope A/R simulation with various values of p and b.

with the normalizing constant $c_f = (2\pi)^{-1/2}$. Consider the bilateral exponential density (Laplace distribution) with scale parameter $\alpha > 0$ of the form

$$g(x) = g_f \exp(-\alpha|x|) = c_f g^*(x), \qquad x > 0. \tag{4.22}$$

with the normalizing constant $c_g = \frac{1}{2}\alpha$. Figure 4.2 plots these functions with various values of α.

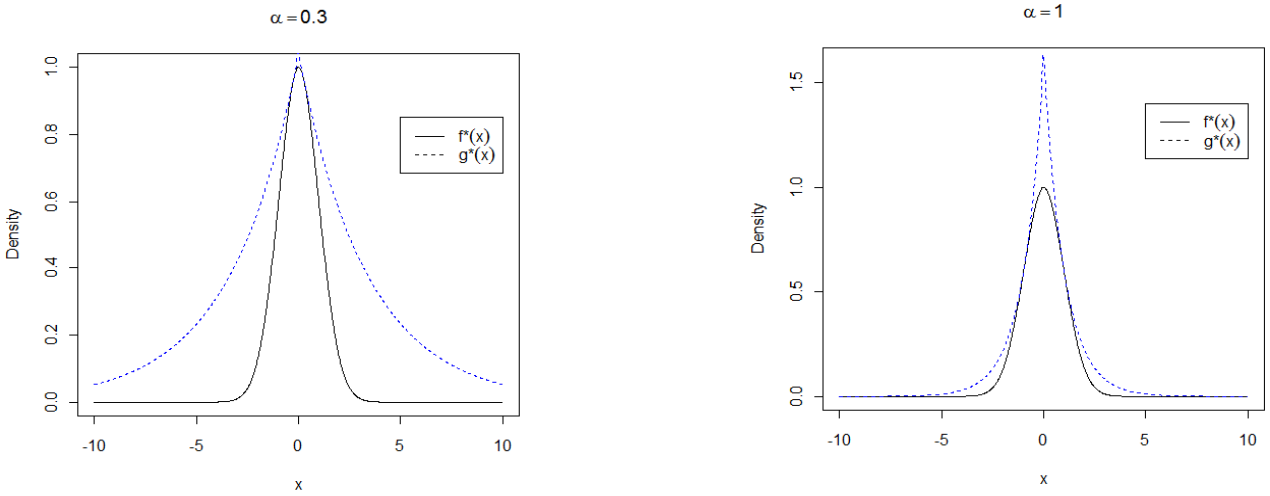

Figure 4.2: The unnormalized standard normal function $f^*(x)$ and the unnormalized bilateral envelope function $g^*(x)$ versus x. In the left panel $\alpha = 0.3$ whereas in the right panel $\alpha = 1.0$.

Let $b > 0$ be an arbitrary constant and let $\alpha = \frac{1}{2}b$. Also if we set $u = x^2$, then the inequality

(4.9) yields

$$-\frac{1}{2}x^2 \leq \Big(\alpha^2 - \alpha|x|\Big). \tag{4.23}$$

Exponentiating this inequality yields the Type II envelope

$$\underbrace{\exp\big(-\frac{1}{2}x^2\big)}_{f^*(x)} \leq \exp(\alpha^2)\underbrace{\exp\big(-\alpha|x|\big)}_{g^*(x)}, \tag{4.24}$$

with starred bound

$$M^* = \exp(\alpha^2). \tag{4.25}$$

Moreover, since the normalizing constants for both distribution are known we can calculate the full bound of the simulation

$$\begin{aligned} M(\alpha) &= \frac{c_f}{c_g}M^* \\ &= \Big(\frac{2}{\pi}\Big)^{1/2}\alpha^{-1}\exp(\alpha^2). \end{aligned} \tag{4.26}$$

If we minimize M with respect to α we find $\alpha = \sqrt{1/2}$, with optimal bound $M = 1.315489$. Note that the bilateral exponential is slightly more efficient than the Cauchy envelope.

4.6 Simulation from the Real Bingham Distribution with ACG Envelope

The Bingham distribution $\text{Bing}_q(\mathbf{A})$ lies on the unit sphere S^{q-1} in $\mathbb{R}^q$ and is parameterized by $q \times q$ symmetric concentration matrix $\mathbf{A}$. The density of a directional random vector $\mathbf{x}$ with respect to the surface measure $w_q(d\mathbf{x})$ on S^{q-1} is given by

$$f_{\text{Bing}}(\mathbf{x};\mathbf{A}) = {}_1F_1\left(\frac{1}{2};\frac{q}{2};\mathbf{A}\right)^{-1}\exp\big(-\mathbf{x}^T\mathbf{A}\mathbf{x}\big) = c_{\text{Bing}}(\mathbf{A})\, f^*_{\text{Bing}}(\mathbf{x};\mathbf{A}), \tag{4.27}$$

where the normalizing constant $c_{\text{Bing}}(\mathbf{A}) = {}_1F_1\Big(\frac{1}{2};\frac{q}{2};\mathbf{A}\Big)$ is the hypergeometric function of matrix argument (Mardia and Jupp [59], p. 289). Further, for simulation purposes, we can assume that $\mathbf{A} = \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_q)$ where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_q = 0$ (Mardia et al. [72], p. 181 and Kume and Walker [52]). Because of the constraint $\mathbf{x}^T\mathbf{x} = 1$, the matrix $\mathbf{A}$ and $\mathbf{A} + \alpha I_q$ for any real number α, define the same real Bingham distribution with $c_{\text{Bing}}(\mathbf{A} + \alpha I_q) = \exp(\alpha)c_{\text{Bing}}(\mathbf{A})$.

The family of angular central Gaussian (ACG) distributions is an alternative to the family of Bingham distributions for modelling antipodal symmetric directional data (Tyler [93]). An angular central Gaussian (ACG) distribution on the $(q-1)$-dimensional sphere S^{q-1} can be generating by projecting a multivariate Gaussian distribution in $\mathbb{R}^q$, $q \geq 2$ with mean zero onto S^{q-1} with radius one. That is, if $\mathbf{y}$ has a multivariate normal distribution in $\mathbb{R}^q$ with mean vector zero and variance covariance matrix Ψ, then $\mathbf{x} = \mathbf{y}/\|\mathbf{y}\|$ has an ACG distribution on S^{q-1} with $q \times q$ symmetric positive definite parameter matrix Ψ (Mardia and Jupp [70], p. 46 and p. 182). The probability density density for $\mathbf{x}$ which is denoted by $\mathrm{ACG}_q(\Psi)$ is given by

$$g_{\mathrm{ACG}}(\mathbf{x};\Psi) = w_q^{-1}|\Psi|^{-1/2}\big(\mathbf{x}^T\Psi^{-1}\mathbf{x}\big)^{-q/2} = c_{\mathrm{ACG}}(\Psi)\, g^*_{\mathrm{ACG}}(\mathbf{x}), \tag{4.28}$$

with respect to the surface measure on S^{q-1}. The constant $w_q = 2\pi^{q/2}/\Gamma(q/2)$ represents the surface area on S^{q-1}, $c_{\mathrm{ACG}}(\Psi) = w_q^{-1}|\Psi|^{-1/2}$ is the normalizing constant (Tyler [93] and Kent and Tyler [44]) and Ψ is a $q \times q$ symmetric positive-definite parameter matrix. Note that Ψ is only identifiable up to multiplication by a positive scalar since for any $c > 0$, $g_{\mathrm{ACG}}(\mathbf{x};\Psi) = g_{\mathrm{ACG}}(\mathbf{x};c\Psi)$.

Next we show that a possible choice of an envelope for the Bingham distribution in an A/R algorithm is the angular central Gaussian (ACG) distribution. Set $u = \mathbf{x}^T\mathbf{A}\mathbf{x}$ and $\Psi^{-1} = I_q + \frac{2}{b}\mathbf{A}$, so that $\mathbf{x}^T\Psi^{-1}\mathbf{x} = 1 + \frac{2}{b}\mathbf{x}^T\mathbf{A}\mathbf{x}$. Then for $b < q$, the inequality (4.7) yields

$$-\mathbf{x}^T\mathbf{A}\mathbf{x} \leq -\frac{q}{2}\log\Big(\mathbf{x}^T\Psi^{-1}\mathbf{x}\Big) + \frac{q}{2}\log\Big(\frac{p}{b}\Big) - \frac{1}{2}\big(q-b\big). \tag{4.29}$$

Exponentiating this inequality yields the type II envelope

$$\underbrace{\exp\Big(-\mathbf{x}^T\mathbf{A}\mathbf{x}\Big)}_{f^*_{\mathrm{Bing}}(\mathbf{x};\mathbf{A})} \leq \Big(\frac{q}{b}\Big)^{q/2}\, \exp\Big[-\frac{1}{2}\big(q-b\big)\Big]\, \underbrace{\Big(\mathbf{x}^T\Psi^{-1}\mathbf{x}\Big)^{-q/2}}_{g^*_{\mathrm{ACG}}(\mathbf{x};\Psi)}, \tag{4.30}$$

with starred bound

$$M^*(q,b) \;=\; \Big(\frac{q}{b}\Big)^{q/2}\, \exp\Big[-\frac{1}{2}\big(q-b\big)\Big], \tag{4.31}$$

and full bound

$$\begin{aligned} M(q,b;\mathbf{A},\Psi) &= \frac{c_{\mathrm{Bing}}(\mathbf{A})}{c_{\mathrm{ACG}}(\Psi)} M^*(q,b) \\ &= \frac{2\pi^{q/2}}{{}_1F_1\Big(\frac{1}{2};\frac{q}{2};\mathbf{A}\Big)|\Psi^{-1}|^{1/2}\Gamma\Big(\frac{q}{2}\Big)}\Big(\frac{q}{b}\Big)^{q/2}\, \exp\Big[-\frac{1}{2}\big(q-b\big)\Big]. \end{aligned} \tag{4.32}$$

It is well-known that under high concentration the Bingham distribution $\text{Bing}_q(\mathbf{A})$ is asymptotically multivariate normal $\text{MVN}_p(\Sigma)$ where $\mathbf{A} = \frac{1}{2}\Sigma^{-1}$ and the angular central Gaussian distribution $\text{ACG}_q(\Psi)$ with definite positive matrix Ψ is asymptotically multivariate Cauchy $\text{MVC}_p(\Phi)$ with definite positive matrix Φ (Tyler [93] and Auderset et. al. [3]). So, we expect their efficiencies to be similar for the same b.

The inequality (4.30) involves a tuning constant b. This can be found analytically for the multivariate normal distribution with a multivariate Cauchy envelope, and this value can be used as an approximate optimal value of b for the Bingham distribution with an ACG envelope. The optimal value of b is then approximately $b = (q+2)/2$. The true optimal value of b for the Bingham distribution with an ACG envelope can be found by simple numerical optimization.

It is known in the case $q = 2$ that the Bingham distribution is a 2-wrapped von Mises by doubling angles (Mardia and Jupp [70], p. 54 and p. 182). That is, for a 2×2 symmetric matrix $\mathbf{A}$ in $\mathbb{R}^2$, the exponent term for the Bingham density in the polar co-ordinates $(1, \theta)$, becomes

$$-\mathbf{x}^T\,\mathbf{A}\,\mathbf{x} \;=\; \begin{bmatrix}\cos\theta & \sin\theta\end{bmatrix}\begin{bmatrix}\lambda & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}\cos\theta\\ \sin\theta\end{bmatrix} = -\lambda\cos^2\theta = -\frac{\lambda}{2}\Big(1+\cos 2\theta\Big), \tag{4.33}$$

where we used the double angle formula $\cos^2\theta = (1+\cos 2\theta)/2$. Thus the Bingham distribution becomes

$$\begin{aligned} f_{\text{Bing}}(\mathbf{x};\mathbf{A}) &\propto \exp\big(-\mathbf{x}^T\mathbf{A}\mathbf{x}\big)\\ &\propto \exp\Big(-\frac{\lambda}{2}\cos 2\theta\Big)\\ &= \exp\Big[\frac{\lambda}{2}\cos\Big(2\theta-\pi\Big)\Big] = \exp\big[\kappa\cos\big(2\theta-\pi\big)\big], \end{aligned} \tag{4.34}$$

which is the doubly-wrapped von Mises distribution with mode or mean direction at $\mu = \pi$ and a normalizing constant $c_{\text{Bing}}(\kappa) = [2\pi\, I_0(\kappa)]^{-1}$ and $\kappa = \lambda/2$.

Next it is possible analytically to find the optimal value of b that minimizes M as follows. The inverse of the 2×2 symmetric positive definite parameter matrix Ψ^{-1} is $\Psi^{-1} = I_2 + \frac{2}{b}\mathbf{A} = \text{diag}(1+\frac{4}{b}\kappa, 1)$ and $|\Psi^{-1}| = 1+\frac{4}{b}\kappa$. Hence the full bound in (4.32) becomes

$$M(b,\kappa) \;=\; \frac{1}{I_0(\kappa)\sqrt{1+\frac{4}{b}\kappa}}\Big(\frac{2}{b}\Big)\,\exp\Big[-\frac{1}{2}\Big(2-b\Big)\Big]. \tag{4.35}$$

The log of the full bound M can be written as

$$\log M(b,\kappa) \;=\; \log\Big(\frac{2}{b}\Big) - \frac{1}{2}\Big(2-b\Big) - \log\Big[\sqrt{1+\frac{4}{b}\kappa}\Big] - \log\, I_0(\kappa). \tag{4.36}$$

Minimizing $\log M(b,\kappa)$ with respect to b shows that the optimal value of b is

$$b(\kappa) = 1 + \sqrt{1+4\kappa^2}, \tag{4.37}$$

and $M(\kappa) \geq 1$ becomes

$$\begin{aligned} M(\kappa) &= \frac{1}{I_0(\kappa)}\left(\frac{2}{1+\sqrt{1+4\kappa^2}}\right)\left(\frac{1+\sqrt{1+4\kappa^2}}{1+\sqrt{1+4\kappa^2}+4\kappa}\right)^{1/2} \\ &\quad \exp\Big[-\frac{1}{2}\Big(1-\sqrt{1+4\kappa^2}\Big)\Big]. \end{aligned} \tag{4.38}$$

For the case $q \geq 3$, it is not easy in practice to calculate the full bound of the algorithm M since the normalizing constant for the Bingham distribution involves the multivariate hypergeometric function which is analytically cumbersome. R functions to implement this simulation scheme are available in Appendix A. Overall, we have the following accept-reject algorithm.

(1) For given $q \times q$ symmetric concentration matrix $\mathbf{A}$ find Ψ.

(2) Generate q-vectors from the angular central Gaussian (ACG) distribution on the sphere, $g^*_{\text{ACG}}(\mathbf{x})$ with parameter matrix Ψ in equation (4.28).

(3) Generate a random variable U from the Uniform$(0,1)$ distribution.

(4) If $U \leqslant \dfrac{f^*_{\text{Bing}}(\mathbf{x};\mathbf{A})}{M^* g^*_{\text{ACG}}(\mathbf{x};\Psi)}$, accept $\mathbf{x}$; otherwise reject $\mathbf{x}$ and repeat from step 1.

Table 4.2 gives some values of the simulated efficiencies with various values of λ_1, λ_2, λ_3 and $b = 2.5 < q = 3$. The simulated efficiency rate under this A/R algorithm is found satisfactory for generating pseudo random sample of size $n = 10000$ from the real Bingham distribution with low and high concentration parameters. The simulated efficiency rates are very reasonable under low concentrations. Under high concentrations, $\lambda_1 = \lambda_2 = 100$ and $\lambda_3 = 0$, the simulated efficiency rate is close to the efficiency rate in Table 4.1 for simulating the multivariate normal distribution with a multivariate Cauchy envelope.

λ_1	λ_2	λ_3	Simulated Efficiency Rate	Standard Errors
1	0	0	91.2%	0.00283
1	1	0	89.7%	0.00304
10	0	0	84.1%	0.00366
10	1	0	81.7%	0.00387
10	10	0	77.4%	0.00418
100	0	0	57.6%	0.00494
100	10	0	53.4%	0.00497
100	100	0	46.1%	0.00499

Table 4.2: Simulated efficiencies rates and their standard errors for Bingham/Angular central Gaussian (ACG) Envelope A/R simulation with $n = 10000$ and various values of λ_1, λ_2 and $\lambda_3 = 0$.

4.7 Simulation from the von Mises Distribution with a Wrapped Cauchy Envelope

For the special case on the circle, $q = 2$, consider the von Mises distribution on the unit circle with zero mean direction and probability density function given by

$$f(\theta; 0, \kappa) = c_f \exp\{\kappa \cos\theta\} = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos\theta\}, \qquad \kappa > 0, \quad -\pi \leq \theta \leq \pi, \tag{4.39}$$

where κ is known as a concentration parameter, $c_f = \left[2\pi I_0(\kappa)\right]^{-1}$ is the normalizing constant and $I_0(\kappa)$ denotes the modified Bessel function of the first kind and order zero.

Best and Fisher [5] use a wrapped Cauchy distribution $WC(0, \rho)$,

$$g(\theta; 0, \rho) \;=\; \frac{1}{2\pi} \frac{1 - \rho^2}{1 + \rho^2 - 2\rho\cos\theta}, \quad -\pi \leq \theta \leq \pi, \tag{4.40}$$

where ρ represents to a scale parameter with $0 \leq \rho < 1$. As indicated by its name, the location-scale family of wrapped Cauchy distribution is generated by wrapping the location-scale family of Cauchy distributions on the line about the unit circle, or in other words by expressing the latter mod(2π). The distribution function of the wrapped Cauchy distribution is

$$G(\theta) = \frac{1}{2\pi} \cos^{-1}\left\{ \frac{(1 + \rho^2)\cos\theta - 2\rho}{1 + \rho^2 - 2\rho\cos\theta} \right\}, \quad -\pi \leq \theta \leq \pi \tag{4.41}$$

The optimal value of ρ, and the associated sampling efficiency, M^{-1}, are determined by

$$\begin{aligned} M(\rho,\kappa) &= \min_{\rho}\max_{\theta}\{f(\theta)/g(\theta)\} \\ &= \min_{\rho}\max_{\theta}\left\{\frac{\exp\{\kappa\cos\theta\}(1+\rho^2-2\rho\cos\theta)}{(1-\rho^2)I_0(\kappa)}\right\} \\ &= \frac{\big(2\rho/\kappa\big)\exp\Big[\kappa(1+\rho^2)/2\rho-1\Big]}{\big(1-\rho^2\big)I_0(\kappa)}, \end{aligned} \tag{4.42}$$

where $\rho = \big(\tau-(2\tau)^{1/2}\big)/2\kappa$ and $\tau = 1+\big(1+4\kappa^2\big)^{1/2}$. The corresponding (θ,ρ) is therefore a saddlepoint of $f(\theta)/g(\theta)$. Best and Fisher [5] showed that

$$Mg(\theta) = \frac{(2\rho/\kappa)\exp\big\{\kappa(1+\rho^2)/2\rho-1\big\}}{2\pi I_0(\kappa)(1+\rho^2-2\rho\cos\theta)}, \tag{4.43}$$

is the best upper envelope for the von Mises density $f(\theta)$ in (4.39) and that the choice of ρ maximizes the acceptance ratio M^{-1} of this acceptance rejection scheme.

Take $\gamma(\theta) = \exp\{\kappa\cos\theta\}(1+\rho^2-2\rho\cos\theta)$ where $-\pi \le \theta \le \pi$. An optimal value of ρ will be determined below. Hence

$$\gamma^{'}(\theta) = 2\rho\sin\theta\exp\{\kappa\cos\theta\} - (1+\rho^2-2\rho\cos\theta)\kappa\sin\theta\exp\{\kappa\cos\theta\}, \tag{4.44}$$

and thus $\gamma^{'}(\theta) = 0$ when $\sin\theta = 0$ or $\cos\theta = (1+\rho^2-(2\rho/\kappa))/2\rho$. By examining $\gamma^{''}(\theta)$, we find that $\gamma(\theta)$ has a local maximum value $\gamma_1 \equiv (1-\rho)^2\exp(\kappa)$ at $\sin\theta = 0$ if

$$2\rho/(1-\rho)^2 < \kappa, \tag{4.45}$$

and a local maximum value $\gamma_2 \equiv (2\rho/\kappa)\exp\big\{\kappa(1+\rho^2)/2\rho-1\big\}$ at $\cos\theta = (1+\rho^2-(2\rho/\kappa))/2\rho$, if

$$2\rho/(1+\rho)^2 < \kappa < 2\rho/(1-\rho)^2. \tag{4.46}$$

In order to choose the best value of ρ each of the above maxima will now be examined to determine the value of ρ, in terms of κ, which minimizes M. Best and Fisher [5] computed the reciprocals of acceptance ratios for the two maxima. For $\sin\theta = 0$ the reciprocal of acceptance ratio $A_1(\rho_1)$, say, attains its minimum value at $\rho = \rho_1 = (\kappa+1-(1+2\kappa)^{1/2})/\kappa$. For $\cos\theta = (1+\rho^2-(2\rho/\kappa))/2\rho$ the reciprocal of acceptance ratio $A_2(\rho_2)$, say, attains its minimum value at $\rho = \rho_2 = \big(\tau-(2\tau)^{1/2}\big)/2\kappa$ where $\tau = 1+\big(1+4\kappa^2\big)^{1/2}$. Finally $A_2(\rho_2) < A_1(\rho_1)$, that is, the reciprocal of the acceptance ratios in the algorithm is minimized by choosing $\rho = \rho_2$ and $Mg(\theta) = \gamma_2/\big\{2\pi I_0(\kappa)(1+\rho^2-2\rho\cos\theta)\big\}$ is the best envelope for this ρ.

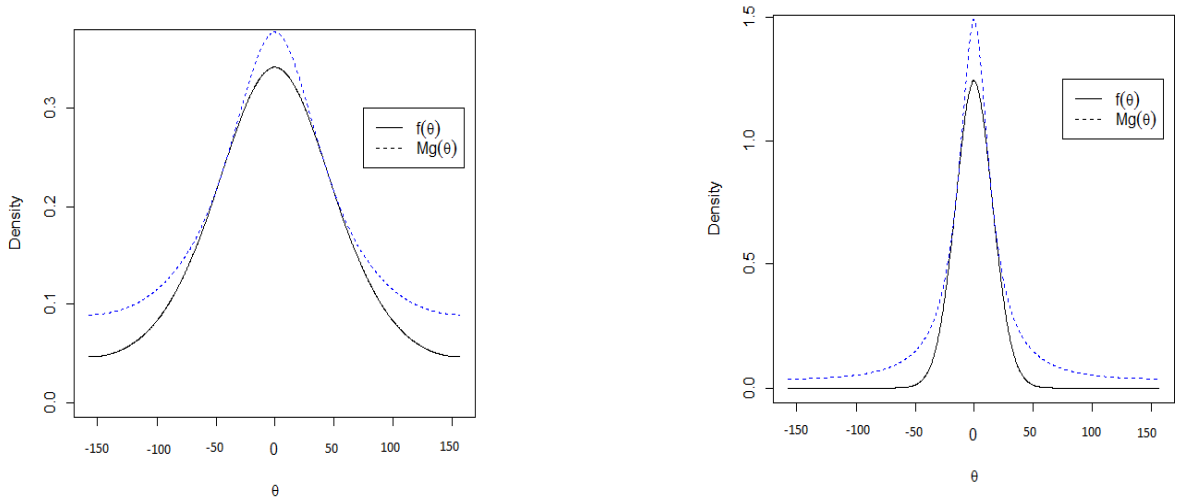

Figure 4.3: The von Mises function $f(\theta)$ and the the simulation proportional envelope of the wrapped Cauchy distribution $Mg_\Theta(\theta)$. In the left panel $\kappa = 1.0$, $\tau = 3.236$, $\rho = 0.346$ and $M = 1.152$ whereas in the right panel $\kappa = 10$, $\tau = 21.02$, $\rho = 0.727$ and $M = 1.481$.

Figure 4.3 plots the target probability function $f_\Theta(\theta)$ and the simulation proportional envelope $Mg_\Theta(\theta)$ with various values of κ, τ, ρ and M.

For large κ, one can use an asymptotic expansion approximation for $I_0(\kappa)$ (see Mardia & Jupp [70], p. 40) with $q = 2$.

$$I_0(\kappa) \cong \exp(\kappa)/\sqrt{2\pi\kappa} \tag{4.47}$$

and the full bound M becomes

$$M(\rho, \kappa) = \frac{2\sqrt{\pi/\kappa}\, \exp\Big\{\big(\kappa(1+\rho^2)/2\rho - 1\big) - \kappa\Big\}}{1-\rho^2}. \tag{4.48}$$

4.8 Link Between the Best-Fisher Method and the Concave Inequality

The majorizing inequality (4.3) for the Best-Fisher method can also be obtained from our concave inequality (4.7). Firstly note that there are at least 3 expressions for the probability density function of the wrapped Cauchy distribution. The first expression is given in (4.38) and the second expression

can be written as

$$\begin{aligned} f_{\mathrm{WC}}(\theta;0,\alpha) &= \frac{1-\rho^2}{2\pi(1+\rho^2)}\frac{1}{1-\frac{2\rho}{1+\rho^2}\cos\theta} \\ &\propto \frac{1}{1-\alpha\cos\theta}, \quad 0\le\alpha\le 1,\ -\pi\le\theta\le\pi, \end{aligned} \tag{4.49}$$

where $\alpha = 2\rho/1+\rho^2$. Moreover, if we substitute α by $\beta/1+\beta$ we get the third expression of the probability density function for the wrapped Cauchy distribution

$$f_{\mathrm{WC}}(\theta;0,\beta) \propto \frac{1}{1+\beta-\beta\cos\theta}, \quad 0\le\beta<\infty,\ -\pi\le\theta\le\pi. \tag{4.50}$$

The angular central Gaussian density can be transformed to a wrapped Cauchy distribution (Tyler [93], Kent and Tyler [44] and Mardia and Jupp [70], p. 52). Here we have

$$\begin{aligned} \mathbf{x}^T\Psi^{-1}\mathbf{x} &= \begin{bmatrix}\cos\theta & \sin\theta\end{bmatrix}\begin{bmatrix}\psi^{11} & 0\\ 0 & \psi^{22}\end{bmatrix}\begin{bmatrix}\cos\theta\\ \sin\theta\end{bmatrix} \\ &= \begin{bmatrix}\cos\theta & \sin\theta\end{bmatrix}\begin{bmatrix}1+\frac{4\kappa}{b} & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}\cos\theta\\ \sin\theta\end{bmatrix} \\ &= \left(1+\frac{4\kappa}{b}\right)\cos^2+\sin^2\theta = \psi^{11}\cos^2\theta+\psi^{22}\sin^2\theta, \end{aligned} \tag{4.51}$$

where $\Psi^{-1} = \mathrm{diag}(\psi^{11},\psi^{22}) = I_2 + \frac{2}{b}\mathbf{A}$ and $\mathbf{A} = \mathrm{diag}(2\kappa,0)$. Hence the angular central Gaussian density becomes

$$\begin{aligned} g_{\mathrm{ACG}}(\mathbf{x};\Psi) &\propto \frac{1}{\mathbf{x}^T\Psi^{-1}\mathbf{x}} \\ &= \frac{1}{\psi^{11}\cos^2\theta+\psi^{22}\sin^2\theta} \\ &= \frac{1}{\alpha_1(1+\cos 2\theta)+\alpha_2(1-\cos 2\theta)} \\ &= \frac{1}{(\alpha_1+\alpha_2)+(\alpha_1-\alpha_2)\cos 2\theta} \\ &= \frac{1}{(\alpha_1+\alpha_2)-(\alpha_1-\alpha_2)(\cos 2\theta-\pi)} \\ &\propto \frac{1}{1-\alpha(\cos 2\theta-\pi)}, \quad 0\le\alpha\le 1, \end{aligned} \tag{4.52}$$

where we used the double angle formulas $\sin^2\theta = (1-\cos 2\theta)/2$ and $\cos^2\theta = (1+\cos 2\theta)/2$, $\psi = 2\theta$, $\omega_1 = \psi^{11}/2 = (b+4\kappa)/2b$, $\omega_2 = \psi^{22}/2 = 1/2$ and $\alpha = \alpha_1 - \alpha_2/(\alpha_1+\alpha_2) = 2\kappa/(b+2\kappa)$. Note that the last expression of the probability density function of the angular central Gaussian distribution is

proportional to the second expression (4.49) of the doubly-wrapped Cauchy distribution with mode or mean direction at $\mu = \pi$. Further, $\alpha = 2\rho/1 + \rho^2 = 2\kappa/(b + 2\kappa)$ and solving this equality with respect to ρ yields

$$\rho(\kappa, b) = \frac{b - \sqrt{2b}}{2\kappa} = \frac{1 + \sqrt{1 + 4\kappa^2} - \Big(1 + \sqrt{1 + 4\kappa^2}\Big)^{1/2}}{2\kappa}. \tag{4.53}$$

Note that both the values of $\rho(\kappa)$ and $b(\kappa)$ in (4.53) and (4.37), respectively, are identical to the values of $\rho(\kappa)$ and $\tau(\kappa)$ for the Best-Fisher rejection algorithm to simulate random samples from the von Mises distribution with a wrapped Cauchy envelope.

We can use the strategy of simulation from the Bingham distribution with ACG envelope to link to the Best-Fisher method of simulation from the von Mises distribution with a wrapped cauchy envelope as follows. Set $q = p = 2$, $u = \mathbf{x}^T\mathbf{A}\mathbf{x} = \kappa(1 - \cos 2\theta)$ and $\Psi^{-1} = I_2 + \frac{2}{b}\mathbf{A}$, so that $\mathbf{x}^T\Psi^{-1}\mathbf{x} = 1 + \frac{2}{b}u = 1 + \frac{2\kappa}{b}\big(1 - \cos 2\theta\big)$ and finally return 2θ back to θ. Then the inequality (4.7) yields

$$-\kappa + \kappa\cos\theta \le \log\bigg[1 + \frac{2\kappa}{b}\Big(1 - \cos\theta\Big)\bigg]^{-1} + \log\Big(\frac{2}{b}\Big) - \frac{1}{2}\Big(2 - b\Big). \tag{4.54}$$

Exponentiating this inequality yields the Type II envelope

$$\exp(-\kappa)\exp\big(\kappa\cos\theta\big) \le \Big(\frac{2}{b}\Big)\ \exp\Big[-\frac{1}{2}\Big(2 - b\Big)\Big]\bigg[\frac{1}{1 + \frac{2\kappa}{b} - \frac{2\kappa}{b}\cos\theta}\bigg]. \tag{4.55}$$

Multiply both sides of the inequality by $1/2\pi I_0(\kappa)$ and $\exp(\kappa)$ where $I_0(\kappa)$ is the modified Bessel function we get

$$\frac{1}{2\pi I_0(\kappa)}\exp\big(\cos\theta\big) \le \bigg\{\frac{2\exp\big[-\frac{1}{2}\big(2 - b\big) + \kappa\big]}{b\, I_0(\kappa)}\bigg\}\frac{1}{2\pi}\bigg[\frac{1}{1 + \frac{2\kappa}{b} - \frac{2\kappa}{b}\cos\theta}\bigg]. \tag{4.56}$$

Set $\beta = 2\kappa/b$, where $\kappa \ge 0$ and $\beta \ge 0$ and multiply and divide the right hand side of the inequality (4.56) by $(1 - \beta)$ we get

$$\underbrace{\frac{1}{2\pi I_0(\kappa)}\exp\big(\cos\theta\big)}_{f(\theta;0,\kappa)} \le \underbrace{\bigg\{\frac{\big(\beta/\kappa\big)\exp\Big[\kappa(1 + \beta)/\beta - 1\Big]}{(1 - \beta)\, I_0(\kappa)}\bigg\}}_{M}\underbrace{\frac{1}{2\pi}\bigg[\frac{1 - \beta}{1 + \beta - \beta\cos\theta}\bigg]}_{g(\theta;0,\beta)}, \tag{4.57}$$

which is identical to the results of Best-Fisher method of simulation from the von Mises distribution with the first expression of a wrapped Cauchy envelope, that is,

$$\underbrace{\frac{1}{2\pi I_0(\kappa)}\exp\big(\cos\theta\big)}_{f(\theta;0,\kappa)} \ \le\ \underbrace{\frac{\big(2\rho/\kappa\big)\exp\Big[\kappa(1 + \rho^2)/2\rho - 1\Big]}{\big(1 - \rho^2\big)I_0(\kappa)}}_{M}\underbrace{\frac{1}{2\pi}\frac{1 - \rho^2}{1 + \rho^2 - 2\rho\cos\theta}}_{g(\theta;0,\rho)}. \tag{4.58}$$

4.9 Simulation from the Matrix Fisher Distribution

4.9.1 The Matrix Fisher Probability Density Function

Let $SO(3)$ denote the space of 3×3 orthogonal matrices with positive determinant. In geometric terms, $SO(3)$ consists of the proper rotations of three dimensional space.

The probability density function for the matrix Fisher distribution can be written as

$$\begin{aligned} f(\mathbf{X};\mathbf{F}) &= \left\{ {}_0F_1\left(\frac{3}{2};\frac{1}{4}\mathbf{F}^T\mathbf{F}\right)\right\}^{-1} \exp\left\{\text{trace}(\mathbf{F}X^T)\right\} \\ &= c_{\text{MF}}(\mathbf{F}) \exp\left\{\text{trace}(\mathbf{F}\mathbf{X}^T)\right\},\ X \in SO(3), \end{aligned} \tag{4.59}$$

with respect to $[d\mathbf{X}]$, the Haar measure scaled to have unit mass. Here $c_{\text{MF}}(\mathbf{F})$ is the normalizing constant, $\mathbf{F}$ is a 3×3 parameter matrix and X is a 3×3 rotation matrix so that $\mathbf{X}^T\mathbf{X} = \mathbf{X}\mathbf{X}^T = I_3$ with $|\mathbf{X}| = 1$, where I_3 is the identity matrix (Downs [16], Mardia and Jupp [59], p. 289 and Green and Mardia [28]). Moreover, ${}_0F_1$ is the hypergeometric function of matrix argument defined as

$${}_0F_1\left(\frac{3}{2};\frac{1}{4}\mathbf{F}^T\mathbf{F}\right) = \int_{SO(3)} \exp\left\{\text{trace}(\mathbf{F}\mathbf{X}^T)\right\}[d\mathbf{X}]. \tag{4.60}$$

4.9.2 Simulation Scheme

Let $\mathbf{x} \in S^3$ be distributed as Bingham with parameter matrix $\Lambda = \text{diag}(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$, $\text{B}(\Lambda)$, where $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4 = 0$ (Mardia et al. [72], p. 181 and Kume and Walker [52]), so that $\mathbf{x}$ has the density

$$\begin{aligned} g_{\text{Bing}}(\mathbf{x};\Lambda) &= {}_1F_1\left(\frac{1}{2};2;\Lambda\right)^{-1} \exp\left(-\mathbf{x}^T\Lambda\mathbf{x}\right) \\ &= {}_1F_1\left(\frac{1}{2};2;\Lambda\right)^{-1} \exp\Big(-\sum_{j=1}^{4}\lambda_j x_j^2\Big), \end{aligned} \tag{4.61}$$

with respect to $[dx]$, the surface measure on S^3 in $\mathbb{R}^4$.

There is a 2-to-1 mapping from the sphere S^3 in $\mathbb{R}^4$ to the rotation group $SO(3)$ which sends $\pm\mathbf{x} = \pm(x_1, x_2, x_3, x_4)^T$ in S^3 to

$$\mathbf{X} = \begin{pmatrix} x_1^2 + x_4^2 - x_2^2 - x_3^2 & 2(x_1x_2 - x_3x_4) & 2(x_2x_4 + x_1x_3) \\ 2(x_1x_2 + x_3x_4) & x_2^2 + x_4^2 - x_1^2 - x_3^2 & 2(x_2x_3 - x_1x_4) \\ 2(x_1x_3 - x_2x_4) & 2(x_2x_3 + x_1x_4) & x_3^2 + x_4^2 - x_1^2 - x_2^2 \end{pmatrix}, \tag{4.62}$$

a 3×3 rotation matrix with trace $4x_4^2 - 1$ and $|X| = 1$. Also $\mathbf{x}$ is uniformly distributed on a unit hemisphere in $\mathbb{R}^4$ if and only if $\mathbf{X}$ is uniformly distributed on $SO(3)$ (Prentice [80] and Wood [99]). An alternative expression for $\mathbf{X}$ is

$$\mathbf{X} = I_3 + 2x_1\mathbf{B}(\mathbf{y}) + 2\mathbf{B}(\mathbf{y})^2, \tag{4.63}$$

(Mardia and Jupp [59], pp. 285-286) where $\mathbf{y} = (x_2, x_3, x_4)^T$ and

$$\mathbf{B}(\mathbf{y}) = \begin{pmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{pmatrix}. \tag{4.64}$$

Let

$$\mathbf{F} = \mathbf{U}\mathbf{D}_\phi\mathbf{V}^T \tag{4.65}$$

be the singular value decomposition (SVD) of the parameter matrix $\mathbf{F}$ of the matrix Fisher distribution in (4.59), where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices on $SO(3)$ and $\mathbf{D}_\phi = \text{diag}(\phi_1, \phi_2, \phi_3)$ is definite matrix with $\phi_1 \leq \phi_2 \leq \phi_3$. With these parameters, the density of matrix Fisher $\mathbf{X} \sim \text{MF}(\mathbf{F})$ can be written as

$$\begin{aligned} f(\mathbf{X}|\mathbf{U}, \mathbf{D}_\phi, \mathbf{V}) &\propto \exp\left\{\text{trace}(\mathbf{V}\mathbf{D}_\phi\mathbf{U}^T\mathbf{X})\right\} \\ &= \exp\left\{\text{trace}(\mathbf{D}_\phi\mathbf{U}^T\mathbf{X}\mathbf{V})\right\} \\ &= \exp\left\{\text{trace}(\mathbf{D}_\phi\mathbf{Y})\right\}, \end{aligned} \tag{4.66}$$

where $\mathbf{Y} = \mathbf{U}^T\mathbf{X}\mathbf{V} \sim \text{MF}(\mathbf{D}_\phi)$ since the Haar measure is invariant under the rotation matrices $\mathbf{U}$ and $\mathbf{V}$. Note also that the density in (4.59) is maximized at $\mathbf{X} = \mathbf{U}\mathbf{V}^T$, which can be interpreted as the modal orientation of samples from the population. The entries of $\mathbf{D}_\phi$ can be interpreted as concentration parameters, describing how close the samples are to the mode $\mathbf{M}$, say, where $\mathbf{M}$ is the polar part of $\mathbf{F}$ (Hoff [29]). As a consequence we may, without loss of generality, take $\mathbf{U} = \mathbf{V} = \mathbf{I}_3$, where $\mathbf{I}_3$ is the 3×3 identity matrix so that $\mathbf{F} = \mathbf{D}_\phi = \text{diag}(\phi_1, \phi_2, \phi_3)$ (Wood [99]). Thus,

$$c_{\text{MF}}(\mathbf{F}; SO(3)) = c_{\text{MF}}(\mathbf{D}_\phi; SO(3)) = {}_0F_1\left(\frac{3}{2}; \frac{1}{4}\mathbf{D}_\phi^2\right). \tag{4.67}$$

Next the Bingham distribution on the unit sphere S^3 in $\mathbb{R}^4$ becomes the matrix Fisher distribution on $SO(3)$ (Prentice [80] and Wood [99]). Hence our simulated Bingham random vector via the ACG envelope can be transformed into a simulated matrix Fisher rotation matrix.

The link between parameterizations of the matrix Fisher and Bingham densities can be constructed as follows. Given $\phi_1 \leq \phi_2 \leq \phi_3$ we are looking for $c \in \mathbb{R}$ and $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4 = 0$ such that

$$\text{trace}(\mathbf{D}_\phi \mathbf{X} + c) = -\mathbf{x}^T \Lambda \mathbf{x}. \tag{4.68}$$

If so, then

$$\text{MF}(\mathbf{D}_\phi) \equiv \text{B}(\Lambda). \tag{4.69}$$

Using (4.62) we note that the left hand side in (4.68) is simplified as

$$\begin{aligned} \text{trace}(\mathbf{D}_\phi \mathbf{X} + c) &= \phi_1\big(x_1^2 + x_4^2 - x_2^2 - x_3^2 + c\big) \\ &\quad +\phi_2\big(x_2^2 + x_4^2 - x_1^2 - x_3^2 + c\big) \\ &\quad +\phi_2\big(x_3^2 + x_4^2 - x_1^2 - x_2^2 + c\big) \\ &= x_1^2\big(\phi_1 - \phi_2 - \phi_3 + c\big) + x_2^2\big(-\phi_1 + \phi_2 - \phi_3 + c\big) \\ &\quad +x_3^2\big(-\phi_1 - \phi_2 + \phi_3 + c\big) + x_4^2\big(\phi_1 + \phi_2 + \phi_3 + c\big). \end{aligned} \tag{4.70}$$

If we set $c = -\phi_1 - \phi_2 - \phi_3$, then the the left hand side in (4.68) becomes

$$\begin{aligned} \text{trace}(\mathbf{D}_\phi \mathbf{X} + c) &= -2x_1^2\big(\phi_2 + \phi_3\big) - 2x_2^2\big(\phi_1 + \phi_3\big) \\ &\quad -2x_3^2\big(\phi_1 + \phi_2\big) + x_4^2(0). \end{aligned} \tag{4.71}$$

Moreover, the right hand side in (4.68) is simplified as

$$-\mathbf{x}^T \Lambda \mathbf{x} = -\Big\{\lambda_1 x_1^2 + \lambda_2 x_2^2 + \lambda_3 x_3^2 + (0) x_4^2\Big\}. \tag{4.72}$$

Hence the relationship between the seven parameters of both distributions is stated as

$$\lambda_1 = 2(\phi_2 + \phi_3), \;\; \lambda_2 = 2(\phi_1 + \phi_3), \;\; \lambda_3 = 2(\phi_1 + \phi_2), \;\; \lambda_4 = 0. \tag{4.73}$$

Overall, the simulation procedure is summarized as follows:

(1) For a given $\phi_1 \leq \phi_2 \leq \phi_3$, obtain $\mathbf{F} = \mathbf{D}_\phi = \text{diag}(\phi_1, \phi_2, \phi_3)$.

(2) Compute λ_1, λ_2, λ_3 and $\lambda_4 = 0$ as in (4.73).

(3) Sample $\mathbf{x} = (x_1, x_2, x_3, x_4)^T$ from the real Bingham distribution with concentration parameter matrix $\mathbf{A} = \Lambda = \text{diag}(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$. This can be done by simulating from the angular central Gaussian (ACG) distribution with matrix Ψ as described in section (4.6).

(4) Evaluate $\mathbf{X}$ and the resulting $\mathbf{X}$ has the target matrix Fisher density.

Table 4.3 gives some values of the simulated efficiencies and their standard errors for matrix Fisher/Angular central Gaussian (ACG) Envelope A/R simulation with various values of ϕ_1, ϕ_2, ϕ_3, λ_1, λ_2, λ_3 and $\lambda_4 = 0$. The simulated efficiency rates under this A/R algorithm are found satisfactory for generating pseudo random sample of size $n = 1000$ from the matrix Fisher distribution with low and high concentration parameters. Under high concentrations the rejection scheme suggested gives efficiency rates close to 50%.

ϕ_1	ϕ_2	ϕ_3	λ_1	λ_2	λ_3	λ_4	Simulated Efficiency Rate	Standard Errors
1	1.5	2.0	7	6	5	0	90.1%	0.00944
1	2.0	2.5	9	7	6	0	87.4%	0.01049
1	2.5	3.0	11	8	7	0	84.3%	0.01150
10	15	20	70	60	50	0	81.1%	0.01238
10	20	25	90	70	60	0	79.5%	0.01277
10	25	30	110	80	70	0	75.3%	0.01363
50	75	80	310	260	250	0	73.1%	0.01402
50	80	90	340	280	260	0	70.4%	0.01444
50	90	100	380	300	280	0	68.8%	0.01465
100	110	120	460	440	420	0	59.2%	0.01557
100	120	130	500	460	440	0	51.7%	0.01580
100	130	150	560	500	460	0	47.6%	0.01579

Table 4.3: Simulated efficiencies rates and their standard errors for matrix Fisher/Angular central Gaussian (ACG) Envelope A/R simulation with $n = 1000$ and various values of ϕ_1, ϕ_2, ϕ_3, λ_1, λ_2, λ_3 and $\lambda_4 = 0$.

Chapter 5

General Techniques of Simulation from Directional and Shape Models

5.1 Introduction

The main goals in this chapter are (1) to review some standard existing methods for simulation from some directional and shape distributions, (2) to link and compare some new simulation methods from some directional densities to those in the last chapter and (3) to develop new envelope for an acceptance-rejection method which is both simple to program and fast for all the values of the parameters of the complex Bingham quartic (CBQ) distribution.

A developed accept-reject algorithm based on Bingham density as an envelope is given to generate samples from the von Mises distribution on the circle. Ulrich's simulation algorithm from the von Mises-Fisher distribution with an envelope proportional to Beta distribution is stated. For the circular case, a comparison is given between the efficiencies of the modified Ulrich's algorithm (Wood [100]), the Bingham envelope and that of the Best-Fisher scheme.

Two methods of simulation based on the acceptance-rejection principle are discussed for the five parameter Fisher-Bingham (FB5) distribution. A comparison is given between their efficiencies and the efficiency of the Kent and Hamelryck [45] simulation algorithm with truncated double exponential envelope.

For the shape models for two dimensional landmark data, a simulation scheme from a complex Bingham distribution (truncation to the simplex) is reviewed. Further, so far no simulation method is produced for the complex Bingham quartic (CBQ) distribution for two dimensional landmark

shaped data. In this chapter we propose an acceptance-rejection simulation algorithm from the complex Bingham quartic (CBQ) distribution. The problem of simulating from this complex shape distribution reduces to simulation from a mixture of two standard multivariate normal distributions with reasonable efficiency.

5.2 A Brief Historical Survey of Simulation Methods

This section surveys the literature relating to the historical simulation methods for some interesting directional and shape distributions.

(1) *von Mises distribution*: Best and Fisher [5] proposed a rejection scheme to generate samples from the von Mises distribution using an envelope proportional to the wrapped Cauchy distribution. It can be used efficiently for all κ. The acceptance ratio tends to unity as κ tends to zero and, as $\kappa \to \infty$, tends to 0.66. They found this to be faster than a uniform target distribution method, suggested by Seigerstetter [88]. Seigerstetter's method employs a crude envelope function which is reasonable only for $\kappa < 1$. The acceptance ratio for this method is $\exp(-\kappa)I_0(\kappa)$ which tends to zero as $\kappa \to \infty$. For $\kappa = 0$ the acceptance ration is unity, for $\kappa = 1$ it is 0.47 and for $\kappa = 10$ it is 0.13. Mardia [59](pp. 66-67) suggested an approximate method using pseudo-random variate from the wrapped normal $\mathrm{WN}(0, V)$, with mean zero and variance V chosen to be $-2\ln\big[I_1(\kappa)/I_0(\kappa)\big]$. This approximate method is good in the extreme cases as $\kappa \to 0$ and $\kappa \to \infty$ but the situation for intermediate κ values is not so clear (Best and Fisher [5]). Dagpunar [13] developed a rejection algorithm using Forsythe's method. For $\kappa \leq 0.5$, only one interval is required and the procedure was the fastest of all methods investigated, whether κ was fixed or reset between calls. When $\kappa > 0.5$, several intervals are required, necessitating numerical integrations. Although faster when κ is fixed between calls the numerical component of this algorithm makes it more difficult to implement. Yuan and Kalbeisch [101] suggested a rejection algorithm using the beta distribution as an envelope. Yuan and Kalbeisch's method requires Bessel $\big(\texttt{generate}\ X \leftarrow \texttt{Bessel}(0, \kappa)\big)$ and beta $\big(\texttt{generate}\ B \leftarrow \texttt{beta}(X + 1/2, 1/2)\big)$ random variates. The acceptance ratio for this method is $1/\big(1 + \exp(-2\kappa\sqrt{B})\big)$. The expected time of the algorithm is uniformly bounded over all choices of κ. Under high concentration this method is not efficient. Thus so far Best and Fisher's method is standard and recommended when reasonable speed, efficient and ease

of implementation are all of importance.

(2) *von Mises-Fisher distribution*: Ulrich [94] proposed a general method for simulating unit vector from rotationally symmetric unimodal distributions on the $(p-1)$-sphere such as von Mises-Fisher distribution. The basic idea is ingenious, but Wood [100] discovered empirically that Ulrich's algorithm VMF does not work correctly and it is not obvious what corrections are required to make the algorithm work. Therefore, Wood [100] suggested a modified specification to Ulrich's algorithm with high efficiency rate even under high concentration.

(3) *Bingham distribution*: Johnson [32] developed a rejection scheme to generate samples from the Bingham distribution in $\mathbb{R}^3$ using Atkinson's [2] bipartite method by dividing the range of the random variable θ into two parts and then using a different envelope for each part. This method has reasonable efficiency rate under low concentration but not under high concentration. Wood [100] developed a rejection method for simulating the Bingham distribution in $\mathbb{R}^4$ by adapting the modified specification of Ulrich's algorithm the von Mises-Fisher distribution. Kume and Walker [52] proposed a general simulation method to generate samples from the Bingham distribution on the unit sphere S^{p-1} in $\mathbb{R}^p$ based on the Gibbs sampler. If $\mathbf{x} = (x_1, x_2, \ldots, x_p)^T$ is a random vector from the Bingham distribution with parameter matrix $\mathbf{A}$ such that $\mathbf{x}^T\mathbf{x} = 1$ then $\mathbf{x}^2 = (x_1^2, x_2^2, \ldots, x_p^2)^T$ lies on a simplex. They suggested to transform $\mathbf{x}$ to variables (ω, s) and studied the marginal and conditional distributions of ω and s, where $s_i = x_i^2$ and $\omega_i = x_i/\|x_i\|$, so ω can either be 1 or -1. They started the Gibbs sampler by first sampling $s_1, s_2, \ldots, s_{p-1}$ from some suitable density function and introducing two latent variables (v, w) where the full conditional densities $f(v|w, s)$ and $f(w|v, s)$ can be sampled via uniform random variables. While straightforward, such an approach can result in a slowly mixing Markov chain because the full conditionals are so highly constrained. Therefore, Hoff [29] introduced an alternative Markov chain Monte Carlo (MCMC) version.

(4) *Fisher-Bingham distribution*: Wood [98] considered rejection procedures in the Fisher-Bingham subfamily known as FB4, FB5 and FB6 using different types of envelopes. The 4-parameter Fisher-Bingham and the 5-parameter Fisher-Bingham distributions are bounded by an envelope proportional to a mixture of two Fisher densities with a specific mixing proportion. The 6-parameter Fisher-Bingham is bounded by an envelope proportional to a mixture of two FB4 densities with a specific mixing proportion. He also suggested simulating samples from the

full Fisher-Bingham (FB8) using an envelope proportional to an appropriately chosen FB6 density. Kent and Hamelryck [45] suggested a high efficient simulation method to simulating the 5-parameter Fisher-Bingham (FB5) distribution based on a truncated double exponential envelope. Kume and Walker [53] introduced a sampling method to generate samples from the Fisher-Bingham in $\mathbb{R}^p$ by introducing two latent variables (v, w), say; they used Gibbs sampling to draw samples using the conditional distributions of these latent variables.

(5) *Matrix Fisher distribution*: The matrix Fisher distribution on $SO(3)$ is equivalent to the Bingham distribution in $\mathbb{R}^4$ on the unit sphere S^3 (Prentice [80]). Hence the rejection algorithm of Wood [100] for simulating the Bingham distribution on the unit sphere S^3 can be used to generate samples from the matrix Fisher distribution on $SO(3)$. Hoff [29] introduced general methods to generate samples from matrix Fisher density based on a uniform envelope. Green and Mardia [28] and Hoff [29] introduced general methods to generate samples from the matrix Fisher density based on Gibbs sampling. R functions to implement Hoff's MCMC method are available in the R package `rstiefel`.

(6) *Complex Bingham distribution*: Kent et al. [43] proposed a rejection scheme to generate samples from the complex Bingham distribution. They found that the problem of simulating from this distribution reduces to simulation from a truncated multivariate exponential distribution. They described three possible simulation methods namely truncation to the simplex, acceptance-rejection on the simplex and uniform on simplex and truncated Gamma on $[0, 1]$.

(7) *Complex Bingham quartic distribution*: So far no suggested simulation method is given to generate random samples from this distribution.

5.3 Simulation from the von Mises and von Mises-Fisher Distributions

5.3.1 von Mises Distribution with Real Bingham Envelope

The standard simulation method is the Best-Fisher method, based on an envelope proportional to wrapped Cauchy distribution which was discussed in the previous chapter. As another proposal we

use the real Bingham distribution as an envelope to generate samples from the von Mises distribution. A comparison between their efficiencies is also stated.

For $p = 2$, the probability density function for the Bingham distribution reduces to the doubly wrapped von Mises distribution obtaining by doubling the angles (Mardia and Jupp [70], p. 54 and p. 182) i.e. for $\theta = 2\phi$ and concentration parameter equals to $\kappa/4$; the Bingham distribution (Bingham [7]) has a doubled von Mises distribution with probability density function

$$g(\theta; 0, \kappa/4) = c_g \exp\Big\{\frac{\kappa}{4}\cos 2\theta\Big\} = \frac{1}{2\pi I_0(\kappa/4)} \exp\Big\{\frac{\kappa}{4}\cos 2\theta\Big\}, \quad \kappa > 0, \quad -\pi \leq \theta \leq \pi. \tag{5.1}$$

We can use the trigonometric double angle formula of cosine to link it to the simulation inequality $f^* \leq M^* g^*$ as follows.

$$\begin{aligned} \cos 2\theta &= 2\cos^2\theta - 1 \\ \cos^2\theta &= \left(\frac{1+\cos 2\theta}{2}\right). \end{aligned} \tag{5.2}$$

First note that $(\cos\theta - 1)^2 = \cos\theta^2 - 2\cos\theta + 1 \geq 0$ can be written as

$$\cos\theta \leq \frac{1}{2} + \frac{1}{2}\cos^2\theta. \tag{5.3}$$

Hence

$$\begin{aligned} \kappa\cos\theta &\leq \frac{\kappa}{2} + \frac{\kappa}{2}\left[\frac{1+\cos 2\theta}{2}\right] \\ \exp\big(\kappa\cos\theta\big) &\leq \exp\Big(\frac{3\kappa}{4}\Big)\exp\left(\frac{\kappa}{4}\cos 2\theta\right) \\ \frac{1}{2\pi I_0(\kappa)}\exp\big(\kappa\cos\theta\big) &\leq \frac{\exp\left(\frac{3\kappa}{4}\right)}{2\pi I_0(\kappa)} 2\pi I_0(\kappa/4).\frac{1}{2\pi I_0(\kappa/4)}\exp\Big(\frac{\kappa}{4}\cos 2\theta\Big). \end{aligned} \tag{5.4}$$

From the last result it is clear that $f(\theta)$ follows the von Mises distribution $\mathrm{VM}(\theta; \kappa, 0)$ and $g(\theta)$ reduces also to Bingham distribution which is a von Mises distribution of doubled angles $\mathrm{VM}(2\theta; \kappa/4, 0)$.

We can directly compute the starred bound M^* and the full bound M which is needed to give a theoretical assessment of the efficiency of the algorithm. Figure 5.1 plots the target probability function $f(\theta)$ of the von Mises distribution and the proportional simulation envelope of Bingham distribution $Mg(\theta)$ with $\kappa = 0.5$ and $M = 1.20$.

The expected number of iterations of the algorithm required until θ is successfully generated from g^* is exactly the bounding constant $M = \sup h(\theta) = \sup\big(f(\theta)/g(\theta)\big)$ and given by

$$M(\kappa) = \frac{I_0(\kappa/4)}{I_0(\kappa)}\exp\Big(\frac{3\kappa}{4}\Big). \tag{5.5}$$

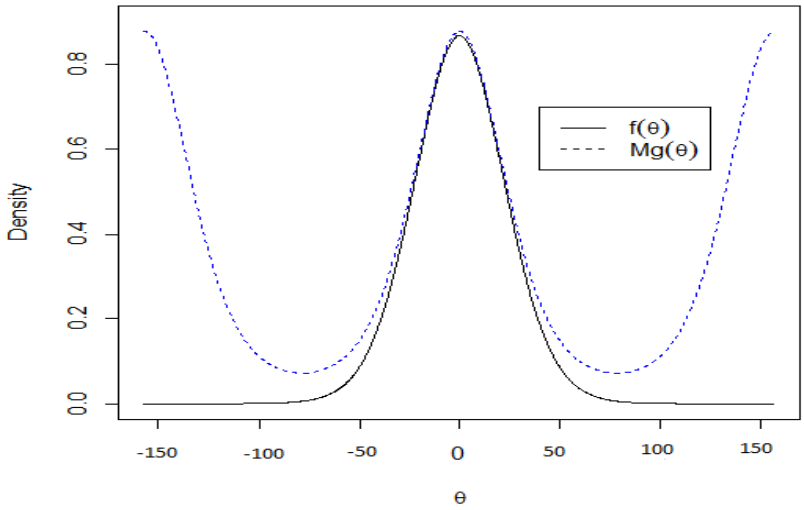

Figure 5.1: Envelope rejection for the von Mises function $f(\theta)$ with Bingham target distribution $g(\theta)$ for $\kappa = 5.0$ and $M = 1.20$.

Note that $M(0) = 1$. For large κ, one can use asymptotic expansion approximation for $I_0(\kappa)$ in (4.46) and the full bounding constant M becomes

$$M = \exp\left(\frac{3\kappa}{4}\right)\frac{\sqrt{2\pi\kappa}\exp(\kappa/4)}{\sqrt{2\pi(\kappa/4)}\exp(\kappa)} = 2. \tag{5.6}$$

For high concentration note that the mean number of trials to success this simulation is approximately $M = 2$ i.e. to produce one von Mises random variable, this A/R algorithm requires on the average 2 Bingham variables. The probability of acceptance of this simulation under Bingham candidate is given by $M^{-1} = 0.50$ i.e. the efficiency rate for this simulation is around 50%. Overall, for high concentration it is recommended to use the simulation algorithm strategy of Fisher-Best to simulate the von Mises density via a wrapped Cauchy envelope (Best and Fisher [5]).

5.3.2 von Mises-Fisher Distribution with Beta Envelope

Let $\mathbf{x}$ be a unit random vector of dimension $p \times 1$ with unit length (i.e. $\mathbf{x}^T\mathbf{x} = 1$). This p vector has von Mises-Fisher distribution with concentration parameter $\kappa \geq 0$, modal direction $\mu = (1, 0, \ldots, 0, 0)^T$ and probability density function with respect to uniform measure on a sphere $\mathbf{S}^{p-1}$

$$f(\mathbf{x}; \kappa, \mu) = \Gamma\left(\frac{p}{2}\right)\left(\frac{\kappa}{2}\right)^{1-(p/2)} I_{(p/2)-1}(\kappa)^{-1}\exp\left(\kappa\mathbf{x}^T\mu\right), \tag{5.7}$$

(see Mardia and Jupp [70], p. 168). $\kappa \geq 0$ and $\mu^T\mu = 1$. It follows from Ulrich [94] Theorem 1 that the unit p-vector $\mathbf{x}$ has a von Mises-Fisher distribution (5.7) with mean direction $\mu = (1, 0, \ldots, 0, 0)^T$ if and only if $\mathbf{x}^T = ((1-W^2)^{1/2}\mathbf{v}; W)$ where $\mathbf{v}$ is a unit $(p-1)$-vector which is uniformly distributed ($\mathbf{v} \sim U_{p-1}$), and W is a scalar random variable with probability density function

$$f(w) = c_{\mathrm{W}}(p,\kappa)^{-1}\big(1-w^2\big)^{(p-3)/2}\exp(\kappa w), \qquad -1 \leq w \leq 1, \tag{5.8}$$

where $c_{\mathrm{W}}(p,\kappa)$ is a constant of normalization and it is given by.

$$\begin{aligned} c_{\mathrm{W}}(p,\kappa) &= \int_{-1}^{1}\exp(\kappa w)\big(1-w^2\big)^{(p-3)/2}dw \\ &= \frac{\Gamma(1/2)\Gamma\big[((p-2)/2)+1/2\big]\big(k/2\big)^{(p-2)/2}}{\Gamma(1/2)\Gamma\big[((p-2)/2)+1/2\big]\big(k/2\big)^{(p-2)/2}}\int_{-1}^{1}\exp(\kappa w)\Big(1-w^2\Big)^{\left(\frac{p-2}{2}-\frac{1}{2}\right)}dw \\ &= \Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{p-1}{2}\right)\left(\frac{\kappa}{2}\right)^{1-\frac{p}{2}}I_{(p/2)-1}(\kappa) \\ &= \mathbf{B}\left(\frac{1}{2},\frac{p-1}{2}\right)\Gamma\left(\frac{p}{2}\right)\left(\frac{\kappa}{2}\right)^{1-\frac{p}{2}}I_{(p/2)-1}(\kappa), \end{aligned} \tag{5.9}$$

since

$$\mathbf{B}\left(\frac{1}{2},\frac{p-1}{2}\right) = \frac{\Gamma\big(\frac{1}{2}\big)\Gamma\big(\frac{p-1}{2}\big)}{\Gamma\big(\frac{p}{2}\big)}. \tag{5.10}$$

Ulrich [94] proposed an acceptance-rejection technique to generate random sample from von Mises-Fisher density $f(\mathbf{x};\kappa,\mu)$. The entire procedure is repeated until a variate is accepted so there is an average of M trials required for each accepted x. The following probability density function is suggested to use as an envelope to generate samples from (5.7).

$$g(x,b) = \frac{2b^{(p-1)/2}}{\mathbf{B}\Big(\frac{p-1}{2},\frac{p-1}{2}\Big)}\,\frac{(1-x^2)^{(p-3)/2}}{\Big[(1+b)-(1-b)x\Big]^{p-1}}, \quad -1 \leq x \leq 1, \quad 0 \leq b \leq 1. \tag{5.11}$$

For this envelope we want to choose the value of b to minimize the expected number of trials M (efficiency Bound). It is quite smooth to generate from $g(x,b)$ since if $Y \sim \mathrm{Beta}(\frac{1}{2}(p-1), \frac{1}{2}(p-1))$, then $X \sim g(x,b)$ if and only if $X = \big(1-(1+b)Y\big)/\big(1-(1-b)Y\big)$ (Ulrich [94]). Ulrich calculated the value of x_0 which maximizes the function $\big(1-x^2\big)^{(p-3)/2}\exp(\kappa x)$ and also the corresponding value of b_0 as

$$x_0 = \frac{-(p-1)+(4\kappa^2+(p-1)^2)^{1/2}}{2\kappa^2}, \tag{5.12}$$

and

$$b_0 = \frac{1 - x_0}{1 + x_0} = \frac{-2\kappa + (4\kappa^2 + (p-1)^2)^{1/2}}{p-1}. \tag{5.13}$$

Moreover, the optimal bound M, say, is given by

$$M_{\text{UW}} = \Gamma\left(\frac{p}{2}\right)\left(\frac{\kappa}{2}\right)^{1-\frac{p}{2}} \exp\big(\kappa x_0\big)(1 - x_0)^{(p-1)/2}. \tag{5.14}$$

Table 5.1 gives some values of the full bounds M for the three methods of simulation from von Mises distribution ($p = 2$) with various values of κ, τ, ρ (function of κ) and x_0. The acceptance ratio tends to unity as κ tends to zero, as $\kappa \to \infty$, tends around 0.66 for the Best-Fisher and Ulrich-Wood methods. These simulation schemes have high efficiencies under lower concentration since the acceptance ratio of this simulation scheme tends to unity as $\kappa \to 0$ as well as being easy to generate a realization angle θ. On the other hand, under high concentrations the efficiency of a Bingham envelope is around 50%. Overall, Best-Fisher and Ulrich-Wood methods are more reasonable than that of the Bingham envelope under high concentrations.

Wood [100] suggested the following modified accept-reject algorithm for simulation from the von Mises-Fisher distribution.

1. Calculate $b_0 = \big[-2\kappa + (4\kappa^2 + (p-1)^2)^{1/2}/p - 1\big]$,

2. Put $x_0 = (1 - b_0)/(1 + b_0)$ and $c = \kappa x_0 + (p-1)\ln(1 - x_0^2)$.

3. Generate $Z \sim \text{Beta}(\frac{p-1}{2}, \frac{p-1}{2})$, $U \sim U(0, 1)$ and calculate

$$W = \frac{1 - (1 + b_0)Z}{1 + (1 - b_0)Z}$$

4. If $\kappa W + (p-1)\ln(1 - x_0 W) - c < \ln(U)$ then go to step 3.

5. Generate an $(p-1) \times 1$ spherical uniform vector $\mathbf{v}$, and return $\mathbf{x}^T = ((1 - W^2)^{1/2}\mathbf{v}; W)$.

Then $\mathbf{x}$ has the von Mises-Fisher distribution with modal direction $(1, 0, \ldots, 0, 0)^T$ and concentration parameter $\kappa \geq 0$. In Figure(5.2) two random samples, 200 random points each, have been drawn from the von Mises-Fisher distribution with $p = 3$ around $\mu = [1\ \ 0\ \ 0]$ and $\kappa = 2, 10$ from left to right, respectively. The effect of the concentration parameter κ is clearly illustrated.

κ	τ	ρ	x_0	M_{BF}	M_{UW}	M_{Bing}	Efficiency Rate BF	Efficiency Rate UW	Efficiency Rate Bing
0.1	2.019	0.049	12.587	1.003	1.011	1.02	99.7%	98.9%	98.2%
0.2	2.077	0.097	11.527	1.009	1.019	1.04	99.1%	98.9%	96.1%
0.5	2.414	0.216	10.581	1.053	1.054	1.06	95.0%	94.9%	94.2%
1.0	3.236	0.346	9.156	1.152	1.154	1.09	86.8%	86.5%	91.6%
5.0	11.05	0.635	8.353	1.441	1.465	1.20	69.4%	68.4%	83.2%
10.0	21.02	0.727	7.142	1.481	1.490	1.88	67.5%	67.1%	53.2%
20.0	41.01	0.799	6.879	1.512	1.512	2.00	66.1%	66.1%	50.0%
50.0	101.0	0.868	5.132	1.515	1.516	2.00	66.0%	65.9%	50.0%
100.0	201.0	0.905	4.785	1.517	1.519	2.00	65.9%	65.8%	50.0%

Table 5.1: Analytical efficiencies for Von Mises/Wrapped Cauchy, Beta and Real Bingham Envelopes A/R simulation A/R simulation with various values of κ, τ, ρ and x_0.

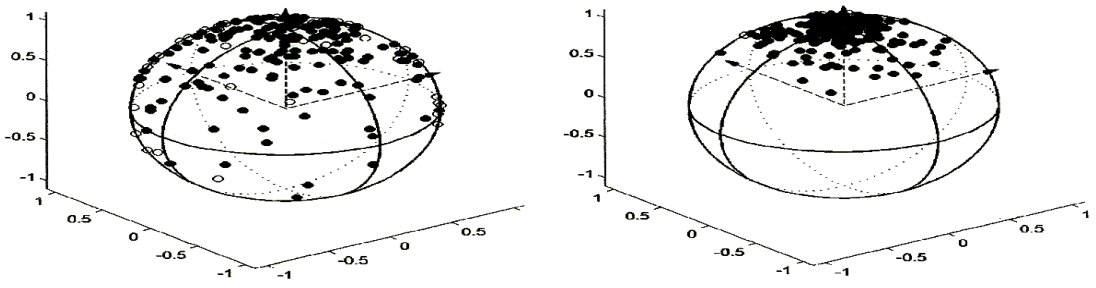

Figure 5.2: Spherical plot of two random samples with 200-points each from von Mises-Fisher distribution around $\mu = [1\ 0\ 0]$ and $\kappa = 2, 10$ from left to right, respectively.

5.4 Bipartite Rejection Scheme for the Real Bingham Distribution on the Sphere

In terms of the polar co-ordinates (θ, ϕ), consider the Bingham distribution on the unit sphere S^2 in $\mathbb{R}^3$ with probability density function

$$\begin{aligned} f_{\text{Bing}}(\theta, \phi; \kappa) &= \left[4\pi d(\kappa)\right]^{-1} \exp\Big[(\kappa_1 \cos^2\phi + \kappa_2 \sin^2\phi)\sin^2\theta\Big] \sin\theta \\ &= c_{\text{Bing}}\, f^*_{\text{Bing}}(\theta, \phi; \kappa) \qquad 0 \le \theta < \pi,\ 0 \le \phi < 2\pi, \end{aligned} \tag{5.15}$$

where $\kappa_1 \ge 0$ and $\kappa_2 \ge 0$ are concentration parameters, $\kappa = \frac{1}{2}\big(\kappa_1 + \kappa_2\big)$, the proportionality constant $d(\kappa) = {}_1F_1\Big(\frac{1}{2}; \frac{3}{2}; \kappa\Big)$, the confluent hypergeometric function with matrix argument κ and $c_{\text{Bing}} = \left[4\pi d(\kappa)\right]^{-1}$ is the normalizing constant. Note that the two angles θ and ϕ determine a point on the surface of the sphere (Bingham [7]). The confluent hypergeometric function with matrix argument κ is also known as the Kummer function $M\Big(\frac{1}{2}; \frac{3}{2}; \kappa\Big)$ which is the normalizing constant for the Watson distribution (Watson [96], Mardia and Jupp [70], p. 351 and Abramowitz and Stegun [1], p. 505).

We can apply Atkinson's bipartite rejection scheme (Atkinson [2]) by firstly dividing the range of the random variable θ into two parts. Recognize that $f_{\text{Bing}}(\pi/2 - \theta, \phi) = f_{\text{Bing}}(\pi/2 + \theta, \phi)$, so

that f is symmetric about $\theta = \pi/2$. Splitting the interval $(0, \pi/2)$ into the two intervals $(0, \pi/3]$ and $(\pi/3, \pi/2)$, the possible two envelopes (Johnson [32], p. 47) are

$$g_1(\theta, \phi; \kappa) = \exp(\kappa \sin^2 \theta) \sin(2\theta), \qquad 0 \le \theta \le \pi/3, \tag{5.16}$$

and

$$g_2(\theta, \phi; \kappa) = \exp(\kappa) \sin(\theta), \qquad \pi/3 \le \theta \le \pi/2, \tag{5.17}$$

where $\kappa = \max(\kappa_1, \kappa_2)$. Note that $f^*_{\text{Bing}}(\theta, \phi; \kappa) \le g_1(\theta, \phi; \kappa)$ and $f^*_{\text{Bing}}(\theta, \phi; \kappa) \le g_2(\theta, \phi; \kappa)$.

The function $g_1(\theta, \phi; \kappa)$ approximates $f_{\text{Bing}}(\theta, \phi; \kappa)$ and Θ can be generated by inverting the distribution function. Solving $G_1(\theta, \phi; \kappa) = U_1$ for θ yields

$$\theta = \sin^{-1} \sqrt{\kappa^{-1} \ln\big[U_1(\exp(3\kappa/4) - 1)\big]}, \tag{5.18}$$

where U_1 is uniform $(0, 1)$ (Johnson [32], p. 48). Since ϕ does not enter explicitly in either dominating function, we recognize its distribution to be uniform on $(0, 2\pi)$ and independent of θ. Hence, a variate (Θ, Φ) is generated from $g_1(\theta, \phi; \kappa)$ for Θ as in (5.18) and for Φ as $\Phi = 2\pi U_2$ where U_2 is also uniform $(0, 1)$.

The function $g_2(\theta, \phi; \kappa)$ also approximates $f_{\text{Bing}}(\theta, \phi; \kappa)$ and Θ can be generated directly from $\sin \theta$.

Atkinson's bipartite rejection scheme requires the calculation of the following quantities.

$$\Delta_1 = \int_0^{2\pi} \int_0^{\pi/3} g_1(\theta, \phi; \kappa) d\theta d\phi = \frac{2\pi\big[\exp(3\kappa/4) - 1\big]}{\kappa}, \tag{5.19}$$

$$\Delta_2 = \int_0^{2\pi} \int_0^{\pi/3} g_2(\theta, \phi; \kappa) d\theta d\phi = \pi \exp(\kappa), \tag{5.20}$$

$$S_1 = \sup_{\theta, \phi} \frac{f^*_{\text{Bing}}(\theta, \phi; \kappa)}{g_1(\theta, \phi; \kappa)} = 1 \tag{5.21}$$

$$S_2 = \sup_{\theta, \phi} \frac{f^*_{\text{Bing}}(\theta, \phi; \kappa)}{g_2(\theta, \phi; \kappa)} = \exp(-\kappa). \tag{5.22}$$

The starred bound of the algorithm is

$$M^* = S_1\Delta_1 + S_2\Delta_2 = \frac{2\pi\big[\exp(3\kappa/4)\big] + \pi(\kappa - 2)}{\kappa}, \quad \kappa > 0, \tag{5.23}$$

and the efficiency of the algorithm is

$$M^{-1} = \frac{1}{c_{\text{Bing}} M^*} = \frac{2\pi\big[\exp(3\kappa/4)\big] + \pi(\kappa - 2)}{\kappa\big[4\pi d(\kappa)\big]^{-1}}. \tag{5.24}$$

Table 5.2 gives some values of the full bound M with various values of $\kappa = \max(\kappa_1, \kappa_2)$. We use the R `kummerM` function in the `fAsianOptions` package to calculate the confluent hypergeometric function of the 1st kind with matrix argument κ, $d(\kappa) = {}_1F_1\Big(\frac{1}{2}; \frac{3}{2}; \kappa\Big)$. It is clear that Atkinson's bipartite rejection scheme is efficient under low concentrations but inefficient as κ tends to infinity.

κ	M	Efficiency Rate
0.1	1.14	87.9%
0.2	1.17	85.5%
0.3	1.24	80.6%
0.4	1.31	76.7%
0.5	1.56	64.1%
1.0	1.79	55.9%
5.0	2.09	47.9%
10.0	3.14	31.8%
100.0	4.88	20.5%

Table 5.2: Analytical efficiencies for Bingham/Bipartite Envelopes A/R simulation with various values of κ.

5.5 Simulation from the Fisher Distribution with Bingham Envelope

The Fisher density on the unit sphere in $\mathbb{R}^3$ takes the standardised form

$$\begin{aligned} f(\theta, \phi) &= \frac{\kappa}{4\pi \sinh \kappa} \exp\big\{\kappa \cos\theta\big\}, \;\; 0 \le \theta \le \pi, \;\; 0 \le \phi < 2\pi \\ &= \frac{\kappa}{2\pi(e^{\kappa} - e^{-\kappa})} \exp\big\{\kappa \cos\theta\big\} \\ &= \frac{\kappa}{2\pi(1 - e^{-2\kappa})} \exp\big\{\kappa(\cos\theta - 1)\big\} = c_{\text{F}}\, f^*(\theta, \phi), \end{aligned} \tag{5.25}$$

with respect to $\sin\theta d\theta d\phi$. Here $\kappa > 0$ is the concentration parameter (Mardia and Jupp [70], p. 170 and Fisher et. al [23], p.86-87) and $c_{\mathrm{F}} = \kappa/2\pi(1 - e^{-2\kappa})$ is the normalizing constant [1].

A possible type of envelope, $g(\theta, \phi)$, is the real Bingham distribution on the unit sphere in $\mathbb{R}^3$ with concentration matrix A, that is,

$$\begin{aligned} g(\theta,\phi) &= \frac{1}{4\pi\ {}_1F_1\Big(\frac{1}{2};\frac{3}{2};\kappa\Big)} \exp\{\mathbf{x}^T\ \mathbf{A}\ \mathbf{x}\} \\ &= \frac{1}{4\pi\ {}_1F_1\Big(\frac{1}{2};\frac{3}{2};\kappa\Big)} \exp\Big[\frac{\kappa}{2}(\cos^2\theta)\Big] \\ &= \frac{e^{\kappa/2}}{4\pi\ {}_1F_1\Big(\frac{1}{2};\frac{3}{2};\kappa\Big)} \exp\Big[\frac{\kappa}{2}(\cos^2\theta - 1)\Big] = c_{\mathrm{Bing}}\ g^*(\theta,\phi), \end{aligned} \tag{5.26}$$

with respect to $\sin\theta d\theta d\phi$. Here ${}_1F_1\Big(\frac{1}{2};\frac{3}{2};\kappa\Big)$ is the confluent hypergeometric function with matrix argument κ, $c_{\mathrm{Bing.}}$ is the normalizing constant. Note that for a 3×3 symmetric matrix $\mathbf{A}$ in $\mathbb{R}^3$, the exponent term for the Bingham density in the polar co-ordinates (θ, ϕ), becomes

$$\mathbf{x}^T\ \mathbf{A}\ \mathbf{x} = \begin{bmatrix}\cos\theta & \sin\theta\cos\phi & \sin\theta\sin\phi\end{bmatrix} \begin{bmatrix} \frac{\kappa}{2} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \cos\theta \\ \sin\theta\cos\phi \\ \sin\theta\sin\phi \end{bmatrix} = \frac{\kappa}{2}\cos^2\theta. \tag{5.27}$$

Note that

$$2(\cos\theta - 1) \le \cos^2\theta - 1 \tag{5.28}$$

holds, hence

$$\begin{aligned} \kappa(\cos\theta - 1) &\le \frac{\kappa}{2}\big(\cos^2\theta - 1\big) \\ \underbrace{\frac{\kappa}{2\pi(1-e^{-2\kappa})}\exp\Big[\kappa(\cos\theta-1)\Big]}_{f(\theta,\phi)} &\le \frac{2\kappa\ {}_1F_1\Big(\frac{1}{2};\frac{3}{2};\kappa\Big)}{(1-e^{-2\kappa})} \underbrace{\frac{e^{\kappa/2}}{4\pi\ {}_1F_1\Big(\frac{1}{2};\frac{3}{2};\kappa\Big)}\exp\Big[\frac{\kappa}{2}\big(\cos^2\theta-1\big)\Big]}_{g(\theta,\phi)}. \end{aligned} \tag{5.29}$$

The last inequality satisfies the main simulation scheme as describes in section (4.2). The expected number of iterations of the algorithm required until θ and ϕ are successfully generated from g is

[1] A direct approximation is possible for the normalizing constant of the Fisher distribution in (5.25) using the standard second order saddlepoint approximations.

exactly the bounding constant M where

$$M = \frac{2\kappa \, {}_1F_1\left(\frac{1}{2}; \frac{3}{2}; \kappa\right)}{(1 - e^{-2\kappa})}. \tag{5.30}$$

Note that, when $p = 3$ and if we use spherical polar coordinates then θ and ϕ are independent and ϕ is uniform on the unit circle (Mardia and Jupp [70], p. 170). Moreover, under high concentrations the efficiency rate with a Bingham envelope is around 50% and Ulrich-Wood method with a beta envelope is more reasonable than that of Bingham envelope.

5.6 Simulation from the Fisher-Bingham (FB5) Distribution

In this section we consider two methods to sample from the FB5 distribution. One of interest is using the uniform envelope and another with an envelope proportional to the real Bingham distribution. Their A/R efficiencies are compared. Kent and Hamelryck [45] suggested simulation approach based on truncated double exponential distribution.

5.6.1 Background

The 5-parameter Fisher-Bingham FB5 distribution, otherwise known as the Kent distribution, belongs to the family of spherical distributions in directional statistics. In particular the FB5 serves as an extension of the von Mises-Fisher (VMF) distribution. The FB5 is more flexible since it has oval density contours as opposed to the VMF's circular contours. Therefore the Fisher distribution will not succeed in describing data that have originated from oval density contours. However, the FB5 distribution will be capable of describing data that have originated from a distribution with circular density contours i.e. the VMF distribution (Mammasis and Stewart [57]).

The five parameter Fisher-Bingham (FB5) distribution is defined by the probability density function

$$f(\mathbf{x}; \Theta) = c(\kappa, \beta)^{-1} \exp\{\kappa \gamma_1^T \mathbf{x} + \beta[(\gamma_2^T \mathbf{x})^2 - (\gamma_3^T \mathbf{x})^2]\}, \quad \mathbf{x} \in \Omega_3 \tag{5.31}$$

where $\Omega_3 = \{\mathbf{x} \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 = 1\}$ denotes the unit sphere in $\mathbb{R}^3$ and $c(\kappa, \beta)$ is the normalizing constant. We use the notation FB5(κ, β, Γ) to define the distribution. Here $\Theta = (\kappa, \beta, \gamma_1, \gamma_2, \gamma_3)^T$

is the parameter vector. The parameters can be interpreted as follows: $\kappa \geq 0$ represents the concentration, β with $0 \leq 2\beta < \kappa$ determines the ovalness or the ellipticity of the contours of equal probability and a (3×3) orthogonal matrix $\Gamma = (\gamma_1, \gamma_2, \gamma_3)^T$. The parameter γ_1 is the vector of the directional cosines that define the mean direction or pole or centre of the distribution. The parameters γ_2 and γ_3 relate to the orientation of the distribution. If we visualize the elliptical contours of the distribution on the surface of the sphere, then γ_2 and γ_3 define the directions of the major and minor axes, respectively, of the ellipses. Also, if $\beta = 0$ then (5.31) reduces to a Fisher density (Kent [40]).

In order to sample from the FB5 distribution we can without loss of generality consider the density rotated to the standard frame of reference (Kent [40]). In terms of the polar coordinates (θ, ϕ) defined by

$$x_1 = \cos\theta, \qquad x_2 = \sin\theta\cos\phi, \qquad x_3 = \sin\theta\sin\phi, \tag{5.32}$$

and the probability density function of the FB5 distribution takes the form

$$\begin{aligned} f(\theta,\phi) &= c(\kappa,\beta)^{-1}\exp\left\{\kappa\cos\theta + \beta\sin^2\theta\cos 2\phi\right\}\sin\theta \\ &= c(\kappa,\beta)^{-1}\exp\left\{\kappa\cos\theta + \beta\sin^2\theta[\cos^2\phi - \sin^2\phi]\right\}\sin\theta, \end{aligned} \tag{5.33}$$

with respect to $d\theta d\phi$ where $0 \leq \theta \leq \pi$, $0 \leq \phi < 2\pi$.

5.6.2 FB5 Distribution with Uniform Envelope

Let $f(\theta, \phi)$ take the form

$$f(\theta,\phi) = c_f f^*(\theta,\phi), \tag{5.34}$$

where

$$f^*(\theta,\phi) = \exp\left\{\kappa\cos\theta + \beta\sin^2\theta\cos 2\phi\right\}\sin\theta. \tag{5.35}$$

A possible choice of an envelope, $g(\theta, \phi)$, is the uniform density on the sphere,

$$\begin{aligned} g(\theta,\phi) &= c_g\, g^*(\theta,\phi) \\ &= \frac{\Gamma\left(\frac{3}{2}\right)}{2\pi^{3/2}}\sin\theta = \frac{1}{4\pi}\sin\theta, \end{aligned} \tag{5.36}$$

where $g^*(\theta,\phi) = \sin\theta$ and $\Gamma(n+\frac{1}{2}) = (2n-1)!\,2^{-n}\sqrt{\pi}$ (Abramowitz and Stegun [1]). The expected number of iterations of the algorithm required until θ and ϕ are successfully generated from $g(\theta,\phi)$ is exactly the bounding constant $M = \sup h(\theta,\phi)$. Note that

$$h^*(\theta,\phi) = \frac{f^*(\theta,\phi)}{g^*(\theta,\phi)} = \exp\big\{\kappa\cos\theta + \beta\sin^2\theta\cos 2\phi\big\}, \tag{5.37}$$

and we compute its maximum; which must occur at those values of θ and ϕ which maximize the exponent in (5.37). Then the starred bound M^* is given by

$$M^* = \sup h^*(\theta,\phi) = \exp\big(\kappa+\beta\big), \tag{5.38}$$

and the full bound M is

$$\begin{aligned}
M &= \frac{c_f M^*}{c_g} \\
&= \frac{c(\kappa,\beta)^{-1}M^*}{(4\pi)^{-1}} \\
&= \frac{4\pi\exp\big(\kappa+\beta\big)}{c(\kappa,\beta)} \\
&= \frac{4\pi\exp\big(\kappa+\beta\big)}{2\pi\sum_{j=1}^{\infty}\frac{\Gamma(j+\frac{1}{2})}{\Gamma(j+1)}\beta^{2j}\big(\frac{1}{2}\kappa\big)^{-2j-\frac{1}{2}}I_{2j+\frac{1}{2}}(\kappa)} \\
&\approx \frac{4\pi\exp\big(\kappa+\beta\big)}{2\pi e^{\kappa}[(\kappa-2\beta)(\kappa+2\beta)]^{-1/2}}, \qquad \text{(large } \kappa \text{ with } 2\beta/\kappa < 1 \text{ fixed)} \\
&= 2\Big[e^{\beta}(\kappa-2\beta)(\kappa+2\beta)\Big]^{1/2},
\end{aligned} \tag{5.39}$$

(see, for example, Kent [40], p. 73 and Kent et al. [46]). Note that M depends upon the normalizing constant of the five parameter Fisher-Bingham (FB5) distribution. Note also that $c(0,0) = 4\pi$, the surface area of the sphere, $M = 1$ and $c(\kappa,0) = 4\pi\kappa^{-1}\sinh\kappa$, the normalizing constant for the Fisher distribution.

5.6.3 FB5 Distribution with Truncated Exponential Envelope

In this subsection we consider the Kent-Hamelryck method to sample from the FB5 distribution otherwise known as the Kent distribution (Kent and Hamelryck [45]). For the purpose of simulation it is helpful to use an equal area projection. Set

$$x_1 = r\cos\theta, \qquad x_2 = r\sin\theta, \quad r = \sin(\theta/2), \tag{5.40}$$

so that $(2x_1, 2x_2)$ represents an equal-area projection of the sphere.

In (x_1, x_2) coordinates, the Jacobian factor $\sin\theta$ disappears and the pdf (with respect to $dx_1 dx_2$ in the unit disk $x_1^2 + x_2^2 < 1$) takes the form

$$\begin{aligned} f(x_1, x_2) &\propto \exp\big\{ -2\kappa r^2 + 4\beta(r^2 - r^4)(\cos^2\phi - \sin^2\phi)\big\} \\ &= \exp\big\{ -2\kappa(x_1^2 + x_2^2) + 4\beta[1 - (x_1^2 + x_2^2)(x_1^2 - x_2^2)]\big\} \\ &= \exp\big\{ -\frac{1}{2}[ax_1^2 + bx_2^2 + \gamma(x_1^4 - x_2^4)]\big\}, \end{aligned} \tag{5.41}$$

where the new parameters

$$a = (4\kappa - 8\beta), \quad b = (4\kappa + 8\beta), \quad \gamma = 8\beta, \tag{5.42}$$

satisfy $0 \le a \le b$ and $\gamma \le b/2$. Here we have used the double angle formulas, $\cos\theta = 1 - 2\sin^2(\theta/2)$, $\sin\theta = 2\sin(\theta/2)\cos(\theta/2)$.

Note that the pdf splits into a product of a function of x_1 alone and x_2 alone. Hence x_1 and x_2 would be independent except for the constraint $x_1^2 + x_2^2 < 1$. The Kent-Hamelryck method of simulation, as sketched below, will be to simulate $|x_1|$ and $|x_2|$ separately by acceptance-rejection using a truncated exponential envelope $g(x)$,

$$g(x) = \frac{\alpha \exp(\alpha x)}{1 - \exp(-\alpha)}, \quad 0 \le x \le 1, \quad \alpha > 0, \tag{5.43}$$

and then additionally to reject any values lying outside the unit disk. The starting point of the Kent-Hamelryck's simulation method is the inequality

$$\frac{1}{2}(\sigma|w| - \tau)^2 \ge 0 \tag{5.44}$$

for any parameters σ, $\tau > 0$ and for all w, hence

$$-\frac{1}{2}\sigma^2 w^2 \le \frac{1}{2}\tau^2 - \sigma\tau|w|. \tag{5.45}$$

After exponentiation, this inequality provides the basis for simulating a Gaussian random variable from a double exponential random variable by acceptance-rejection criteria. For x_1 we need to apply (5.45) twice, first with $\sigma = \gamma^{1/2}$, $\tau = 1$ and $w = x_1^2$, and the second with $\sigma = (a + 2\gamma^{1/2})^{1/2}$, $\tau = 1$ and $w = x_1$, to get

$$\begin{aligned} -\frac{1}{2}(ax_1^2 + \gamma x_1^4) &\le \frac{1}{2} - \frac{1}{2}(a + 2\gamma^{1/2})x_1^2 && (5.46) \\ &\le c_1 - \lambda_1|x_1|, && (5.47) \end{aligned}$$

where $c_1 = 1$ and $\lambda_1 = (a + 2\gamma^{1/2})^{1/2}$. Again, to develop a suitable envelope for x_2 recall that $0 \leq 2\gamma \leq b$. To begin with suppose $b > 0$. From (5.46) with $\sigma = (b-\gamma)^{1/2}$, $\tau = (b/(b-\gamma))^{1/2}$ and $w = x_2^2$,

$$-\frac{1}{2}(bx_2^2 + \gamma x_2^4) \leq -\frac{1}{2}(a-\gamma)x_2^2 \leq c_2 - \lambda_2|x_2|, \tag{5.48}$$

where $c_2 = b/\{2(b-\gamma)\} \leq 1$ and $\lambda_2 = b^{1/2}$. If $b = 0$ and so $\gamma = 0$ then (5.48) continues to hold with $\lambda_2 = 0$ and $c_2 = 0$.

For $\beta = 0$ the Fisher-Bingham (FB5) distribution reduces to the Fisher distribution (Kent [40]). Table 5.3 gives numerical results for the analytical simulation bounds of the three methods of simulation from the Fisher-Bingham (FB5) distribution with various values of κ and $\beta < \kappa/2$. Under low concentrations all the three methods are efficient. On the other hand, under high concentration the rejection simulation method of uniform envelope is not efficient. The real Bingham envelope has a good efficiency bound close to 1.00 as $\kappa \to 0$ and it is expected to has an efficiency bound close to 2.00 as $\kappa \to \infty$. The Kent-Hamelryck method is more reasonable than both other simulation methodologies in which the efficiencies are high for any range of various values of κ and β. This simulation scheme has high efficiency either under low or high concentrations since the acceptance ratio of this simulation scheme tends to unity as $\kappa \to 0$ as well as it is easy to generate the realization angles θ and ϕ.

5.7 Simulation from the Complex Bingham Distribution

5.7.1 The Complex Bingham Density Function

First of all we consider the case where we have the probability distribution on the pre-shape sphere $S^{m(k-1)-1}$, corresponding to k landmarks in m dimensions $(k-1)$. For the $m = 2$ dimensional case and using complex notation we have seen that $S^{2k-3} \equiv \mathbb{C}S^{k-2}$. The pre-shape $z = (z_1, \ldots, z_{k-1})^T$ lies on the unit complex sphere $\mathbb{C}S^{k-2}$, where $\mathbf{z}^*\mathbf{z} = \sum_{j=1}^{k-1} |z_j|^2 = 1$. One way of constructing an appropriate distribution is by conditioning the complex multivariate normal distribution with probability density function proportional to $\exp(-\frac{1}{2}\mathbf{z}^*\Sigma^{-1}\mathbf{z})$ where Σ is Hermitian (i.e. $\Sigma = \Sigma^*$). Conditioning it on $\mathbf{z}^*\mathbf{z} = 1$ gives rise to the following complex Bingham distribution (Dryden and Mardia [18], pp. 111-112). The complex Bingham distribution with canonical parameter matrix A

κ	β	M_{KH}	M_U	M_{Bing}
0.10	0.00	1.01	1.09	1.05
0.25	0.10	1.04	1.34	1.09
0.30	0.12	1.05	1.57	1.15
0.60	0.24	1.12	2.36	1.36
1.00	0.40	1.15	4.09	1.45
1.50	0.60	1.18	8.19	1.67
3.00	1.20	1.21	68.7	1.82
10	4.00	1.28	103.9	1.94
20	8.00	1.35	273.1	2.02

Table 5.3: Analytical efficiencies for the FB5 distribution based on the Kent-Hamelryck, the uniform and the real Bingham envelopes A/R simulation with various values of κ and β ($\kappa > 2\beta$).

has probability density function

$$f(\mathbf{z}) = c(\mathbf{A})^{-1} \exp(\mathbf{z}^* \mathbf{A} \mathbf{z}), \qquad z \in \mathbb{C}S^{k-2}, \tag{5.49}$$

where the $(k-1) \times (k-1)$ matrix $\mathbf{A}$ is Hermitian and $c(\mathbf{A})$ is the normalizing constant. We write

$$\mathbf{z} \sim \mathbb{C}B_{k-2}(\mathbf{A}). \tag{5.50}$$

A complex random vector $\mathbf{z}$ is said to have complex symmetry if its distribution is invariant under scalar rotation, so that $\mathbf{z}$ and $\exp(i\theta)\mathbf{z}$ have the same distribution for all θ i.e. $f(\mathbf{z}) = f(e^{i\theta}\mathbf{z})$, the complex Bingham distribution has this property. This property therefore makes the distribution suitable for statistical shape analysis (location and scale were previously removed because z is on the pre-shape sphere). The complex Bingham distribution provides a very elegant framework for the analysis of two dimensional shape data (see Kendall, [42]; Kent, [41]).

Since $\mathbf{z}^*\mathbf{z} = 1$ for $\mathbf{z}$ in $\mathbb{C}S^{k-2}$, the parameter matrices $\mathbf{A}$ and $\mathbf{A} + \alpha I$ define the same complex Bingham distribution and $c(\mathbf{A} + \alpha I) = c(\mathbf{A}) \exp \alpha$ for any complex number α. It is convenient to remove this non-identifiability by setting $\lambda_{\max}(\mathbf{A}) = 0$, where $\lambda_{\max}(A)$ denotes the largest eigenvalue of $\mathbf{A}$. Hence, without loss of generality, we may shift the eigenvalues of $\mathbf{A}$ so that they are nonpositive with the largest one equalling 0. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{k-1} = 0$ denote the eigenvalues of $-\mathbf{A}$. Denote

the corresponding standardized eigenvectors by the columns of the unitary matrix $\Gamma = (\gamma_1, \ldots \gamma_{k-1})$, with $\Gamma^*\Gamma = I$. Write $\lambda = (\lambda_1, \ldots, \lambda_{k-2})$ for the vectors of the first $k-2$ eigenvalues. The $\{\lambda_j\}$ can be thought of as concentration parameters.

5.7.2 Simulation Schemes

In order to simulate from $\mathbb{C}B_{k-2}(\mathbf{A})$, it is convenient to rotate to principal axes. Kent [41] proposed some non-standard polar coordinates on the pre-shape sphere. Given a point $(z_1, \ldots, z_{k-1})^T$ on $\mathbb{C}S^{p-1}$, $\sum_{j=1}^{k-1} |z_j|^2 = 1$, we transform to $(s_1, s_2, \ldots, s_{k-2}, \theta_1, \theta_2, \ldots, \theta_{k-1})$, where

$$\mathrm{Re}(z_j) = s_j^{1/2} \cos\theta_j, \quad \mathrm{Im}(z_j) = s_j^{1/2} \sin\theta_j. \tag{5.51}$$

Hence the components $\mathbf{z} = (z_1, z_2, \ldots, z_{k-1})^T$ can be expressed as

$$\begin{aligned} z_j &= s_j^{1/2} \cos\theta_j + i\, s_j^{1/2} \sin\theta_j \\ &= s_j^{1/2}[\cos\theta_j + i \sin\theta_j] \\ &= s_j^{1/2} e^{i\theta_j}, \end{aligned} \tag{5.52}$$

for $j = 1, 2, \ldots, k-1$, $s_j = |z_j|^2 \geq 0$, $0 \leq \theta_j = \arg(z_j) < 2\pi$ and $s_{k-1} = 1 - \sum_{j=1}^{k-2} s_j$. The coordinates $s_1, s_2, \ldots, s_{k-2}$ are on the $k-2$ dimensional unit simplex, $\mathcal{S}_{k-2}$, where

$$\mathcal{S}_{k-2} = \Big\{ s = (s_1, s_2, \ldots, s_{k-2})^T \in \mathbb{R}^{k-2} : \quad \sum_{j=1}^{k-2} s_j \leq 1 \Big\}. \tag{5.53}$$

Under the complex Bingham distribution $\theta_1, \ldots, \theta_{k-1}$ are uniformly distributed on $[0, 2\pi)$ independently of one another and of $s = (s_1, \ldots, s_{k-2})$, which has a truncated multivariate exponential distribution (Kent et al. [43]) with probability density function

$$f(s) = d(\lambda)^{-1} \exp\Big(\sum_{j=1}^{k-2} -\lambda_j s_j \Big), \quad s \in \mathcal{S}_{k-2}, \tag{5.54}$$

where the normalizing constants in (5.49) and (5.54) is related by $c(\mathbf{A}) = 2\pi^{k-1} d(\lambda)$.

Initially, $(k-2)$ truncated exponentials are generated subject to a linear constraint, and then these random variables are expressed in polar coordinates to deliver a complex Bingham distribution. Consider the continuous random variable X has exponential distribution truncated to the interval $[0, 1]$ and denoted by $\mathrm{TExp}(\lambda)$ with probability density function

$$f(x) = \lambda \exp(\lambda x)\big(1 - \exp(-\lambda)\big)^{-1}, \quad 0 \leq x \leq 1. \tag{5.55}$$

Here λ is the rate parameter $\lambda > 0$ and its cumulative distribution function is given by

$$F(x) = \big(1 - \exp(-\lambda x)\big)\big(1 - \exp(-\lambda)\big)^{-1}. \tag{5.56}$$

One way to do the simulation from TExp(λ) is by the inversion method. If $U \sim U[0,1]$, then

$$X = F^{-1}(U) = -(1/\lambda)\log\big(1 - U(1 - e^{-\lambda}\big), \tag{5.57}$$

has truncated exponential distribution and we can use the following algorithm.

Step 1 Simulation of TExp(λ)

(1) Simulate uniform random variable $U[0,1]$.

(2) Calculate $X = -(1/\lambda)\log\big(1 - U(1 - e^{-\lambda}\big)$.

The method for simulating the complex Bingham distribution uses $(p-1)$ truncated exponentials to generate a p vector with a complex Bingham distribution. We can use the following algorithm.

Step 2 Simulation of the complex Bingham distribution $\mathbb{C}B_{p-1}(A)$

(1) Generate $S = (S_1, \ldots, S_{k-2})$ where where $S_j \sim \text{TExp}(\lambda_j), j = 1, 2, \ldots, k-2$ are independent random variables simulated using Algorithm 1 with $S_j = -(1/\lambda_j)\log\big(1 - U(1 - e^{-\lambda_j}\big)$.

(2) If $\sum_{j=1}^{k-2} S_j < 1$, write $S_{k-1} = 1 - \sum_{j=1}^{k-2} S_j$. Otherwise, return to step 1.

(3) Generate independent angles $\theta_j \sim U[0, 2\pi), j = 1, 2, \ldots, k-1$.

(4) Calculate $Z_j = S_j^{1/2} e^{i\theta_j}, j = 1, 2, \ldots, k-1$.

The algorithm delivers a $(k-1)$ vector $(Z_1, Z_2, \ldots, Z_{k-1})^T$; which has a complex Bingham distribution. Note that $(S_j^{1/2}; \theta_j)$ are essentially polar coordinates for complex number Z_j. If we define the truncation probability $p_T = \int_X g(x)dx$, then the number of iteration N_T, say, of the algorithm required until S and θ are successfully generated from TExp and $U[0, 2\pi)$ is exactly the boundary constant (efficiency) and has a geometric distribution with parameter p_T. That is, $P(N_T = p) = q_T^{p-1} p_T, 1 \leq k-1 \leq \infty$, with $E(N_T) = p_T^{-1}$. For all the eigenvalues of $\mathbf{A}$ are equal

$\lambda_1 = \ldots = \lambda_{k-2} = \lambda$ and $\lambda_{k-1} = 0$, Kent et al. [43] calculated the probability of acceptance under this simulation scheme as

$$p_T = \left\{ e^{\lambda} - \sum_{j=0}^{k-3} \frac{\lambda^j}{j!} \right\} e^{-\lambda}. \tag{5.58}$$

In each cycle the number of uniform random variables used is $p-1$. Thus if we let M be the number of uniform random variable needed in this algorithm, then

$$M = (k-2)p_T^{-1} = (k-2)e^{\lambda} \left\{ e^{\lambda} - \sum_{j=0}^{k-3} \frac{\lambda^j}{j!} \right\}^{-1}. \tag{5.59}$$

Further, M is bounded by $(k-1) \times (k-1)$ as $\lambda \to \infty$. Kent et al. [43] suggested also two other simulation methods from the complex Bingham distribution. One of them is an acceptance-rejection on the simplex and the other is uniform on simplex and truncated gamma on $[0, 1]$. Generally, the simulation scheme described in this section (truncation to the simplex) will be preferred for large concentrations and the acceptance-rejection on the simplex for small concentrations.

The $(k-2)$ dimensional complex Bingham distribution can be regarded as a special case of a $(2k-3)$ dimensional real Bingham distribution (Dryden and Mardia [18], p. 113). So, we can use the angular central Gaussian distribution as an envelope to generate random samples from the complex Bingham distribution and compare the efficiency of the new method of section (4.6) with that of a truncated multivariate exponential envelope. Table 5.4 gives some numerical values for the simulated bound M of the rejection simulation method of Kent et al. [43] and for the rejection scheme of the angular central Gaussian (ACG) envelope with k varying and with a common concentration value λ. It is clear from the table that the rejection scheme of a truncated multivariate exponential envelope is more efficient than that of the ACG envelope with various values of k and with a common concentration value λ.

5.8 Simulation from the Complex Bingham Quartic Distribution

Let $\mathbb{C}S^{k-1} = \{\mathbf{z} = (z_1, z_2, \ldots, z_{k-1})^T : \sum_{j=1}^{k-1} |z_j|^2 = 1\} \subset \mathbb{C}^{k-1}$ denote the unit complex sphere in $\mathbb{C}^{k-1}$. The complex Bingham quartic (CBQ) density centred at the north pole $(0, 0, \ldots, 1)^T$ on $\mathbb{C}S^{k-2}$ with concentration symmetric matrix Ω can be written in the form

$$f_{\mathrm{CBQ}}(\mathbf{z}) = c_{\mathrm{CBQ}}(\Omega)^{-1} \exp\Big\{ -\frac{1}{2} \big(\mathbf{x}^T \Omega \mathbf{x} - (\mathbf{x}^T \mathbf{x}) \mathbf{x}^T \Omega^{(\mathrm{as})} \mathbf{x} \big) \Big\}, \quad 0 \le |\mathbf{x}| \le 1. \tag{5.60}$$

Truncated Multivariate Exponential Envelope						
k	$\lambda = 0.01$	$\lambda = 0.1$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 10$	$\lambda = 20$
3	1.12	1.14	1.17	1.21	1.32	1.36
4	1.16	1.18	1.28	1.34	1.59	1.68
5	1.27	1.38	1.46	1.57	1.85	1.97
Angular Central Gaussian (ACG) Envelope						
k	$\lambda = 0.01$	$\lambda = 0.1$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 10$	$\lambda = 20$
3	1.17	1.19	1.22	1.25	1.41	1.66
4	1.18	1.23	1.39	1.44	1.71	1.91
5	1.23	1.43	1.58	1.74	1.98	2.11

Table 5.4: Simulated efficiencies M of truncated multivariate exponential envelope and angular central Gaussian envelope needed for the simulation methods from the complex Bingham distribution with k varying and with a common concentration value λ.

where Ω is positive semidefinite (Kent et al. [46]). The complex symmetric and anti-complex symmetric matrices are defined in (3.35) and (3.36), respectively, such that $\Omega = \Omega^{(\mathrm{cs})} + \Omega^{(\mathrm{as})}$.

Let

$$\Omega^{\perp} = \Omega^{(\mathrm{cs})} - \Omega^{(\mathrm{as})},\ \Psi(s) = -\frac{1}{2}(s\Omega - s^2\Omega^{(\mathrm{as})}) \quad \text{and} \quad \Psi^{\perp}(s) = -\frac{1}{2}(s\Omega + s^2\Omega^{(\mathrm{as})}), \tag{5.61}$$

where $\Omega^{\perp}$ is also positive semidefinite, $|\Omega| = |\Omega^{\perp}|$, $\Psi(s)$ and $\Psi^{\perp}(s)$ are also real symmetric $(2k-4)\times(2k-4)$ matrices with $0 < s < 1$. Then, the probability density function for the complex Bingham quartic distribution becomes proportional to

$$\begin{aligned}
f^{*}_{\mathrm{CBQ}}(\mathbf{x}) &= \exp\left\{-\frac{1}{2}\left[\mathbf{x}^T\Omega\mathbf{x} - (\mathbf{x}^T\mathbf{x})\mathbf{x}^T\Omega^{(\mathrm{as})}\mathbf{x}\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\mathbf{x}^T(\Omega^{(\mathrm{cs})} + \Omega^{(\mathrm{as})})\mathbf{x} - (\mathbf{x}^T\mathbf{x})\mathbf{x}^T\Omega^{(\mathrm{as})}\mathbf{x}\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\mathbf{x}^T\Omega^{(\mathrm{cs})}\mathbf{x} + \mathbf{x}^T\Omega^{(\mathrm{as})}\mathbf{x} - (\mathbf{x}^T\mathbf{x})\mathbf{x}^T\Omega^{(\mathrm{as})}\mathbf{x}\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\mathbf{x}^T\Omega^{(\mathrm{cs})}\mathbf{x} + (1 - \mathbf{x}^T\mathbf{x})\mathbf{x}^T\Omega^{(\mathrm{as})}\mathbf{x}\right]\right\}, \quad \|\mathbf{x}\| < 1 \\
&\leq \exp\left\{-\frac{1}{2}\left[\mathbf{x}^T\Omega^{(\mathrm{cs})}\mathbf{x} + \mathbf{x}^T\Omega^{(\mathrm{as})}\mathbf{x}\right]\right\} + \exp\left\{-\frac{1}{2}\left[\mathbf{x}^T\Omega^{(\mathrm{cs})}\mathbf{x} - \mathbf{x}^T\Omega^{(\mathrm{as})}\mathbf{x}\right]\right\},\ \forall\, \mathbf{x} \in \mathbb{R}^{2k-4} \\
&= \exp\left(-\frac{1}{2}\mathbf{x}^T\Omega\mathbf{x}\right) + \exp\left(-\frac{1}{2}\mathbf{x}^T\Omega^{\perp}\mathbf{x}\right).
\end{aligned} \tag{5.62}$$

To justify the inequality note that if $-\frac{1}{2}\mathbf{x}^T\Omega^{(as)}\mathbf{x} \geq 0$, then

$$-\frac{1}{2}(1-\mathbf{x}^T\mathbf{x})\mathbf{x}^T\Omega^{(\text{as})}\mathbf{x} \leq -\frac{1}{2}\mathbf{x}^T\Omega^{(\text{as})}\mathbf{x}, \tag{5.63}$$

whereas, if $-\frac{1}{2}\mathbf{x}^T\Omega^{(\text{as})}\mathbf{x} \leq 0$, then

$$-\frac{1}{2}(1-\mathbf{x}^T\mathbf{x})\mathbf{x}^T\Omega^{(\text{as})}\mathbf{x} \leq 0 \leq -\frac{1}{2}\mathbf{x}^T\Omega^{(\text{as})}\mathbf{x}. \tag{5.64}$$

A possible choice of an envelope, $g_{\text{MMN}}(\mathbf{x})$, is a simple mixture of two multivariate normal densities truncated to the unit disc with variance-covariance matrices Ω and $\Omega^{\perp}$, respectively, that is,

$$g_{\text{MMN}}(\mathbf{x}) = \frac{1}{2}|2\pi\Omega|^{-1/2}\exp\Big(-\frac{1}{2}\mathbf{x}^T\Omega\mathbf{x}\Big) + \frac{1}{2}|2\pi\Omega^{\perp}|^{-1/2}\exp\Big(-\frac{1}{2}\mathbf{x}^T\Omega^{\perp}\mathbf{x}\Big). \tag{5.65}$$

Under high concentrations the complex Bingham quartic (CBQ) distribution converges to the multivariate normal density i.e. $f_{\text{CBQ}}(\mathbf{x};\,\Omega) \approx N_{2k-4}(0,\Sigma)$ where $\Sigma = \Omega^{-1}$ and $g_{\text{MMN}}(\mathbf{x}) \approx \frac{1}{2}\Big[N_{2k-4}(0,\Sigma) + N_{2k-4}(0,\Sigma^{\perp})\Big]$. Thus, letting $g^*_{\text{MMN}}(\mathbf{x};\,\Omega,\,\Omega^{\perp})$ denote the probability density function in (5.65) without the normalizing constants $|2\pi\Omega|^{-1/2}$ and $|2\pi\Omega^{\perp}|^{-1/2}$, then (5.62) can be recast as

$$\underbrace{\exp\Big\{-\frac{1}{2}\big(\mathbf{x}^T\Omega\mathbf{x} - (\mathbf{x}^T\mathbf{x})\mathbf{x}^T\Omega^{(\text{as})}\mathbf{x}\big)\Big\}}_{f^*_{\text{CBQ}}(\mathbf{x};\,\Omega)} \;\leq\; 2\,\underbrace{\Big[\exp\Big(-\frac{1}{2}\mathbf{x}^T\Omega\mathbf{x}\Big) + \exp\Big(-\frac{1}{2}\mathbf{x}^T\Omega^{\perp}\mathbf{x}\Big)\Big]}_{g^*_{\text{MMN}}(\mathbf{x};\,\Omega,\,\Omega^{\perp})}, \tag{5.66}$$

for all $\mathbf{x} \in \mathbb{R}^{2k-4}$. The simulation bound for this algorithm is so cumbersome to compute analytically. Instead the simulated bounds can be calculated within the simulated R functions for the CBQ distribution. The efficiency rate with an envelope of mixture multivariate normal densities is around 50% under high concentrations. Overall, we have the following accept-reject algorithm.

Step 1 Simulation of the mixtures of two multivariate normals $\alpha N_{2k-4}(0,\Omega) + (1-\alpha)N_{2k-4}(0,\Omega^{\perp})$ (Rizzo [85], p. 78)

(1) Generate $U \sim \text{Uniform}[0,1]$.

(2) If $U \leq \alpha$ generate $\mathbf{x}$ from $N_{2k-4}(0,\Omega)$. Otherwise generate $\mathbf{x}$ from $N_{2k-4}(0,\Omega^{\perp})$ with arbitrary $\alpha = \frac{1}{2}$.

Step 2 Simulation of the complex Bingham quartic distribution

(1) Generate $\mathbf{x}$ from the mixture of normal densities, $g^*_{\text{MMN}}(\mathbf{x})$ with parameter matrices Ω and $\Omega^{\perp}$, respectively.

(2) Generate a random variable U from the Uniform$(0,1)$ distribution.

(3) If $U \leqslant \dfrac{f^*_{\text{CBQ}}(\mathbf{x})}{2g^*_{\text{MMN}}(\mathbf{x})}$, accept $\mathbf{x}$; otherwise reject $\mathbf{x}$ (In particular, reject any value lying outside the unit disk ($\|\mathbf{x}\| = \mathbf{x}^T\mathbf{x} > 1$) and repeat from step 1.

When $k = 3$, the complex projective $\mathbb{C}P^1$ can be isometrically identified with a sphere of radius $\frac{1}{2}$ in $\mathbb{R}^3$. Thus, up to factor of $\frac{1}{2}$, $\mathbb{C}P^1$ can be identified with S^2. The expression of this mapping is given in Kent [41] and Kent et al. [46]. Kent [41], and Dryden and Mardia [18], p. 117, showed that the CB distribution on $\mathbb{C}P^1$ (with quadratic terms in the exponent of the probability density function) can be identified with the Fisher distribution (which has linear terms in the the exponent of the probability density function). Kent et al. [46] showed also that the complex Bingham quartic (CBQ) distribution on $\mathbb{C}P^1$ can be identified with the FB5 distribution on S^2 (which has linear and quadratic terms in the the exponent of the probability density function). The complex Bingham quartic (CBQ) interpretation of the FB5 distribution has turned out to be useful for the simulation. Two simulation methods are proposed namely, rejection scheme with the mixture multivariate normals envelope truncated to the unit disc and the Kent-Hamelryck method. The Kent-Hamelryck [45] algorithm is clearly better under high concentration. If $k > 3$, only one method with efficiency approximately 50% under high concentration.

In our case $k = 3$, the 2×2-dimensional real symmetric matrix Ω, the anti-complex symmetric matrix $\Omega^{(\text{as})}$, the complex symmetric matrix $\Omega^{(\text{cs})}$ and $\Psi(s)$ reduce to the expressions (3.83), (3.84), (3.85) and (3.86), respectively. Further, $\Omega^{\perp}$ becomes

$$\Omega^{\perp} = \frac{1}{2}\begin{pmatrix} \lambda_1+\lambda_2 & 0 \\ 0 & \lambda_1+\lambda_2 \end{pmatrix} - \frac{1}{2}\begin{pmatrix} \lambda_1-\lambda_2 & 0 \\ 0 & \lambda_2-\lambda_1 \end{pmatrix} = \begin{pmatrix} \lambda_2 & 0 \\ 0 & \lambda_1 \end{pmatrix}, \tag{5.67}$$

$\Psi^{\perp}(s)$ reduces to

$$\begin{aligned} \Psi^{\perp}(s) &= -\frac{1}{2}\left[s\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} + s^2\begin{pmatrix} \lambda_1-\lambda_2 & 0 \\ 0 & \lambda_2-\lambda_1 \end{pmatrix}\right] \\ &= -\frac{1}{2}\begin{pmatrix} \lambda_1 s\big(1+s\big) - \lambda_2 s^2 & 0 \\ 0 & \lambda_2 s\big(1+s\big) - \lambda_1 s^2 \end{pmatrix}, \qquad 0 < s < 1, \end{aligned} \tag{5.68}$$

and the concentration parameter $\kappa \geq 0$ and the ovalness parameter $0 \leq \beta \leq \kappa/2$ are given by

$$\kappa = \frac{1}{8}(\lambda_1 + \lambda_2), \quad \beta = \frac{1}{16}\big(\lambda_2 - \lambda_1\big), \tag{5.69}$$

(Kent et al. [46]). For $k = 3$, Table 5.5 gives some values of the simulated efficiencies with various values of κ, β, λ_1 and λ_2 with fixing $\beta < \kappa/2$. The simulated efficiency rate under this A/R algorithm is found satisfactory for running a loop of R functions to generate pseudo random sample of size $n = 10000$ from the complex Bingham quartic (CBQ) distribution with lower and higher concentration parameters. Overall, it is clear from the table that the Kent-Hamelryck method with truncated double exponential envelope is more efficient than the rejection scheme with a mixture multivariate normals envelope truncated to the unit disc.

For $k = 4$, Table 5.6 gives some values of the simulated efficiencies with various values of λ_1, λ_2, λ_3 and λ_4. The simulated efficiency rate under this A/R algorithm is also found satisfactory for generating pseudo random sample of size $n = 1000$ from the complex Bingham quartic (CBQ) distribution with lower and higher concentration parameters. Overall, it is clear from the table that the rejection scheme with a mixture multivariate normals envelope truncated to the unit disc has efficiencies approximately 50% under high concentration.

λ_1	λ_2	κ	β	Simulated Efficiency Rate (KH)	S.E. (KH)	Simulated Efficiency Rate (MMN)	S.E. (MMN)
0.4	3.6	0.5	0.2	98.1%	0.00268	93.2%	0.00493
1.6	14.4	2	0.8	95.3%	0.00415	90.1%	0.00585
4	36	5	2	90.4%	0.00577	86.6%	0.00668
8	72	10	4	89.5%	0.00601	81.1%	0.00767
12	108	15	6	87.9%	0.00639	78.7%	0.00802
20	180	25	10	86.1%	0.00678	72.4%	0.00876
44	396	55	22	82.9%	0.00738	63.1%	0.00946
60	540	75	30	77.8%	0.00815	56.1%	0.00973
80	720	100	40	73.5%	0.00865	49.9%	0.00980
100	900	125	50	70.4%	0.00894	47.6%	0.00979
120	1080	150	60	68.2%	0.00913	45.3%	0.00975
140	1260	175	70	66.7%	0.00924	44.1%	0.00973
160	1440	200	80	65.1%	0.00934	42.7%	0.00969
200	1800	250	100	64.7%	0.00937	41.4%	0.00965

Table 5.5: Simulated efficiencies rates and their standard errors for the complex Bingham quartic (CBQ) distribution/Mixture Multivariate Normals and the Kent-Hamelryck envelopes in the case $k = 3$ with various values of κ, β, λ_1, λ_2 and a sample of size $n = 10000$ (fixed $2\beta/\kappa < 1$).

λ_1	λ_2	λ_3	λ_4	Simulated Efficiency Rate (MMN)	S.E. (MMN)
1	0	0	0	91.1%	0.00900
1	1	0	0	89.8%	0.00957
1	1	1	0	88.6%	0.01005
10	0	0	0	86.2%	0.01091
10	10	0	0	85.4%	0.01117
10	10	10	0	82.3%	0.01204
25	0	0	0	79.1%	0.01285
25	25	0	0	75.8%	0.01354
25	25	25	0	71.5%	0.01427
50	0	0	0	68.4%	0.01470
50	50	0	0	62.1%	0.01534
50	50	50	0	57.6%	0.01563
100	0	0	0	53.7%	0.01576
100	100	0	0	47.1%	0.01578
100	100	100	0	45.2%	0.01573

Table 5.6: Simulated efficiencies rates and their standard errors for the complex Bingham quartic (CBQ) distribution/Mixture Multivariate Normals envelope in the case $k = 4$ with various values of λ_1, λ_2, λ_3, λ_4 and a sample of size $n = 1000$.

Part III

Methods of Estimation for Torus Data

Chapter 6

Methods of Estimation for Torus Data

6.1 Introduction

In this chapter we review the moments of the sine model on the torus. Maximum likelihood (ML) and pseudolikelihood (PL) estimators for the sine distribution are briefly discussed. A comparison is given between the three bivariate sine and cosine models based on contours of the log-densities. For each of the three models, the parameters are chosen to match any positive definite inverse covariance matrix.

The multivariate wrapped normal has all marginals wrapped normal as well as the marginal bivariate distributions wrapped normal, and thus has a theoretical advantage. One disadvantage of the wrapped normal torus distribution is that it does not form an exponential family, unlike the sine and cosine models. Another disadvantage of the bivariate wrapped normal is that if we wrap $c\theta_1, c\theta_2$ instead of θ_1, θ_2 then the two wrapped distributions are different though the original correlation is invariant. Furthermore, the maximum likelihood estimators of the parameters, even for the univariate case are not computationally feasible, as is well known, and one has to resort to selecting some moment estimators which are not clear-cut for the bivariate parameters. Instead we calculate the moments for the wrapped normal torus distribution based on the sample variances covariances.

6.2 Bivariate von Mises Torus Distribution

Motivated by problems of modeling torsional angles in molecules, Singh et al. [89] proposed a bivariate circular distribution which is a natural torus analogue of the bivariate normal distribution. The general form is proposed by Mardia [60] and Mardia and Patrangenaru [74] but this particular case has some attractive properties among the minimal redundancy class. Let Θ_1 and Θ_2 be two circular random variables. The proposed probability density function is of the form

$$f(\theta_1,\theta_2) = \{c(\kappa_1,\kappa_2,\mathbf{A})\}^{-1}\exp\Big\{\kappa_1\cos(\theta_1-\mu_1)+\kappa_2\cos(\theta_2-\mu_2) \\ +[\cos(\theta_1-\mu_1),\sin(\theta_1-\mu_1)]\mathbf{A}[\cos(\theta_2-\mu_2),\sin(\theta_2-\mu_2)]^T\Big\}, \tag{6.1}$$

where $\theta_1,\theta_2\in(-\pi,\pi]$ lie on the torus, a square with opposite sides identified, $\kappa_1,\kappa_2>0$, $-\pi\le\mu_1$, $\mu_2<\pi$, the matrix $\mathbf{A}=(a_{ij})$ is 2×2 and $c(\cdot)$ is a normalization constant. This model has eight parameters and allows for dependence between the two angles. Various submodels with five parameters have appeared (Singh et al. [89]) to mimic the bivariate normal distribution. Three submodels are the sine and cosine models with positive interaction, and the cosine model with negative interaction. We will mainly concentrate on the sine model where $a_{11}=a_{12}=a_{21}=0, a_{22}=\eta$. The sine model has the density (Singh et al. [89] and Rivest [84])

$$f(\theta_1,\theta_2) = \{c(\kappa_1,\kappa_2,\eta)\}^{-1}\exp\Big\{\kappa_1\cos(\theta_1-\mu_1)+\kappa_2\cos(\theta_2-\mu_2)+\eta\sin(\theta_1-\mu_1)\sin(\theta_2-\mu_2)\Big\}. \tag{6.2}$$

The parameter $-\infty<\eta<\infty$ is a measure of dependence between θ_1 and θ_2. If $\eta=0$, then Θ_1 and Θ_2 are independent with each having univariate von Mises distributions.

Singh et al. [89] obtained an expression for the normalization constant in (6.2) as

$$c(\kappa_1,\kappa_2,\eta) = 4\pi^2\sum_{m=0}^{\infty}\binom{2m}{m}\Big(\frac{\eta}{2}\Big)^{2m}\kappa_1^{-m}I_m(\kappa_1)\kappa_2^{-m}I_m(\kappa_2). \tag{6.3}$$

This distribution has a natural generalization allowing multiple modes in the marginal distributions which is obtained by replacing by $(\theta_i-\mu_i)$ by $\tau_i(\theta_i-\mu_i), i=1,2$, in (6.2) where τ_1 and τ_2 are positive integers (Mardia et al. [68]).

For the cosine model with positive interaction the density is given by (Mardia et al. [64] and Mardia and Voss [73])

$$f(\theta_1,\theta_2)\propto\exp\Big\{\kappa_1\cos(\theta_1-\mu_1)+\kappa_2\cos(\theta_2-\mu_2)+\gamma_1\cos(\theta_1-\mu_1-\theta_2-\mu_2)\Big\}. \tag{6.4}$$

For the cosine model with negative interaction the density is given by (Kent et al. [48])

$$f(\theta_1,\theta_2) \propto \exp\Big\{\kappa_1\cos(\theta_1-\mu_1)+\kappa_2\cos(\theta_2-\mu_2)+\gamma_2\cos(\theta_1-\mu_1+\theta_2-\mu_2)\Big\}, \tag{6.5}$$

where $\gamma_1 \geq \gamma_2 \geq 0$ are the correlation parameters. This distribution is approximately bivariate normal when the fluctuations in the circular variables are small.

For $\mu_1 = \mu_2 = 0$ and under high concentration the density function of the Sine distribution in (6.2) becomes

$$\begin{aligned}
f_{\sin}(\theta_1,\theta_2) &\propto \exp\Big\{\kappa_1\cos\theta_1+\kappa_2\cos\theta_2+\eta\sin\theta_1\sin\theta_2\Big\}\\
&\approx \exp\Big\{\kappa_1\Big(1-\frac{\theta_1^2}{2}\Big)+\kappa_2\Big(1-\frac{\theta_2^2}{2}\Big)+\eta\theta_1\theta_2\Big\},\\
&\approx \exp\big(\kappa_1+\kappa_2\big)\exp\Big\{-\frac{1}{2}\Big(\kappa_1\theta_1^2+\kappa_2\theta_2^2+2\eta\theta_1\theta_2\Big)\Big\}\\
&\approx c_1\exp\left\{-\frac{1}{2}\begin{bmatrix}\theta_1 & \theta_2\end{bmatrix}\begin{bmatrix}\kappa_1 & -\eta\\ -\eta & \kappa_2\end{bmatrix}\begin{bmatrix}\theta_1\\ \theta_1\end{bmatrix}\right\}\\
&\approx c_1\exp\Big(-\frac{1}{2}\Theta^T\Sigma_1^{-1}\Theta\Big),
\end{aligned} \tag{6.6}$$

since

$$\cos\theta_i = 1-\frac{\theta_i^2}{2!}+\frac{\theta_i^4}{4!}+O(\theta_i^6)\approx 1-\frac{\theta_i^2}{2} \quad\text{and}\quad \sin\theta_i=\theta_i-\frac{\theta_i^3}{3!}+\frac{\theta_i^5}{5!}+O(\theta_i^7)\approx\theta_i \tag{6.7}$$

for $i = 1, 2$ up two terms approximation. For Σ_1 to exist and be positive definite matrix we require $\kappa_1 > 0$, $\kappa_2 > 0$ and $\kappa_1\kappa_2 > \eta^2$ (unimodal).

The cosine model with positive interaction in (6.4) also becomes

$$\begin{aligned}
f_{\cos}(\theta_1,\theta_2) &\propto \exp\Big\{\kappa_1\cos\theta_1+\kappa_2\cos\theta_2+\gamma_1\cos(\theta_1-\theta_2)\Big\}\\
&\approx \exp\Big\{\kappa_1\Big(1-\frac{\theta_1^2}{2}\Big)+\kappa_2\Big(1-\frac{\theta_2^2}{2}\Big)+\gamma_1\Big(1-\frac{(\theta_1-\theta_2)^2}{2}\Big)\Big\},\\
&\approx \exp\big(\kappa_1+\kappa_2+\gamma_1\big)\exp\Big\{-\frac{1}{2}\Big(\kappa_1\theta_1^2+\kappa_2\theta_2^2+\gamma_1(\theta_1^2+\theta_2^2-2\theta_1\theta_2\Big)\Big\}\\
&\approx c_2\exp\left\{-\frac{1}{2}\Big[\big(\kappa_1+\gamma_1\big)\theta_1^2+\big(\kappa_1+\gamma_1\big)\theta_2^2-2\gamma_1\theta_1\theta_2\Big]\right\}\\
&\approx c_2\exp\left\{-\frac{1}{2}\begin{bmatrix}\theta_1 & \theta_2\end{bmatrix}\begin{bmatrix}\kappa_1+\gamma_1 & -\gamma_1\\ -\gamma_1 & \kappa_2+\gamma_1\end{bmatrix}\begin{bmatrix}\theta_1\\ \theta_1\end{bmatrix}\right\}\\
&\approx c_2\exp\Big(-\frac{1}{2}\Theta^T\Sigma_2^{-1}\Theta\Big).
\end{aligned} \tag{6.8}$$

For Σ_2 to exist and be a positive definite matrix we require $\kappa_1 + \gamma_1 > 0$, $\kappa_2 + \gamma_1 > 0$ and $(\kappa_1 + \gamma_1)(\kappa_2 + \gamma_1) > \gamma_1^2$ (Mardia et al. [68]) i.e., $\gamma_1 < \kappa_1\kappa_2/(\kappa_1 + \kappa_2)$ (unimodal case). So, under high concentrations about $(\theta_1, \theta_2) = (0, 0)$, each of the models behaves as a bivariate normal distribution with inverse covariance matrix of the form

$$\Sigma_1^{-1} = \begin{bmatrix} \kappa_1 & -\eta \\ -\eta & \kappa_2 \end{bmatrix}, \quad \Sigma_2^{-1} = \begin{bmatrix} \kappa_1 + \gamma_1 & -\gamma_1 \\ -\gamma_1 & \kappa_2 + \gamma_1 \end{bmatrix} \quad \Sigma_3^{-1} = \begin{bmatrix} \kappa_1 - \gamma_2 & \gamma_2 \\ \gamma_2 & \kappa_2 - \gamma_2 \end{bmatrix}. \tag{6.9}$$

The geometry of the torus implies that it is not possible to get a single fully satisfactory analogue of the bivariate normal distribution. Some interim comparison between the cosine and sine models follow.

1. By construction, each of the sine and cosine models is symmetric, $f(\theta_1, \theta_2) = f(-\theta_1, -\theta_2)$. However, for each of the three models, f has a further symmetry since it is a continuous function on the torus,

$$f(\theta_1, \pi) = f(\theta_1, -\pi), \quad f(\pi, \theta_2) = f(-\pi, \theta_2).$$

   This latter property means that an elliptical pattern in the contours of constant probability for f will generally become distorted as (θ_1, θ_2) approaches the boundary of the square on which f is defined. In particular, this distortion complicates the development of efficient simulation algorithms using a 2-dimensional envelope since the density will not necessarily be monotonically decreasing on the rays from the origin to the edge of the square (Kent et al. [48]).

2. In most situations there is little difference between the sine and cosine models. Further, under high concentration, using either model is equivalent to fitting a bivariate normal distribution in a tangent plane.

3. For routine applications the sine model is somewhat easier to use, since it can be matched to any positive definite matrix Σ^{-1}, whereas the cosine models are limited to the dominated covariance case.

4. The sine and cosine models on the bivariate torus can be easily extended to a higher dimensional torus. It seems the sine model may be more suitable since the multivariate extension is attractive (Mardia et al. [75], Mardia et al. [64] and Kent et al. [48]).

6.2.1 Marginal and Conditional Models

In this subsection, the conditional and marginal probability densities of the random variables (Θ_1, Θ_2) will be derived for the sine and cosine models in (6.2) and (6.4). When $\mu_1 = \mu_2 = 0$, the densities of the sine and cosine with positive interaction, and cosine with negative interaction models are given by

$$\begin{aligned} f_{\text{sin}}(\theta_1, \theta_2) &= \Big\{c(\kappa_1, \kappa_2, \eta)\Big\}^{-1} \exp\Big\{\kappa_1 \cos\theta_1 + \kappa_2 \cos\theta_2 + \eta \sin\theta_1 \sin\theta_2\Big\}, \\ f_{\text{cosinePI}}(\theta_1, \theta_2) &= \Big\{c(\kappa_1, \kappa_2, \gamma_1)\Big\}^{-1} \exp\Big\{\kappa_1 \cos\theta_1 + \kappa_2 \cos\theta_2 + \gamma_1 \cos(\theta_1 - \theta_2)\Big\}, \\ f_{\text{cosineNI}}(\theta_1, \theta_2) &= \Big\{c(\kappa_1, \kappa_2, \gamma_2)\Big\}^{-1} \exp\Big\{\kappa_1 \cos\theta_1 + \kappa_2 \cos\theta_2 + \gamma_2 \cos(\theta_1 + \theta_2)\Big\}, \end{aligned} \tag{6.10}$$

respectively.

Let (Θ_1, Θ_2) be distributed according to the probability density function of the sine distribution in (6.10). When the fluctuations in Θ_1 and Θ_2 are sufficiently small, so that

$$\cos(\theta_i) \approx 1 - \frac{1}{2}\theta_1^2, \quad \sin\theta_i = \theta_i \ \ (i = 1, 2), \tag{6.11}$$

then it follows that (Θ_1, Θ_2) has approximately a bivariate normal distribution with parameters

$$\sigma_1^2 = \frac{\kappa_2}{\kappa_1\kappa_2 - \eta^2}, \ \ \sigma_2^2 = \frac{\kappa_1}{\kappa_1\kappa_2 - \eta^2}, \ \ \rho = \frac{\eta}{\sqrt{\kappa_1\kappa_2}}. \tag{6.12}$$

For these bivariate normal parameters to be meaningful, only the condition $\eta^2 < \kappa_1\kappa_2$.

Define new parameters α and β by $\kappa_2 = \alpha\cos\beta$ and $\eta\sin\theta_1 = \alpha\sin\beta$, so that

$$\alpha = \Big(\kappa_2^2 + \eta^2 \sin^2\theta_1\Big)^{1/2}, \quad \tan\beta = (\eta/\kappa_2)\sin\theta_1. \tag{6.13}$$

Write $\alpha = \alpha(\theta_1)$ and $\beta = \beta(\theta_1)$ to emphasize the dependence on θ_1. Then the marginal probability density function of Θ_1 for the sine model in (6.10) is given by

$$\begin{aligned} f_{\text{sin}}(\theta_1) &= \int_{-\pi}^{\pi} f_{\text{sin}}(\theta_1, \theta_2)\, d\,\theta_2 \\ &= \int_{-\pi}^{\pi} \Big\{c(\kappa_1, \kappa_2, \eta)\Big\}^{-1} \exp\Big\{\kappa_1 \cos\theta_1 + \kappa_2 \cos\theta_2 + \eta \sin\theta_1 \sin\theta_2\Big\}\, d\,\theta_2 \\ &= \Big\{c(\kappa_1, \kappa_2, \eta)\Big\}^{-1} 2\pi I_0(\alpha(\theta_1)) \exp\{\kappa_1 \cos\theta_1\} \ \ (-\pi \le \theta_1 < \pi), \end{aligned} \tag{6.14}$$

where $I_0(\cdot)$ is a modified Bessel function of the first kind of order $\nu = 0$ (see, for example, Abramowitz and Stegun [1], p.376, Mardia [59], p.57 and Mardia and Jupp [70], p.349). The marginal probability density function of Θ_2 is obtained in a similar way.

Given that $\Theta_1 = \theta_1$, the conditional probability density function of Θ_2 is

$$\begin{aligned} f_{\sin}(\theta_2/\theta_1) &= \frac{f_{\sin}(\theta_2, \theta_1)}{f_{\sin}(\theta_1)} \\ &= \frac{1}{2\pi I_0\big(\alpha(\theta_1)\big)} \exp\Big\{\alpha(\theta_1)\cos\Big(\theta_2 - \beta(\theta_1)\Big)\Big\}. \end{aligned} \tag{6.15}$$

Thus the conditional probability density function of Θ_2, given that $\Theta_1 = \theta_1$, is a von Mises distribution with the concentration parameter $\alpha(\theta_1)$ and mean angle $\beta(\theta_1)$. If $\kappa_2 \to \infty$ and $\eta \to \infty$ so that $\eta/\kappa_2 \to \zeta$, then the concentration parameter $\alpha(\theta_1)$ of the conditional von Mises distribution tends to infinity for each given θ_1 (Singh et al. [89] and Mardia et al. [76]).

The marginal probability density function of Θ_2 for the cosine model in (6.10) is also given by

$$\begin{aligned} f_{\cos}(\theta_2) &= \int_{-\pi}^{\pi} f_{\cos}(\theta_1, \theta_2)\, d\,\theta_1 \\ &= \int_0^{2\pi} \Big\{c(\kappa_1, \kappa_2, \gamma_1)\Big\}^{-1} \exp\Big\{\kappa_1 \cos\theta_1 + \kappa_2 \cos\theta_2 + \gamma_1 \cos(\theta_1 - \theta_2)\Big\}\, d\,\theta_1 \\ &= \Big\{c(\kappa_1, \kappa_2, \gamma_1)\Big\}^{-1} 2\pi I_0\big(\alpha(\theta_2)\big) \exp\{\kappa_2 \cos\theta_2\}, \end{aligned} \tag{6.16}$$

where $\alpha(\theta_2) = \Big(\kappa_1^2 + \gamma_1^2 - 2\kappa_1\gamma_1 \cos\theta_2\Big)^{1/2}$. Further, the conditional probability density function $f_{\cos}(\theta_1/\theta_2)$ is a von Mises distribution with the concentration parameter $\alpha(\theta_2)$, mean angle $\beta(\theta_2)$ and $\tan\beta(\theta_2) = \big(\gamma_1/(\kappa_1 - \gamma_1\cos\theta_2)\big)$ (which means that $-\pi/2 < \theta_{2\mu} < \pi/2$).

6.2.2 Comparing Models using Log-Densities

The bivariate densities can be represented by contour plots which can be used to illustrate various statements of comparison. However, the key features are more easily compared by plotting the log of the density and omitting the normalizing constant. The reason for this is that the logarithm of the exponential family does not change if we multiply it by a constant and the plots of the log densities will be clear even in the tails. The contour plots of the joint density provide insight into the manner in which the parameters indexes the amount of probability that concentrates along the curve. For each of the three models, it is possible to choose the parameters to match any positive definite inverse covariance matrix. Figures 6.1 and 6.2 plot the contours for the sine and cosine with positive interaction log-densities and cosine with negative interaction models for $\kappa_1 = \kappa_2 = 3$ and for a range values of η, γ_1 and γ_2. Figures 6.3 and Figures 6.4 illustrate the 3D perspective plots of the three models with a range of their parameters. In comparing the three models we will note:

(a) Changing the sign of η in the sine model causes a reflection in the axes.

(b) It is clear from the 3D perspective plots that the bimodality of the sine model occurs only if $\kappa_1\kappa_2 < \eta^2$.

(c) The bimodality for the cosine model with positive interaction occurs only if $(\kappa_1+\gamma_1)(\kappa_2+\gamma_1) < \gamma_1^2$.

(d) Changing the sign of small γ_1 give an approximate reflection.

(e) For small $\eta \approx -\gamma_1$ the two models are approximately the same.

(f) For large $\gamma_1 = -\gamma_2$ the cosine models with positive interaction is similar to that of negative interaction.

(g) Transforming (θ_1, θ_2) to $(\theta_2, -\theta_1)$ is equivalent to changing the sign of η in the sine model, but also allows for rotations of the cosine model contour plots (which cannot be achieved by changing γ_1 (Mardia et al. [75]).

6.3 Moments and Correlations Under High Concentration

Another approach for comparing the sine and cosine models is to examine their moments. Characteristic functions can be used to estimate the moments of the two distributions and we can numerically compute the correlations for the cosine and sine models in order to study their behaviour in relation to the parameters η and γ_2.

Let ϕ be the characteristic function for the bivariate directional model and let $\mathbb{E}_{\sin}\big(\cos(p\theta_1+q\theta_2)\big)$ and $\mathbb{E}_{\cos}\big(\cos(p\theta_1+q\theta_2)$ are the real parts of the characteristic function for the sine and cosine models, respectively. When $p = q = 1$ the components of the real part of the first moments are of interest, namely $\mathbb{E}\big(\cos(\theta_1)\sin(\theta_2)\big)$, $\mathbb{E}\big(\cos(\theta_1)\cos(\theta_2)\big)$ and $\mathbb{E}\big(\sin(\theta_1)\sin(\theta_2)\big)$.

If $f_{\sin}(\theta_1, \theta_2)$ is the probability density function for the sine model in (6.10), and since $f_{\sin}(\theta_1, \theta_2)$ is reflection symmetric about x and y axes, then

$$\mathbb{E}\big(\cos\theta_1 \sin\theta_2\big) = 0.$$

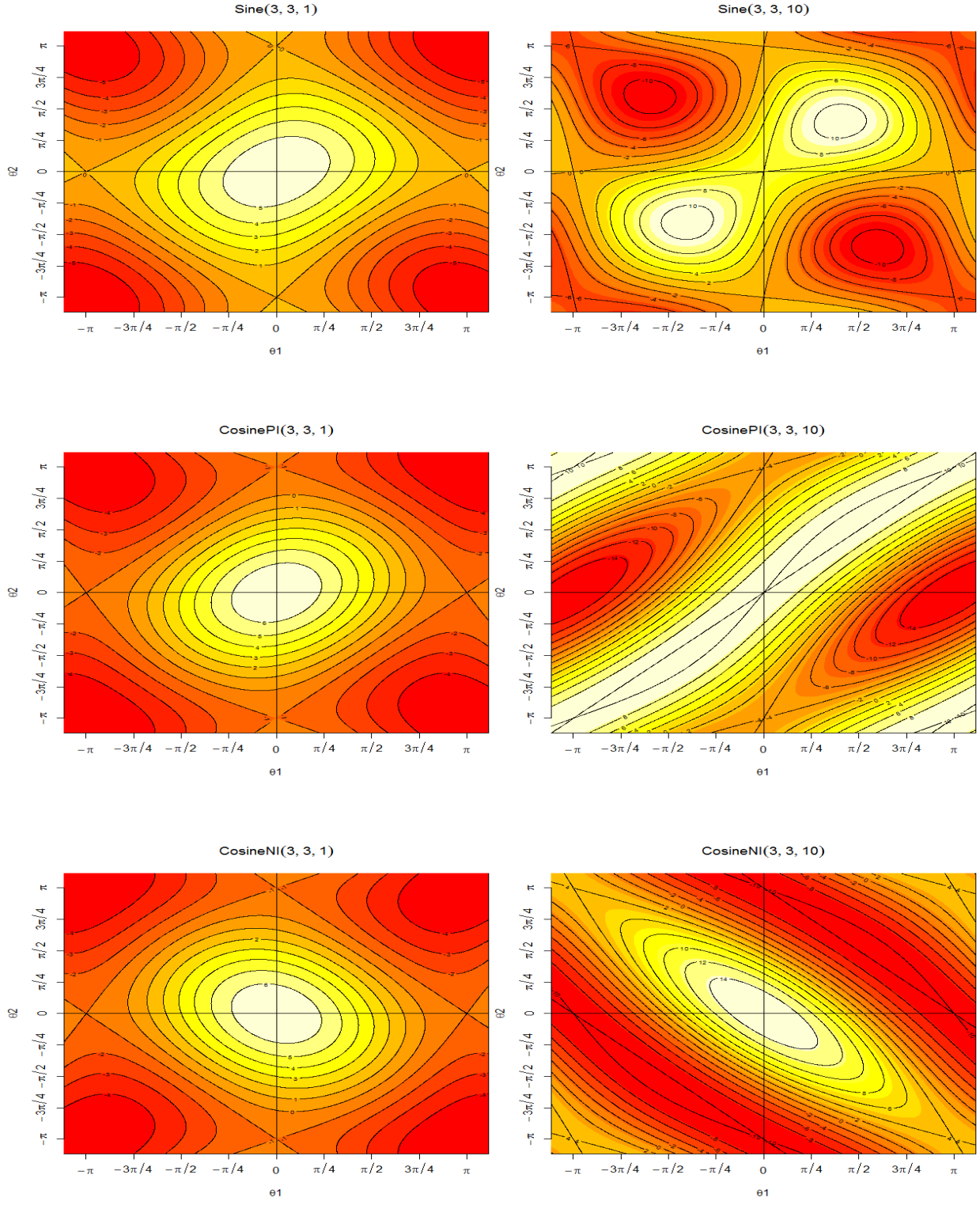

Figure 6.1: Contour plots of log-densities (normalizing constant omitted) for the sine, cosine with positive interaction and cosine with negative interaction models. Captions indicate the vector of parameters: κ_1, κ_2, η (Positive) for the sine model, $\kappa_1, \kappa_2, \gamma_1$ (Positive) for the cosine model with positive interaction model and $\kappa_1, \kappa_2, \gamma_2$ (Positive) for cosine with negative interaction models.

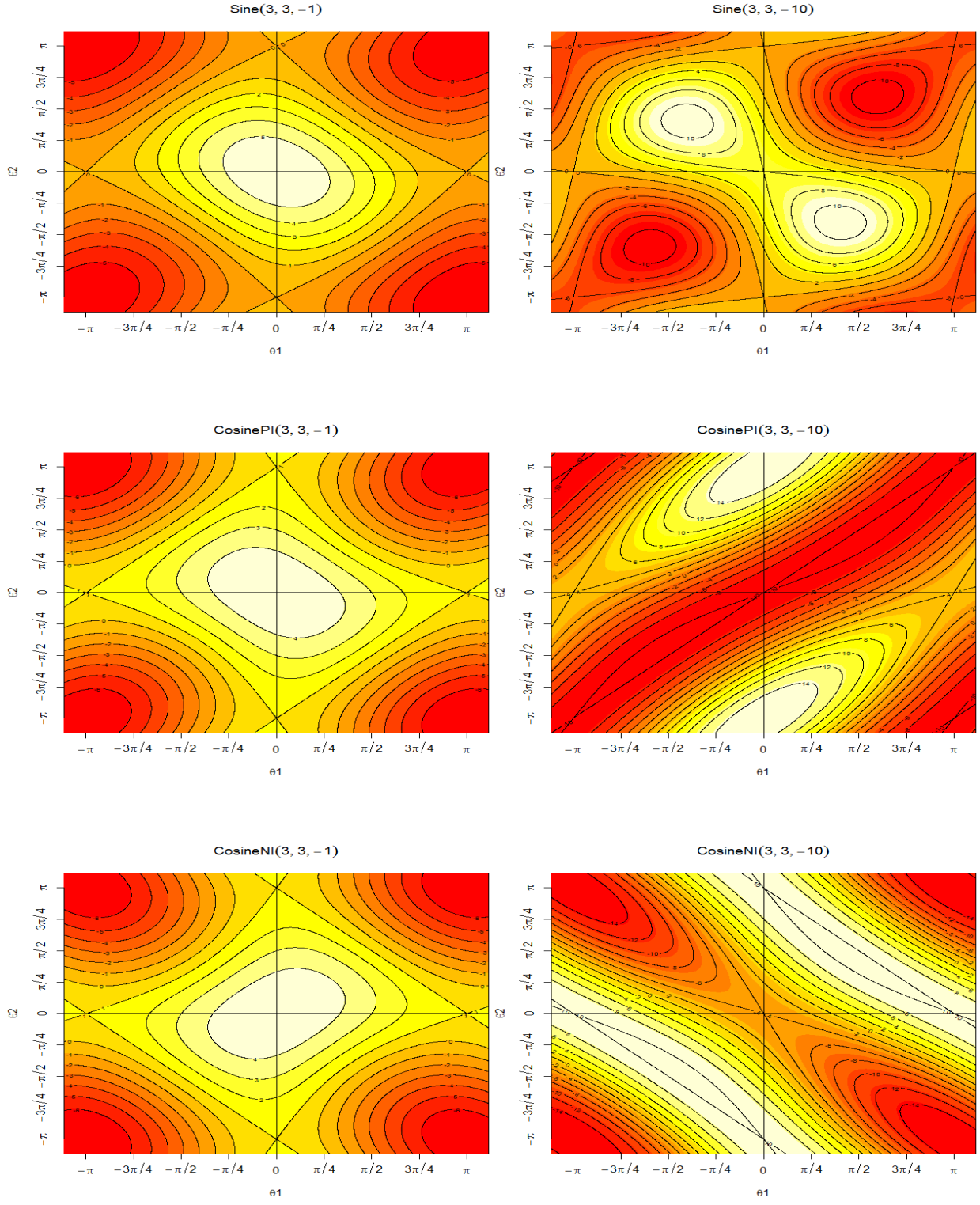

Figure 6.2: Contour plots of log-densities (normalizing constant omitted) for the sine, cosine with positive interaction and cosine with negative interaction models. Captions indicate the vector of parameters: κ_1, κ_2, η (Negative) for the sine model, $\kappa_1, \kappa_2, \gamma_1$ (Negative) for the cosine model with positive interaction model and $\kappa_1, \kappa_2, \gamma_2$ (Negative) for cosine with negative interaction models.

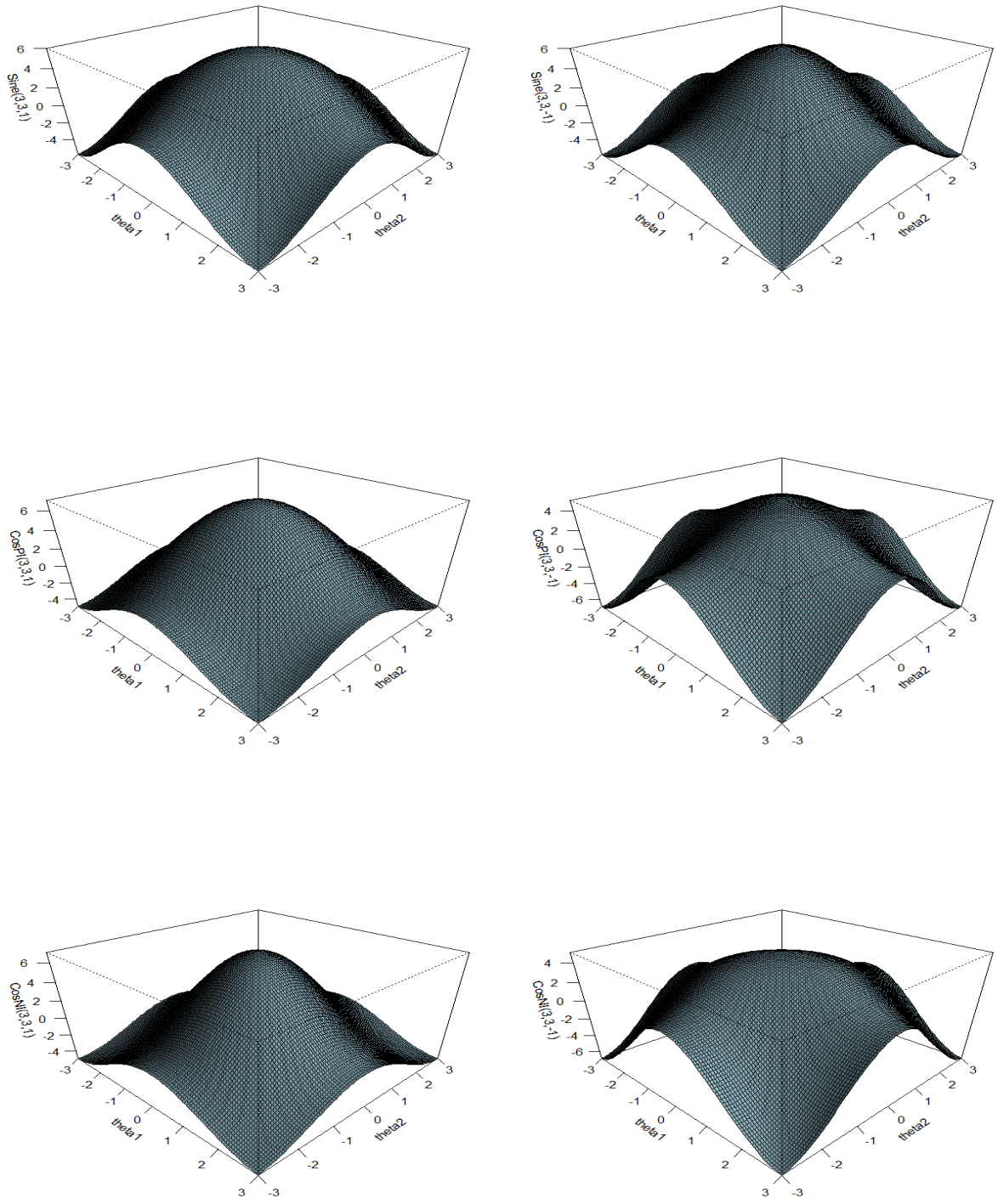

Figure 6.3: 3D perspective plots of the log-densities (normalizing constant omitted) for the sine, cosine with positive interaction and cosine with negative interaction models. Captions indicate the vector of parameters: κ_1, κ_2, η(Positive) for the sine model, $\kappa_1, \kappa_2, \gamma_1$ (Positive) for the cosine model with positive interaction model and $\kappa_1, \kappa_2, \gamma_2$ (Positive) for cosine with negative interaction models.

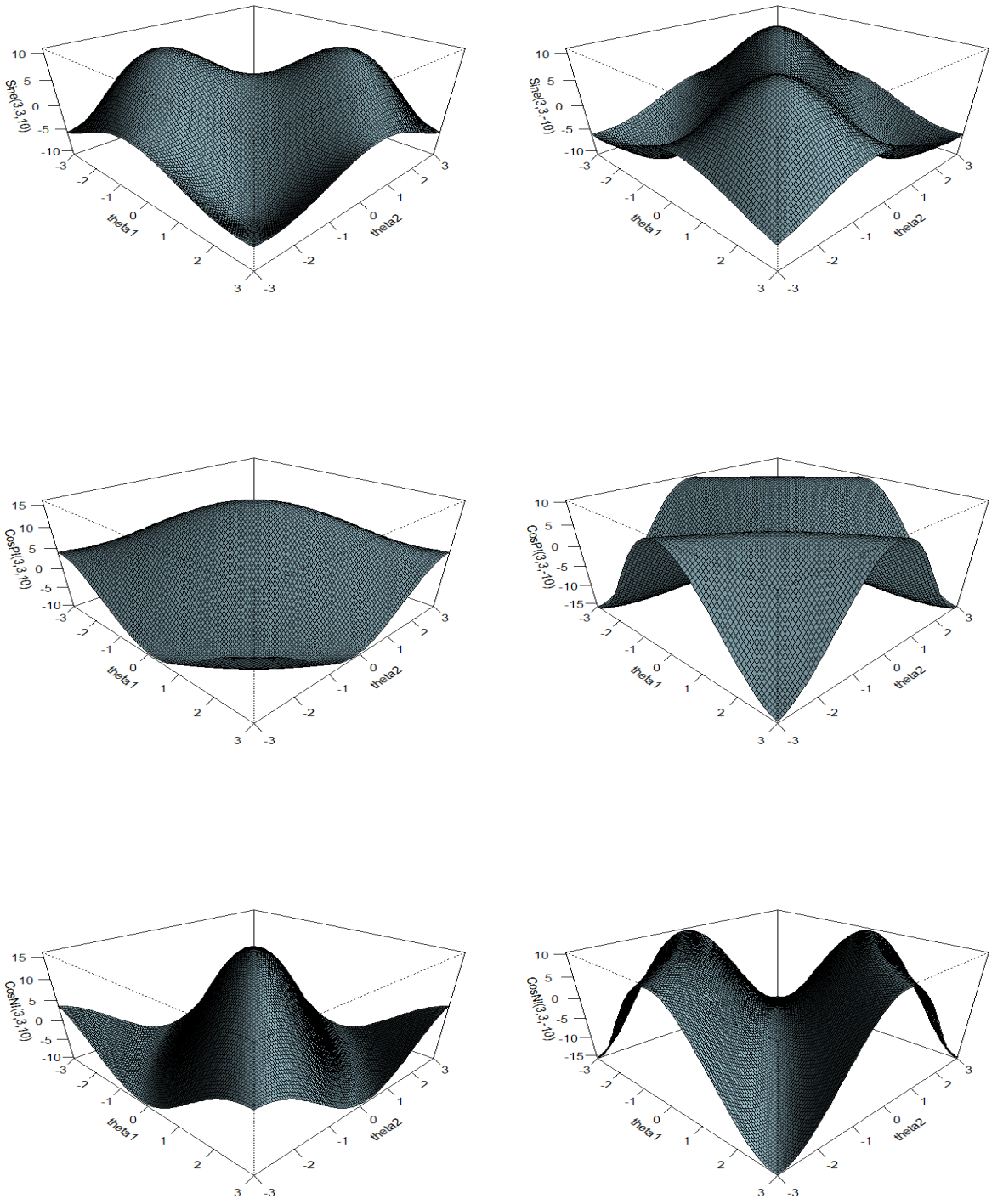

Figure 6.4: 3D perspective plots of the log-densities (normalizing constant omitted) for the sine, cosine with positive interaction and cosine with negative interaction models. Captions indicate the vector of parameters: κ_1, κ_2, η(Positive) for the sine model, $\kappa_1, \kappa_2, \gamma_1$ (Positive) for the cosine model with positive interaction model and $\kappa_1, \kappa_2, \gamma_2$ (Positive) for cosine with negative interaction models.

Under high concentration

$$\begin{aligned}\mathbb{E}\big(\sin\theta_1\sin\theta_2\big) &\approx \mathbb{E}(\theta_1\theta_2)\\ &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\theta_1\theta_2\exp\Big\{\kappa_1\cos\theta_1+\kappa_2\cos\theta_2+\eta\sin\theta_1\sin\theta_2\Big\}d\theta_1\,d\theta_2\\ &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\exp\Big\{\kappa_1\theta_1\theta_2\cos\theta_1+\kappa_2\theta_1\theta_2\cos\theta_2+\eta\theta_1\theta_2\sin\theta_1\sin\theta_2\Big\}d\theta_1\,d\theta_2,\end{aligned}$$

which resembles covariance and

$$\begin{aligned}\mathbb{E}\big(\cos\theta_1\cos\theta_2\big) &\approx \mathbb{E}\left[\Big(1-\frac{\theta_1^2}{2}\Big)\Big(1-\frac{\theta_2^2}{2}\Big)\right]\\ &\approx \mathbb{E}\left[1-\frac{\theta_1^2}{2}-\frac{\theta_2^2}{2}\right]\\ &= 1-\frac{1}{2}\mathbb{E}\Big(\theta_1^2+\theta_2^2\Big)\\ &= 1-\frac{1}{2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\Big(\theta_1^2+\theta_2^2\Big)\exp\Big\{\kappa_1\cos\theta_1+\kappa_2\cos\theta_2+\eta\sin\theta_1\sin\theta_2\Big\}d\theta_1\,d\theta_2,\end{aligned}$$

which resembles the squared covariance. These expectations can be used to obtain the correlations between θ_1 and θ_2 for the sine and cosine moments. It is also of interest to empirically explore the relationship between the bivariate normal distribution and the cosine and sine models when the data are highly concentrated. We estimate the correlation under the assumption of normality induced by high concentrations. We compare this estimated correlation under the assumption of normality with the explicit correlation evaluated from the distributions by integration. Under high concentrations, Mardia et al. [75] obtained an approximate correlation between θ_1 and θ_2, for the cosine model with negative interaction, as

$$\rho_{\text{cosineNI}} = \frac{-\gamma_2}{\sqrt{(\kappa_1-\gamma_2)(\kappa_2-\gamma_2)}} \cong \text{corr}(\sin\theta_1,\sin\theta_2). \tag{6.17}$$

For the sine model, Singh et al. [89] obtained the correlation as

$$\rho_{\sin} = \frac{\eta}{\sqrt{\kappa_1\kappa_2}} \tag{6.18}$$

Mardia et al. [76] gave numerical comparisons between the sine and cosine models according to the correlations between θ_1 and θ_2. Using some selected values of κ_1 and κ_2, they observed that the correlation between $\cos\theta_1$ and $\cos\theta_2$ and between $\sin\theta_1$ and $\sin\theta_2$ were seen to be (mostly) decreasing functions of γ_2 for the cosine model, whereas for the sine model the correlation between $\sin\theta_1$ and $\sin\theta_2$ was a monotonic increasing function of η and between $\cos\theta_1$ and $\cos\theta_2$ was always non-negative and has a U-shaped relationship with η.

6.4 Approaches to Estimation

6.4.1 Maximum Likelihood Method

The usual maximum likelihood estimators (MLE) for the parameters of the sine model of zero means in (6.10) are obtained by maximizing the likelihood function given by

$$\{c(\kappa_1,\kappa_2,\eta)\}^{-n}\exp\Big\{\kappa_1\sum_{i=1}^{n}\cos(\theta_{1i})+\kappa_2\sum_{i=1}^{n}\cos(\theta_{2i})+\eta\sum_{i=1}^{n}\sin(\theta_{1i})\sin(\theta_{2i})\Big\}. \tag{6.19}$$

Mardia et al. [68] computed the maximum likelihood estimators for κ_1, κ_2 and η. The algorithm may start with

$$\hat{\kappa}_1 = A_1^{-1}(\bar{R}_1) \text{ and } \hat{\kappa}_2 = A_1^{-1}(\bar{R}_2), \tag{6.20}$$

where $A_v(y) = I_v(y)/I_0(y)$, $I_v(\cdot)$ is the Bessel function and

$$\bar{R}_1 = \frac{1}{n}\sum_{r=1}^{n}\cos\theta_{1r} \text{ and } \bar{R}_2 = \frac{1}{n}\sum_{r=1}^{n}\cos\theta_{2r}. \tag{6.21}$$

The value of $\hat{\eta}$ is obtained from $\mathbb{E}\big(\sin(\theta_1)\sin(\theta_2)\big)$ using these values of $\hat{\kappa}_1$, $\hat{\kappa}_2$ through (6.3). The values of $\hat{\kappa}_2$ and $\hat{\eta}$ are used in $\mathbb{E}\big(\cos(\theta_1)\big)$ to obtain new $\hat{\kappa}_1$. Using this new value of $\hat{\kappa}_1$ and the value of $\hat{\eta}$, the new value of $\hat{\kappa}_2$ is obtained using $\mathbb{E}\big(\cos(\theta_2)\big)$. The procedure is continued till the convergence to the solution is achieved. They proved that the maximum likelihood estimators and the moment estimators coincide for the parameters of the bivariate sine model.

For the maximum likelihood estimators (MLE), the standard errors are calculated from the Hessian matrix (by taking the square roots of the diagonal elements of the inverse of the Hessian), which is obtained numerically by the `nlm` estimation routine in R. The Hessian matrix H for the Sine distribution is obtained as follows (Mardia et al. [75]). Letting $g = \log(f(\theta_1,\theta_2)$ we obtain

$$\begin{aligned}
G_{\theta_1} = \frac{\partial g}{\partial\theta_1} &= -\kappa_1\sin\theta_1+\eta\cos\theta_1\sin\theta_2\\
G_{\theta_2} = \frac{\partial g}{\partial\theta_2} &= -\kappa_2\sin\theta_2+\eta\sin\theta_1\cos\theta_2,\\
G_{\theta_1\theta_1} = \frac{\partial^2 g}{\partial\theta_1\partial\theta_1} &= -\kappa_1\cos\theta_1-\eta\sin\theta_1\sin\theta_2,\\
G_{\theta_2\theta_2} = \frac{\partial g}{\partial\theta_2\partial\theta_2} &= -\kappa_2\cos\theta_2-\eta\sin\theta_1\cos\theta_2,\\
G_{\theta_1\theta_2} = G_{\theta_2\theta_1} = \frac{\partial^2 g}{\partial\theta_1\partial\theta_2} &= \eta\cos\theta_1\cos\theta_2.
\end{aligned} \tag{6.22}$$

The Hessian matrix H is then given by

$$H = \begin{pmatrix} G_{\theta_1\theta_1} & G_{\theta_1\theta_2} \\ G_{\theta_1\theta_2} & G_{\theta_1\theta_2} \end{pmatrix} = \begin{pmatrix} -\kappa_1 \cos\theta_1 - \eta \sin\theta_1 \sin\theta_2 & \eta \cos\theta_1 \cos\theta_2 \\ \eta \cos\theta_1 \cos\theta_2 & -\kappa_2 \cos\theta_2 - \eta \sin\theta_1 \cos\theta_2. \end{pmatrix} \tag{6.23}$$

6.4.2 Maximum Pseudolikelihood Method

Define the pseudolikelihood (Besag [4]), based on a random sample of n observations of $\Theta = (\theta_1, \theta_2)^T$, by

$$\text{PL} = \prod_{j=1}^{p} \prod_{1}^{n} g_j(\Theta_{ji} | (\Theta_{1i}, \Theta_{2i}, \ldots, \Theta_{pi}); q) \tag{6.24}$$

where $g_j(\cdot | \ldots; q)$ is the conditional distribution whose parameters will depend on j and q is an unknown parameter of length r. For the sine distribution with zero means we have

$$\text{PL} = (2\pi)^{-2p} \prod_{1}^{2} \prod_{1}^{n} \Big[I_0\big(\kappa_{1.\text{rest}}^{(i)}\big) \Big]^{-1} \exp\Big\{ \kappa_{j.\text{rest}}^{(i)} \cos(\theta_{ji}) \Big\} \tag{6.25}$$

where

$$\kappa_{j.\text{rest}}^{(i)} = \left\{ \kappa_j^2 + \Big[\eta \sin(\theta_j) \Big]^2 \right\}^{1/2}. \tag{6.26}$$

Parameter estimation based on the pseudo-likelihood approach proceeds by maximizing the PL in (6.24) with respect to the $p + p(p+1)/2 = 5$ unknown parameters (Mardia el. al. [68]). Here we have 3 unknown parameters κ_1, κ_2 and η since we assume that $\mu_1 = \mu_2 = 0$.

Under high concentrations, the sine distribution is asymptotically bivariate normal with mean 0 and inverse covariance matrix Σ_1^{-1} as in (6.9). For the pseudo-likelihood estimators (PLE), the standard errors are calculated from a jackknife estimate of the covariance matrix (Mardia el. al. [68] and Mardia and Kent [71]).

Mardia et al. [69]) studied the efficiency of pseudo-likelihood method for bivariate von Mises torus distribution. They found that the efficiency tends to unity with increasing concentrations. Simulations support the numerical calculations obtained. With the exception of the bimodal case studied, it is seen that the bias of the pseudolikelihood estimator is very similar to that of the maximum likelihood estimator, and for one parameter configuration, the bias is smaller.

6.5 Wrapped Normal Torus Distribution

6.5.1 Overview and Background

Suppose that Θ is a vector of angles following a wrapped normal torus distribution; that is,

$$\Theta_j = X_j \bmod 2\pi, \tag{6.27}$$

for $j = 1, 2, \ldots, p$ where X has the multivariate probability density function with zero mean vector and $p \times p$ symmetric positive definite variance covariance matrix Σ

$$f(\mathbf{x}; \Sigma) = (2\pi)^{-p/2}|\Sigma|^{-1/2} \exp\Big(-\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\Big), \;\; \mathbf{x} \in \mathbb{R}^p. \tag{6.28}$$

For a $p \times 1$ vector with integer coefficients, $\delta \in \mathbb{R}^p$, and $i = \sqrt{-1}$ the characteristic function $\phi_X(\delta)$ for the multivariate normal distribution (Mardia et al. [72], p. 74, Manley [58], p. 17 and Rencher [83], p. 85) is given by

$$\begin{aligned} \phi_X(\delta) &= \mathbb{E}\Big[\exp(i\delta^T X)\Big] \\ &= \mathbb{E}\big[\cos(\delta^T X)\big] + i\mathbb{E}\big[\sin(\delta^T X)\big] \\ &= \exp\Big(-\frac{1}{2}\delta^T\Sigma\delta\Big). \end{aligned} \tag{6.29}$$

The density corresponding to a p-variate wrapped normal density is given by

$$f_w(\Theta; \Sigma) = \sum_{t_1}^{\infty}\sum_{t_2}^{\infty}\ldots\sum_{t_p}^{\infty} f[(\Theta + 2\pi)\mathbf{k}; \Sigma] \tag{6.30}$$

where the vector of angles, $\Theta = (\theta_1, \theta_2 \ldots \theta_p)^T \in [0, 2\pi)^p$ and $\mathbf{k} = (k_1, k_2, \ldots, k_p)^T$ is set of integers (Coles [10], Johnson and Wehrly [34]).

In the case $p = 2$, suppose that the random variables X_1 and X_2 have the bivariate normal distribution with zero mean vector, variance-covariance matrix ($\sigma_{11} = \sigma_{22} = \sigma^2$ and $\sigma_{12} = \sigma_{21} = \rho\sigma^2$, say)

$$\Sigma = \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}. \tag{6.31}$$

Simulation from the wrapped normal torus distribution is straightforward: X_j is simulated from $f(\mathbf{x}; \Sigma)$ in (6.28) and then Θ_j formed componentwise by (6.27) and this is repeated to obtain a sample of size n. Figures 6.5, 6.6 and 6.7 illustrate the effect of wrapping. For each figure both

the simulated $X = (x_1, x_2)^T$ and $\Theta = (\theta_1, \theta_2)^T$ values are shown. The aim of the inference can be thought of as inferring the bivariate normal structure of data X plotted in each of the (a) figures, from the corresponding wrapped data Θ, plotted in the (b) figures. This is rather more challenging in the case of large variance and weak correlation since the original sample produce a more uniform scatter of points in the wrapped sample from which it would be difficult to guess at the structure in the pre-wrapped data (Coles [10]).

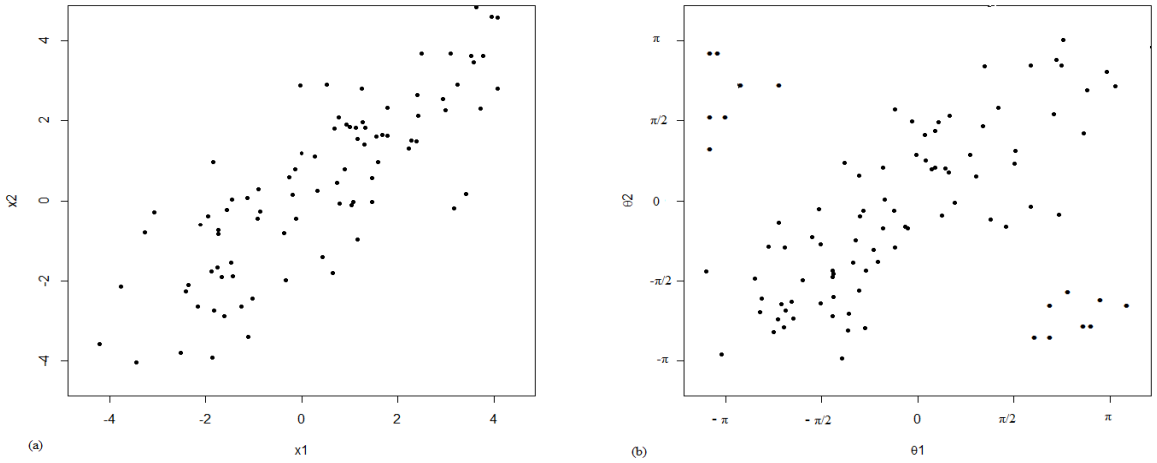

Figure 6.5: Simulated bivariate normal torus data with small variance $\sigma^2 = 1.0$ and strong correlation $\rho = 0.90$ for a sample of size $n = 100$. (a) before wrapping; (b) after wrapping.

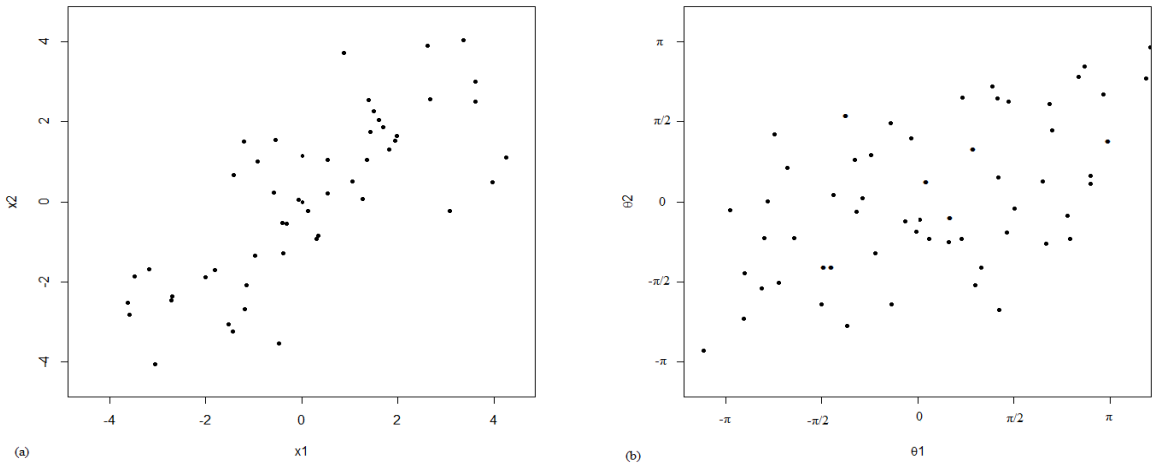

Figure 6.6: Simulated bivariate normal torus data with large variance $\sigma^2 = 50$ and strong correlation $\rho = 0.90$ for a sample of size $n = 100$. (a) before wrapping; (b) after wrapping.

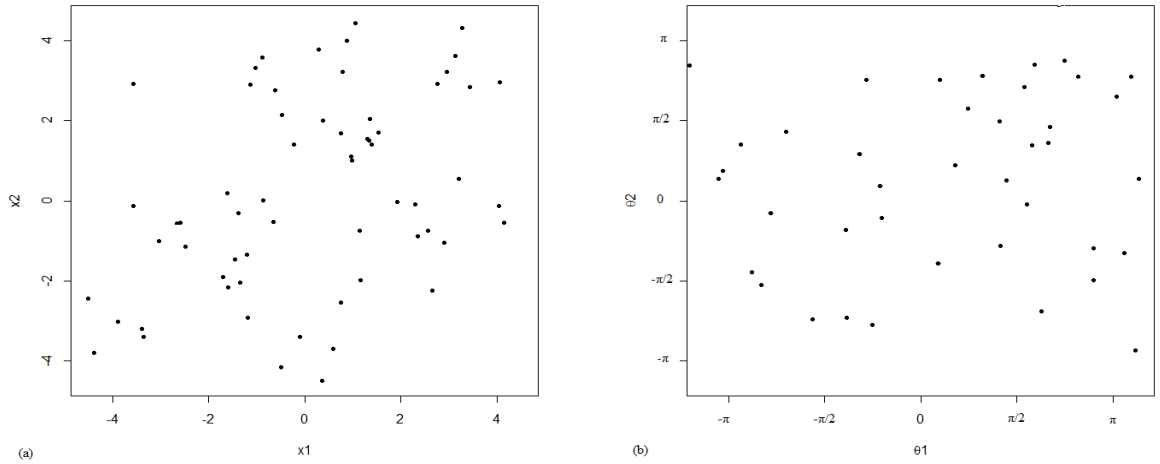

Figure 6.7: Simulated bivariate normal torus data with large variance $\sigma^2 = 50$ and weak correlation $\rho = 0.50$ for a sample of size $n = 100$. (a) before wrapping; (b) after wrapping.

6.5.2 Moments for the Wrapped Normal Torus Distribution

One reason for choosing the wrapped normal torus distribution is that the trigonometric moments have straightforward explicit expressions. If δ is $p \times 1$ vector with integer coefficients, then

$$\begin{aligned} \mathbb{E}\big[\cos(\delta^T\Theta)\big] &= \exp\Big(-\frac{1}{2}\delta^T\Sigma\delta\Big) \\ \mathbb{E}\big[\sin(\delta^T\Theta)\big] &= 0. \end{aligned} \tag{6.32}$$

In particular, for $j, k = 1, 2, \ldots, p$, the first order trigonometric moments are

$$\begin{aligned} \mathbb{E}\big[\cos(\theta_j)\big] &= \exp\Big(-\frac{1}{2}\sigma_{jj}\Big) = c_i, \text{say} \\ \mathbb{E}\big[\sin(\theta_j)\big] &= 0, \\ \mathbb{E}\big[\cos(\theta_j \pm \theta_k)\big] &= \exp\Big(-\frac{1}{2}(\sigma_{jj} \pm 2\sigma_{jk} + \sigma_{kk})\Big), \\ \mathbb{E}\big[\sin(\theta_j \pm \theta_k)\big] &= 0. \end{aligned} \tag{6.33}$$

Let the vector $\mathbf{c} = (c_1, c_2, \ldots, c_p)^T$ and write $\mathbf{D} = \text{diag}(\mathbf{c})$. Combining the two versions of the last two equations yields (Kent and Mardia [49])

$$\begin{aligned} \mathbb{E}\big[\cos(\theta_j) + \cos(\theta_k)\big] &= c_j c_k \cosh(\sigma_{jk}) = a_{jk}, \text{say}, \\ \mathbb{E}\big[\sin(\theta_j) + \sin(\theta_k)\big] &= c_j c_k \sinh(\sigma_{jk}) = b_{jk}, \text{say}, \\ \mathbb{E}\big[\sin(\theta_j) + \cos(\theta_k)\big] &= 0. \end{aligned} \tag{6.34}$$

where

$$\cosh(\sigma_{jk}) = \left(\frac{e^{2\sigma_{jk}}+1}{2e^{\sigma_{jk}}}\right) \quad \text{and} \quad \sinh(\sigma_{jk}) = \left(\frac{e^{2\sigma_{jk}}-1}{2e^{\sigma_{jk}}}\right). \tag{6.35}$$

With the coefficients $\{a_{jk}\}$ and $\{b_{jk}\}$ stored in the matrices

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ \vdots & \ddots & \vdots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ \vdots & \ddots & \vdots & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pp} \end{bmatrix}. \tag{6.36}$$

In matrix form the covariance matrices for the cosines and sines take the form

$$\begin{aligned} \text{var}(\cos\Theta) &= \mathbf{DAD} - \mathbf{cc}^T, \\ \text{var}(\sin\Theta) &= \mathbf{DBD}, \\ \text{cov}(\cos\Theta, \sin\Theta) &= \mathbf{0}. \end{aligned} \tag{6.37}$$

Thus Σ can be recovered from the trigonometric moments through the equation

$$\Sigma_\Theta = \sinh^{-1}(\mathbf{D}^{-1}\text{var}(\sin\Theta)\mathbf{D}^{-1}). \tag{6.38}$$

Here the notation $\sinh^{-1}(\cdot)$ applied to a matrix that the inverse sinh function, $\sinh^{-1}(u) = \log(u + \sqrt{u^2+1}$, is applied to each element of the matrix (Kent and Mardia [49]).

These results suggest a method to estimate Σ from an $n \times p$ matrix of torus data.

(1) Calculate the sample first order trigonometric moments for the p angles, and rotate each angle so that the resultant vector points towards the positive horizontal, x-axis.

(2) Calculate the sample second trigonometric moments corresponding to (6.31) and use (6.35) to produce an estimate of Σ_Θ (Kent and Mardia [49]).

If Σ is small (formally, write $\Sigma = \epsilon\Sigma_0$ for a fixed positive definite matrix Σ_0 and let ϵ get small), then $c_j \approx 1$ for all j and $\Sigma_\Theta \approx \mathbf{B}$. Further the three p-dimensional vectors $\sin\Theta \approx \Theta$ are approximately the same as one another (treating each angle θ_j as a number in $[-\pi, \pi]$, and hence all have approximately the same covariance matrix (Kent and Mardia [49]).

The novelty in this work from Kent and Mardia [49] is to represent the first order and the second order trigonometric moments for the p angles in matrix notation as well as to implement this new method for assessing the efficiency with the application. Table 6.1 displays the simulated

moment estimations for the wrapped bivariate normal torus distribution with various values of $\sigma_{11} = \sigma_{22} = \sigma^2$, ρ and $n = 100$. It is clear from the results that for small or large variances and strong correlations the values of the moment estimators seem close to the true values for the parameters of the wrapped bivariate normal torus distribution.

Parameter	Truth	Moment Estimation	Parameter	Truth	Moment Estimation
σ^2, ρ	1.00, 0.05	0.87, 0.03	σ^2, ρ	10.0, 0.50	9.11, 0.41
σ^2, ρ	1.00, 0.30	0.89, 0.21	σ^2, ρ	10.0, 0.90	9.26, 0.88
σ^2, ρ	10.0, 0.05	8.15, 0.03	σ^2, ρ	50.0, 0.50	45.7, 0.44
σ^2, ρ	10.0, 0.30	8.57, 0.19	σ^2, ρ	50.0, 0.90	48.6, 0.87

Table 6.1: Simulated moment estimations for the parameters of the wrapped bivariate normal torus distribution with a sample of size $n = 100$.

Chapter 7

Conclusions and Future Directions

We subdivided this thesis into three parts or major topics namely, the saddlepoint approximation, new simulation techniques based on concave functions and estimation methods for torus data.

7.1 Saddlepoint Approximation

7.1.1 Motivation

In the first topic we considered the saddlepoint approximations. The motivation for using such an approximation technique is that the saddlepoint approximation has been a valuable tool in the asymptotic analysis and provides an accurate approximation to the density or the distribution of a statistic, even for small tail probabilities and with very small sample sizes. Possibilities of improving the accuracy of the saddlepoint approximation by determining a multiplicative correction to normalize the approximate density to integrate to unity were considered. Another advantage of saddlepoint methods is that the required computational times are essentially negligible compared to simulation (Butler [9], p. 3).

7.1.2 Critical Summary

In Chapter 2 we gave the background to the saddlepoint approximation and we this technique to the normalizing constants of some circular directional distributions such as the von Mises distribution as well as to the normalizing constants for some suitable distributions for spherical and axial data such as the Fisher and the real Bingham distributions.

In a more general setting, it is not always possible to derive an explicit formula for the normalizing constants of some complex shape distributions. In particular, the normalizing constant for the complex Bingham quartic (CBQ) distribution has no known form. In Chapter 3 we explored how the methodology introduced in Chapter 2 can be used to approximate the normalizing constant for the complex Bingham quartic (CBQ) distribution. Two new methods were explored to evaluate this normalizing constant based on the saddlepoint approximations of the Bingham densities namely, the Integrated Saddlepoint (ISP) approximation and the Saddlepoint-Integration (SPI) approximation. Calculating the normalizing constant for the CBQ distribution is based on numerical methods of quadrature.

7.1.3 Future Work

One notable drawback of numerical quadrature is the need to pre-compute (or look up) the requisite weights and nodes. The uniform nodes are not a suitable choice to compute the integrand function for the normalizing constant of the complex Bingham quartic (CBQ) distribution numerically especially under high concentrations. Future work could involve a saddlepoint approximation for the normalizing constant of the complex Bingham quartic (CBQ) distribution using other choices of nodes and weights such as abscissas of the Legendre polynomials and weights depending upon the Legendre polynomials of Gauss-Legendre quadrature method.

7.2 Simulation Techniques

7.2.1 Motivation and Limitation

In the second major topic we attempted to find more efficient simulation methods for sampling from some directional and shape models. Rejection schemes based on concave inequalities are both simple to implement and more reliable. However, the concave inequalities are limited to a critical number of directional models.

7.2.2 Critical Summary

In Chapter 4 we first reviewed the general A/R simulation algorithm. Secondly a general class of inequalities was given based on concave functions. These inequalities were illustrated for the

multivariate normal distribution in $\mathbb{R}^p$ by finding two envelopes, that is the multivariate Cauchy and the multivariate Bilateral exponential distributions. An inequality similar to that was used to show that the angular central Gaussian (ACG) density can be used as an envelope for the Bingham density. The Bingham distribution on S^3 was identified with the matrix Fisher distribution on $SO(3)$. Hence the method of simulation from Bingham distribution analytically gave a method for simulating the matrix Fisher distribution.

The work in Chapter 5 on simulation was carried out to introduce new A/R simulation methods to generate random samples from the von Mises distribution with a Bingham envelope, the Fisher distribution with a Bingham envelope and also the Fisher-Bingham distribution with a Bingham envelope. We also proposed an acceptance-rejection simulation algorithm from the complex Bingham quartic (CBQ) distribution. The problem of simulating from this complex shape distribution reduced to simulation from a mixture of two standard multivariate normal distributions with sensible efficiency.

7.2.3 Future Work

Future work could be done to investigate new simulation methods for the multivariate directional distributions on higher dimension manifold spaces. In particular, distributions on Cartesian products of d copies of spheres S_d^{p-1}, say, such as the Fisher-Bingham distribution on S_d^{p-1}.

7.3 Estimation Methods for Torus Data

7.3.1 Motivation

In the third major topic we attempted to find a new estimation method for the parameters of the wrapped normal torus distribution based on the sample variance-covariances. One reason of choosing the wrapped normal torus distribution is that the trigonometric moments have straightforward explicit expressions.

7.3.2 Critical Summary

In Chapter 6 we considered a review on the sine and cosine bivariate distributions on torus. A comparison was given between three bivariate sine and cosine models based on the contours of the

log-densities. For the wrapped normal torus distribution we developed a new method to estimate the parameters based on the sample sine and cosine moments.

7.3.3 Future Work

Future work could involve a study of the efficiency of the method of sine-cosine moments for the wrapped normal torus distribution compared to other methods such as the maximum likelihood estimators.

Appendix A

Appendix A

A.1 R Functions for Simulating Bingham Distribution using ACG Envelope

```
racg=function(n,PhiI) {
  # simulate n q-vectors from the angular central gaussian distribution
  # by simulating from N_p(0, Sigma) and projecting onto unit sphere
  # PhiI -- inverse covariance matrix, assumed to be symmetric and positive definite
  if(length(PhiI)==1) break("PhiI must have dimension at least 2, as vector or matrix")
  else if(is.vector(PhiI)) Sigmah=diag(sqrt(1/PhiI))
    ee=eigen(PhiI); eval=ee$values; evec=ee$vectors
    Sigmah=evec%*%diag(sqrt(1/eval))%*%t(evec)
    q=nrow(PhiI)
    Y=matrix(rnorm(q*nsim),ncol=q)
    Y=Y%*%Sigmah
    size=sqrt(as.vector((Y^2)%*%rep(1,q)))
    Y=Y*(matrix(1/size,nrow=n,ncol=q)) # Simulation from ACG Density
    Y
}
```

```
rbing=function(n,A) {
  # simulate n q-vectors from the Bingham distribution with minus concentration
  #   matrix A (smallest eigenvalue assumed equal to 0)
  if(is.vector(A)) A=diag(A)
  q=nrow(A); PhiI=diag(q)+2*b^(-1)*A; b = (q + 1)/2; bound=(q/b)^(q/2)*exp(-(q-b)/2)
  accept=0; Yout=matrix(0,nrow=0,ncol=q); rat1=0; nsimtot=0; maxrat=0
  while(accept&nsim) {
    n=nsim-accept; nsimtot=nsimtot+n
    Y=racg(n,PhiI)
    U=runif(n)
    qf1=(Y*(Y%*%A))%*%rep(1,q); qf2=(Y*(Y%*% PhiI))%*%rep(1,q)
    ratio=exp(-qf1)*(qf2^(q/2))/bound; rat1=rat1+sum(ratio&gt)
    maxrat=max(maxrat,max(ratio))
    ar=(U&ratio); Yout=rbind(Yout,Y[ar,]); accept=nrow(Yout)
  }
  cat("rat1:\n")
  cat("maxrat:\n")
  cat("nsimtot = tot number of simulations:\n")
  print(apply(Yout^2,2,mean))
  Yout
}



rbingham.ver3=function(nsim=1,alpha,verbose=FALSE) {
#  compute acceptance ratio for Bingham simulations in R^q.
  alpha=c(alpha,0); q=length(alpha)
  beta=1+2*alpha
  Sigv=1/beta; Sigvh=1/sqrt(beta)
  Y=matrix(rnorm(n*q),ncol=q); U=runif(nsim)
  Y=Y*(matrix(1,nrow=n,ncol=1)%*%t(Sigvh))
  size=sqrt(as.vector((Y^2)%*%rep(1,q)))
  Y=Y*(matrix(1/size,nrow=n,ncol=q)) # Simulation from ACG Density
  qf1=(Y^2)%*%alpha; qf2=(Y^2)%*%beta
```

```
  bound=exp(-qf1+(q-1)/2)*(qf2/q)^(q/2)
  ar=bound
  if(verbose==TRUE) {
    print(alpha)
    print(beta)
    print(Sigv)
    print(Sigvh)
    print(cbind(Y,Y^2%*%rep(1,p)))
    print(cbind(qf1,qf2,exp(-qf1),qf2^(-q/2),bound))
    print(sum(bound & gt,1))
  }
  sum(ar)
}
```

A.2 R Functions for Best-Fisher algorithm of the von Mises Distribution

```
rvm <- function (n, mean, k)
{
    vm <- c(1:n)
    a <- 1 + (1 + 4 * (k^2))^0.5
    b <- (a - (2 * a)^0.5)/(2 * k)
    r <- (1 + b^2)/(2 * b)
    obs <- 1
    while (obs <= n) {
        U1 <- runif(1, 0, 1)
        z <- cos(pi * U1)
        f <- (1 + r * z)/(r + z)
        c <- k * (r - f)
        U2 <- runif(1, 0, 1)
```

```
        if (c * (2 - c) - U2 > 0) {
            U3 <- runif(1, 0, 1)
            vm[obs] <- sign(U3 - 0.5) * acos(f) + mean
            vm[obs] <- vm[obs]%%(2 * pi)
            obs <- obs + 1
        }
        else {
            if (log(c/U2) + 1 - c >= 0) {
                U3 <- runif(1, 0, 1)
                vm[obs] <- sign(U3 - 0.5) * acos(f) + mean
                vm[obs] <- vm[obs]%%(2 * pi)
                obs <- obs + 1
            }
        }
    }
    vm
}

Best.Fisher.efficiency=function(kappa) {
  # analytically compute Best-Fisher efficiency
  tau=1+sqrt(1+4*kappa^2); rho=(tau-sqrt(2*tau))/(2*kappa)
  kr2=2*rho/kappa; omr2=1-rho^2; opr2=1+rho^2; I0=besselI(kappa,0)
  M=(kr2*exp(opr2/kr2-1))/(omr2*I0); Efficiency = 1/M
  cbind(kappa,tau,rho,I0,M, Efficiency)
}
```

A.3 MATLAB Functions for Simulating von Mises-Fisher Distribution

The MATLAB-implemented functions for simulation from VMF distribution and plotting the sample data on a sphere according to Wood's approach are given as follows:

```
function a=vmises3rnd(n,k)
% Generates N directions with spherical von Mises-Fisher distribution (m = 3)
% around the North pole, with concentration K.
% The data matrix A is in standard format.
%
if nargin~=2
    error('Must have two arguments');
end
a=zeros(n,2);
if k<=0
    error('invalid k.');
else
    b=(-2*k+sqrt(4*k^2+4))/2;
    x0=(1-b)/(1+b);
    c=k*x0+2*log(1-x0^2);
for i=1:n

    z=betarnd(1,1); u=rand(1); w=(1-(1+b)*z)/(1-(1-b)*z);
    while k*w+2*log(1-x0*w)-c < log(u)
        z=betarnd(1,1); u=rand(1); w=(1-(1+b)*z)/(1-(1-b)*z);
    end

    v=rand(1,2); v=[1,1]-2*v;
    v=sqrt(1-w^2)*v;
    [a(i,1),a(i,2),r]=cart2sph(v(1,1),v(1,2),w);

end
a=a*180/pi; a=convlat(a); a=convazi(a);
end



function as=convazi(a)
% Converts azimuth data A in [-180, 180]
```

```
% to longitude in [0, 360]
%
n=size(a,1); as=a;
for i=1:n
    if as(i,1)<0
        as(i,1)=360+as(i,1);
    end
end



function as=convlat(a)
% Converts latitude data A(:,2) in [-90, 90]
% to co-latitude in [0, 180] and vice-versa.
%
n=size(a,1); as=a;
for i=1:n
    as(i,2)=90-as(i,2);
end



function polar3d(azi,elev)
%   POLAR3D(AZI,ELEV) plots spherical data on the unitary sphere.
%   using polar coordinates AZImuth E [0,360[ and ELEVation E [-90, 90].
%
%   POLAR3D(COORD) plots spherical data on the unitary sphere
%   using polar coordinates given by COORD. COORD is a matrix with
%   two columns with the following meaning:
%       - COORD(:,1) are the longitudes E [0,360[.
%       - COORD(:,2) are the colatitudes E [0,180].
%
%   Creates toggle buttons for picture rotation and axes labeling.
%
```

```
if (nargin == 1)
    if size(azi,2) ~= 2
        error('Input argument must be a 2-column matrix');
    end
    elev= 90-azi(:,2);
    azi= azi(:,1);
elseif nargin == 2
    if isstr(azi)
        polar3d_call(azi,elev);
        return;
    end
else
    error('Requires 1 or 2 input arguments.')
end

if isstr(azi) | isstr(elev)
    error('Input arguments must be numeric.');
end

if ~ishold
    newplot;
end

h= gcf;
p=get(h, 'position');
p=[p(1) p(2)+p(4)-p(3) p(3) p(3)];
set(h, 'position',p);

% define equator % 101 points %
th = (0:100)'/100*2*pi;
phi= zeros(size(th));
[xequ, yequ, zequ]= sph2cart(th,phi,1);
equ=[xequ yequ zequ];
```

```
% define meridian % 101 points %
th= [zeros(1,50) ones(1,51)*pi]';
phi= [(-25:25)/50*pi (24:-1:-25)/50*pi]';
[xmer, ymer, zmer]= sph2cart(th,phi,1);
mer=[xmer ymer zmer];

% define second meridian % 101 points %
th= [-ones(1,50) ones(1,51)]'*pi/2;
phi= [(-25:25)/50*pi (24:-1:-25)/50*pi]';
[xmer2, ymer2, zmer2]= sph2cart(th,phi,1);
mer2=[xmer2 ymer2 zmer2];

% Conversion from degrees to radians
azi= azi*pi/180;
elev= elev*pi/180;

% transform data to Cartesian coordinates.
[x,y,z]= sph2cart(azi,elev,1);
data=[x(:) y(:) z(:)];

% set view to 3-D and save data as userdata in current axes
view(3);
ud.equ=equ; ud.mer=mer; ud.mer2=mer2; ud.data=data;
set(gca,'userdata',ud);

% create togglebutton for rotation, if don't exist
ui= findobj(gcf,'type','uicontrol','tag','rotpol3');
if isempty(ui)
    ui=uicontrol('units','normalized', 'position', [0.90 0.90 0.05 0.05], ...
        'style','togglebutton', 'callback', 'polar3d(''button'',gcbo)', ...
        'tag','rotpol3','String','R3D');
end
```

```
% create togglebutton to display axes-labels, if don't exist
ui= findobj(gcf,'type','uicontrol','tag','labelpol3');
if isempty(ui)
    ui=uicontrol('units','normalized', 'position', [0.90 0.84 0.05 0.05], ...
        'style','togglebutton', 'String','Label','tag','labelpol3', ...
        'callback', 'polar3d(''label'',gcbo)');
end

% draw data
polar3d_draw;

return;

function polar3d_draw()
% sub function to draw the plot

% define color to draw ALL objects
col= 'k';

% get data from userdata of current axes
ud= get(gca, 'userdata');

% based on view matrix get points in front and back of "monitor"
aa= allchild(gcf);
a= findobj(aa,'flat','type', 'axes', 'tag', 'rotaObj');

% get current view transformation matrix
if ~isempty(a)
    vv=get(a,'view');
else
    vv=get(gca,'view');
end
```

```
v=(vv-[90 0])*pi/180;
[x,y,z]=sph2cart(v(1),v(2),2);

nn=[x y z]';

d=ud.equ*nn;
ind=find(d>=0);
ind=polar3d_ind(ind);
plot3(ud.equ(ind,1),ud.equ(ind,2),ud.equ(ind,3),col)
hold on
ind=find(d<0);
ind=polar3d_ind(ind);
plot3(ud.equ(ind,1),ud.equ(ind,2),ud.equ(ind,3),[col '.'],'markersize',1)

d=ud.mer*nn;
ind=find(d>=0);
ind=polar3d_ind(ind);
plot3(ud.mer(ind,1),ud.mer(ind,2),ud.mer(ind,3),col)
ind=find(d<0);
ind=polar3d_ind(ind);
plot3(ud.mer(ind,1),ud.mer(ind,2),ud.mer(ind,3),[col '.'],'markersize',1)

d=ud.mer2*nn;
ind=find(d>=0);
ind=polar3d_ind(ind);
plot3(ud.mer2(ind,1),ud.mer2(ind,2),ud.mer2(ind,3),col)
ind=find(d<0);
ind=polar3d_ind(ind);
plot3(ud.mer2(ind,1),ud.mer2(ind,2),ud.mer2(ind,3),[col '.'],'markersize',1)

%countour of the sphere
t=viewmtx(vv(1),vv(2));
```

```
d=inv(t)*[ud.equ(:,1)'; ud.equ(:,2)'; zeros(2, size(ud.equ,1))];
plot3(d(1,:),d(2,:),d(3,:), col)

d=ud.data*nn;
ind=find(d>=0);
h=plot3(ud.data(ind,1),ud.data(ind,2),ud.data(ind,3),[col 'o']);
ind=find(d<0);
plot3(ud.data(ind,1),ud.data(ind,2),ud.data(ind,3),[col 'o']);
set(h,'markerfacecolor',get(h,'color'));

%draw axes
plot3([0 1.15 nan 0 0 nan 0 0],[0 0 nan 0 1.15 nan 0 0], [0 0 nan 0 0 nan 0 1.15])
[x,y,z]=cylinder(1:-1:0,20);
x=x/40; y=y/40; z=z/10;
surface(1.15+z,y,x,zeros(2,21,3));
surface(x,1.15+z,y,zeros(2,21,3));
surface(x,y,1.15+z,zeros(2,21,3));
a= findobj(aa,'flat','type', 'uicontrol', 'tag', 'labelpol3');
polar3d('label',a);
hold off;
axis([-1.1 1.1 -1.1 1.1 -1.1 1.1]), axis off;
set(gca,'dataaspectratio',[1 1 1]);
set(gca, 'view', vv);
set(gca, 'userdata',ud);

function polar3d_call(com, handl)
% Processing callback from rotate button

switch (com)
case 'button'
    if get(handl,'value')
        rotate3d on;
        fig=gcbf;
```

```
        set(fig,'WindowButtonUpFcn', ...
            [get(fig,'WindowButtonUpFcn') '; polar3d(''rot'',gcbf)'])
        s=which('scribefiglisten');
        if ~isempty(s)
            scribefiglisten(gcbf,'off');
        end
    else
        rotate3d off;
    end
case 'rot'
    polar3d_draw;
case 'label'
    h= findobj(gca,'type', 'text', 'tag', 'xyzlabel');
    if (get(handl,'value')==get(handl,'max'))  % show axis labels
        if isempty(h)  % create labels
            h(1)=text(0,1.40,0,'Y','HorizontalAlignment','center','tag','xyzlabel');
            h(2)=text(1.40,0,0,'X','HorizontalAlignment','center','tag','xyzlabel');
            h(3)=text(0,0,1.40,'Z','HorizontalAlignment','center','tag','xyzlabel');
        end
        set(h,'visible','on');
    else    % hide axis labels
        if ~isempty(h)
            set(h,'visible','off');
        end
    end

otherwise
    disp('Command unknown');
end

function ind2=polar3d_ind(ind)
% Reordena indices
```

```
dind= diff(ind);
i= find(dind > 1);
if ~isempty(i)
    ind2=ind([(i+1):length(ind) 1:i]);
else
    ind2=ind;
end
```

A.4 Evaluate $\mathbb{E}\big(\cos\theta_1 \sin\theta_2\big) = 0$ for the Bivariate Sine Distribution

If $f_{\text{sin}}(\theta_1, \theta_2)$ is the probability density function for the sine model in (6.11) then

$$\begin{aligned}
\mathbb{E}\big(\cos\theta_1 \sin\theta_2\big) &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \cos\theta_1 \sin\theta_2 \exp\Big\{\kappa_1 \cos\theta_1 + \kappa_2 \cos\theta_2 + \eta \sin\theta_1 \sin\theta_2\Big\} d\theta_1\, d\theta_2 \\
&= \int_{0}^{\pi}\int_{0}^{\pi} \cos\theta_1 \sin\theta_2 \exp\Big\{\kappa_1 \cos\theta_1 + \kappa_2 \cos\theta_2 + \eta \sin\theta_1 \sin\theta_2\Big\} d\theta_1\, d\theta_2 \\
&\quad + \int_{0}^{\pi}\int_{-\pi}^{0} \cos\theta_1 \sin\theta_2 \exp\Big\{\kappa_1 \cos\theta_1 + \kappa_2 \cos\theta_2 + \eta \sin\theta_1 \sin\theta_2\Big\} d\theta_1\, d\theta_2 \\
&\quad + \int_{-\pi}^{0}\int_{0}^{\pi} \cos\theta_1 \sin\theta_2 \exp\Big\{\kappa_1 \cos\theta_1 + \kappa_2 \cos\theta_2 + \eta \sin\theta_1 \sin\theta_2\Big\} d\theta_1\, d\theta_2 \\
&\quad + \int_{-\pi}^{0}\int_{-\pi}^{0} \cos\theta_1 \sin\theta_2 \exp\Big\{\kappa_1 \cos\theta_1 + \kappa_2 \cos\theta_2 + \eta \sin\theta_1 \sin\theta_2\Big\} d\theta_1\, d\theta_2,
\end{aligned}$$

since $f_{\text{sin}}(\theta_1, \theta_2)$ is reflection symmetric about x and y axes. If we change the circular variable θ_1 to $-\theta_1$, we get

$$\begin{aligned}
&\int_{0}^{\pi}\int_{0}^{\pi} \cos\theta_1 \sin\theta_2 \exp\Big\{\kappa_1 \cos\theta_1 + \kappa_2 \cos\theta_2 + \eta \sin\theta_1 \sin\theta_2\Big\} d\theta_1\, d\theta_2 \\
&= \int_{0}^{\pi}\int_{0}^{\pi} -\cos\theta_1 \sin\theta_2 \exp\Big\{\kappa_1 \cos\theta_1 + \kappa_2 \cos\theta_2 - \eta \sin\theta_1 \sin\theta_2\Big\} d\theta_1\, d\theta_2.
\end{aligned}$$

If we change the circular variable θ_2 to $-\theta_2$, we get

$$\begin{aligned}
&\int_{-\pi}^{0}\int_{0}^{\pi} \cos\theta_1 \sin\theta_2 \exp\Big\{\kappa_1 \cos\theta_1 + \kappa_2 \cos\theta_2 + \eta \sin\theta_1 \sin\theta_2\Big\} d\theta_1\, d\theta_2 \\
&= \int_{0}^{\pi}\int_{0}^{\pi} \cos\theta_1 \sin\theta_2 \exp\Big\{\kappa_1 \cos\theta_1 + \kappa_2 \cos\theta_2 - \eta \sin\theta_1 \sin\theta_2\Big\} d\theta_1\, d\theta_2.
\end{aligned}$$

If we change both the circular variable (θ_1, θ_2) to $(\theta_1, -\theta_2)$, we get

$$\begin{aligned}
&\int_{-\pi}^{0}\int_{-\pi}^{0} \cos\theta_1 \sin\theta_2 \exp\Big\{\kappa_1 \cos\theta_1 + \kappa_2 \cos\theta_2 + \eta \sin\theta_1 \sin\theta_2\Big\} d\theta_1\, d\theta_2 \\
&= \int_{0}^{\pi}\int_{0}^{\pi} -\cos\theta_1 \sin\theta_2 \exp\Big\{\kappa_1 \cos\theta_1 + \kappa_2 \cos\theta_2 + \eta \sin\theta_1 \sin\theta_2\Big\} d\theta_1\, d\theta_2.
\end{aligned}$$

Thus, the sum of the four integrals gives

$$\mathbb{E}\big(\cos\theta_1 \sin\theta_2\big) = 0.$$

Bibliography

[1] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions: with Formulas, Graphs and Mathematical Tables*. Dover, New York, 1972.

[2] A. C. Atkinson. The simulation of generlized inverse Gaussian and hyperbolic random variates. *SIAM Journal of Scientific and Statistical Computation*, 3:502–515, 1982.

[3] C. Auderset, C. Mazza, and E. A. Ruh. Angular gaussian and cauchy estimation. *Journal of Multivariate Analysis*, 93:180–197, 2005.

[4] J. Besag. Statistical analysis of non-lattice data. *The Statistician*, 24:179–195, 1975.

[5] D. J. Best and N. I. Fisher. Efficient simulation of the von Misses distribution. *Journal of Applied Statistics*, 28:152–157, 1979.

[6] P. Billingsley. *Probability and Measure*. John Wiley & Sons, New York, 3rd edition, 1995.

[7] C. Bingham. An antipodally stmmetric distribution on the sphere. *The Annals of Statistics*, 2:1201–1225, 1974.

[8] R. Brent. *Algorithms for Minimization without Derivatives*. Englewood Cliffs, NJ: Prentice-Hall, 1973.

[9] R. W. Butler. *Saddlepoint Approximations with Applications*. Cambridge University Press, Cambridge, 2007.

[10] S. Coles. Inference for circular distributions and process. *Statistics and Computing*, 8:105–113, 1998.

[11] M. J. Crawley. *The **R** Book*. John Wiley & Sons Ltd, Chichester, 2007.

[12] J. Dagpunar. *Principels of random variate generations*. Oxford University Press, New York, 1988.

[13] J. S. Dagpunar. *Methods for generating variates from probability distributions*. PhD thesis, Mathematics, Brunel University, unpublished PhD Thesis, 1983.

[14] H. E. Daniels. Saddlepoint approximations in statistics. *The Annals of Mathematical Statistics*, 25, 1954.

[15] H. A. David and H. N. Nagaraja. *Order Statistics*. John Wiley & Sons, Inc., 2003.

[16] T. D. Downs. Orientation statistics. *Biometrika*, 53:665–676, 1972.

[17] I. L. Dryden and K. V. Mardia. General shape distributions in a plane. *Advances in Applied Probability*, 23:259–276, 1991b.

[18] I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis*. John Wiley and Sons, Chichester, 1998.

[19] F. Esscher. On the probability function in the collective theory of risk. *Skand. akt. Tidsskr*, 15:78–86, 1932.

[20] B. Everitt and D. J. Hand. *Finite Mixture Distributions*. Chapman and Hall, 1st edition, 1981.

[21] W Feller. *An Introduction to Probability Theory and its Applications*. John Wiley & Sons, New York, 1971.

[22] N. I. Fisher, T. Lewis, and B.J.J. Embleton. *Statistical Analysis of Spherical Date*. Cambridge University Press, Cambridge, 1987.

[23] N.I. Fisher, T. Lewis, and M. E. Willcox. Tests of Discordancy for samples from Fisher's distribution on the sphere. *Journal of Applied Statistics*, 30(3):230–237, 1981.

[24] D. Germund and Å. Björck. *Numerical Methods in Scientific Computing*, volume I. Society for Industrial Mathematics, 2008.

[25] M. Gharib. Two characterisations of a gamma mixture distribution. *Bull. Australian Mathematics Society*, 52:353–359, 1995.

[26] C. R. Goodall and K. V. Mardia. The noncentral Bartlett decomposition and shape densities. *Journal of Multivatiate Analysis*, 40:94–108, 1992.

[27] C Goutis and G. Casella. Explaining the saddlepoint approximation. *American Statistical Association*, 53:216–224, 1999.

[28] P. J. Green and K. V. Mardia. Bayesian alignment using hierarchical models, with applications in protein bioinformatics. *Biometrika*, 93:235–254, 2006.

[29] P. D. Hoff. Simulation of the matrix bingham-von mises-fisher distribution with applications to multivariate and relational data. *Journal of Computational and Graphical Statistics*, 18:438–456, 2009.

[30] S. Huzurbazer. Practical saddlepoint approximations. *American Statistical Association*, 53:225–232, 1999.

[31] S. R. Jammalamadaka and A SenGupta. *Topics in Circular Statistics*. World Scientific Publishing Co. Pte. Ltd, Singapore, 2001.

[32] M. E. Johnson. *Multivariate Statistical Simulation*. John Wiley & Sons, Inc., Canada, 1987.

[33] N. L. Johnson and S. Kotz. *Distributions in Statistics: Continuous Multivariate Distributions*. John Wiley & Sons, Inc., New York, 1972.

[34] R. A. Johnson and T. Wehrly. Measures and models for angular correlation and angular-linear correlation. *Journal of Royal Statistical Society*, 39 (2):222–229, 1977.

[35] O. Jones, R. Maillardet, and A. Robinson. *Introduction to Scientific Programming and Simulation Using R*. Chapman & Hall/CRC, 2009.

[36] D. G. Kendall. Shape manifolds, Procrustean metrics and complex projective spaces. *Bulletin of the London Mathematical Society*, 1984.

[37] D. G. Kendall. Exact distributions for shapes of random triangles in convex sets. *Advances in Applied Probability*, 17:308–329, North Ireland, 1985.

[38] D. G. Kendall, D. Bardon, T. K. Carne, and H. Le. *Shape and Shape Theory.* John Wiley & Sons, New York, 1st edition, 1999.

[39] W. J. Kennedy and J. E. Gentle. *Statistical Computing.* Marcel Dekker, New York, 1980.

[40] J. T. Kent. The Fisher-Bingham distribution on the sphere. *Journal of Royal Statistical Society*, B(44):71–80, 1982.

[41] J. T. Kent. The complex Bingham distribution and shape analysis. *Journal of Royal Statistical Society*, B(56):285–299, 1994.

[42] J. T. Kent. Current issues for statistical inference in shape analysis. In K.V.Mardia and C.A. Gill, editors, *Proceedings in Current Issues in Statistical Shape Analysis*, pages 167–175, Leeds, 1995. University of Leeds Press.

[43] J. T. Kent, D. L. Constable, and Fikret ER. Simulation for the complex Bingham distribution. *Journal of Statistics and Computing*, 14:53–57, 2004.

[44] J. T. Kent and Tyler. D. E. Maximum likelihood estimation for the wrapped Cauchy distribution. *Journal of Applied Statistics*, 15:247–254, 1988.

[45] J. T. Kent and T. Hamelryck. Using the Fisher-Bingham distribution in stochastic models for protein structure. In S. Barber P.D. Baxter, K.V.Mardia and R.E. Walls, editors, *Quantitative Biology, Shape Analysis, and Wavelets*, pages 57–60. University of Leeds Press, 2005.

[46] J. T. Kent, K. V. Mardia, and P. McDonnell. The complex Bingham quartic distribution and shape analysis. *Journal of Royal Statistical Society*, B(68):747–765, 2006.

[47] J. T. Kent, K. V. Mardia, and J. S. Rao. A characterization of the uniform distribution on the circle. *Journal of Annulal Statistics*, 7:882–889, 1979.

[48] J. T. Kent, K. V. Mardia, and C. C. Taylor. Modelling strategies for bivariate circular data. In A. Gusnanto S. Barber, P.D. Baxter and K.V. Mardia, editors, *The Art and Science of Statistical Bioinformatics*, pages 70–74. University of Leeds Press, 2008.

[49] J. T. Kent and Mardia. K. V. Principal component analysis for the wapped normal torus model. In C. J. Fallaize P.S. Holmes A. Gusnanto K.V. Mardia, J. T. Kent and J. Voss,

editors, *Statistical Tools for Challenges in Bioinformatics*, pages 39–41. University of Leeds Press, 2009.

[50] J. Kolassa. *Series Approximation Methods in Statistics*. Springer-Verlag, New York, 1994.

[51] J. E. Kolassa. *Series Approximation Methods in Statistics*. Springer Science & Business Media, Inc., Third Edition, 2006.

[52] A. Kume and S. G. Walker. Sampling from compostional and directional distributions. *Statistics and Computing*, 3:261–265, 2006.

[53] A. Kume and S. G. Walker. On the Fisher-Bingham distribution. *Statistics and Computing*, 19:167–172, 2009.

[54] A. Kume and A. T. A. Wood. Saddlepoint approximations for the Bingham and Fisher-Bingham normalising constants. *Biometrika Trust*, 92(2):465–476, 2005.

[55] D. Kuonen. Saddlepoint approximations for distributions of duadratic forms in normal variables. *Biometrika Trust*, 4:929–935, 1999.

[56] S. R. Lele and J. T. Richtsmeier. *An Invariant Approach to Statistical Analysis of Shapes*. Chapman and Hall/CRC, USA, 2001.

[57] K. Mammasis and R. W. Stewart. The FB5 distribution and its application in wireless communications. In *2008 International ITG Workshop on Smart Antennas*, pages 375–381, 2008.

[58] B. F. J. Manley. *Multivariate Statistical Methods: A Primer*. Chapman and Hall, 3rd edition, 2004.

[59] K. V. Mardia. *Statistics of Directional Data*. Academic Press INC. (London) LTD, 1972.

[60] K. V. Mardia. Statistics of directional data (with discussion). *Journal of Royal Statistical Society*, B(37):349–393, 1975a.

[61] K. V. Mardia. Shape analysis of triangles through directional techniques. *Journal of Royal Statitical Society*, 51:449–458, 1989.

[62] K. V. Mardia. The complex Watson distribution and shape analysis. *Journal of Royal Statistical Society*, B(61):913–926, 1994.

[63] K. V. Mardia. Directional statistics and shape analysis. *Journal of Applied Statistics*, 26:949–957, 1999.

[64] K. V. Mardia. On some recent advancements in applied shape analysis and directional statistics. In S. Barber and K.V.Mardia, editors, *Statistical Tools for Challenges in Bioinformatics*, pages 9–17, Leeds, 2007. University of Leeds Press.

[65] K. V. Mardia. Statistical complexity in protein Bioinformatics. In K.V.Mardia A Gusnanto and C. J. Fallaize, editors, *Statistical Tools for Challenges in Bioinformatics*, pages 9–20, Leeds, 2009. University of Leeds Press.

[66] K. V. Mardia and I. L. Dryden. Shape distributions for landmark data. *Advances in Applied Probability*, 21:742–755, 1989a.

[67] K. V. Mardia and I. L. Dryden. Statistical analysis of shape data. *Biometrica*, 76:271–281, 1989b.

[68] K. V. Mardia, G. Hughes, and C. C. Taylor. A multivariate von Mises distribution with applications to Bioinformatics. *Canadian Journal of Statistical Science*, 36:99–109, 2008.

[69] K. V. Mardia, G. Hughes, and C. C. Taylor. Efficiency of the pseudolikelihood for multivariate normal and von Mises distributions. *Submitted*, 2011.

[70] K. V. Mardia and P. E. Jupp. *Directional Statistics*. John Wiley and Sons, Chichester, 2000.

[71] K. V. Mardia and J. T. Kent. Maximum likelihood estimation using composite likelihoods for closed exponential families. *Biometrika*, 4:21–30, 2009.

[72] K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*. San Diego: Academic Press, 1979.

[73] K. V. Mardia and J. Voss. Some fundamental properties of a multivariate von Mises distribution. *Submitted*, 2012.

[74] K.V. Mardia and V. Patrangenaru. Directions and projective shapes. *Annals of Statistics*, 33:1666–1699, 2005.

[75] K.V. Mardia, C. C. Taylor, and Subramaniam. G. K. Bivariate von Mises densities for angular data with applications to protein bioinformatics. *Annals of Statistics*, 35:166–180, 2009.

[76] K.V. Mardia, C. C. Taylor, and Subramaniam. G. K. Protein bioinformatics and mixtures of bivariate von Mises distributions for angular data. *Biometrics*, (in press).

[77] P. McDonnell. *Planar Shape Distributions and Image Averaging.* PhD thesis, Mathematics, University of Leeds, unpublished PhD Thesis, 2001.

[78] V.V. Nikulin and I. R. Shafarevich. *Geometries and Groups.* Springer, USA, 1987.

[79] M. S. Paolella. *Intermediate Probability: A computational Approach.* John Wiley and Sons, New York, 2007.

[80] M. J. Prentice. Orientation statistics without parametric assumptions. *J. Roy. Statist. Soc*, 48:214–222, 1986.

[81] N Ravishanker and D. K. Dipak. *A First Course in Linear Model Theory.* Chapman and Hall, London, 2001.

[82] N. Reid. Saddlepoint methods and statistical inference (with Discussion). *Journal of Statistical Science*, 3:213–238, 1988.

[83] A. C. Rencher. *Methods of Multivariate Analysis.* John Wiley and Sons, Canada, 2nd edition, 2002.

[84] L. P. Rivest. A distribution for dependent unit vectors. *Communications in Statistics*, A17:461–483, 1988.

[85] M. L. Rizzo. *Statistical Computing with* ***R***. Chapman & Hall/CRC, London, 2008.

[86] Y. A. Sahalia and J. Yu. Saddlepoint approximations for continuous-time markov processes. *Journal of Econometrics*, 134:507–551, 2006.

[87] J. L. Schiff. *The Laplace Transform: Theory and Applications.* Springer-Verlag, New York, 1999.

[88] J. Seigerstetter. Discussion of Kendall, D. G. (1974), Pole-seeking Brownian motion and bird navigation. *Journal of Royal Statistical Society*, B36:411–412, 1974.

[89] H. Singh, V. Hnizdo, and E. Demchuk. Probabilistic model for two dependent circular variables. *Biometrika*, 89:719–723, 2002.

[90] J. Stalker. *Complex Analysis: Fundamentals of the Classical Theory of Functions*. Springer, 1998.

[91] H. Tanizaki. *Computational Methods in Statistics and Econometrics*. CRC Press, 2004.

[92] H. Scheffé. *The Analysis of Variance*. John Wiley and Sons, New York, 1959.

[93] D. E. Tyler. Statistical analysis for the angular central Gaussian distribution on the sphere. *Biometrika*, 74:579–589, 1987.

[94] G. Ulrich. Computer generation of distributions on the m-sphere. *Journal of Applied Statistics*, 2:158–163, 1984.

[95] D. A. Walker. The standard error of a proportion for different scores and test length. *Practical Assessment Research and Evaluation*, 10(5):12–20, 2005.

[96] G. S. Watson. *Statistics On Sphere*. John Wiley and Sons, Canada, 1983.

[97] A. Wintner. *The Fourier Transformations of Probability Distributions*. Johns Hopkins University, Baltimore, 1947.

[98] A. T. A. Wood. The simulation of spherical distributions in the Fisher-Bingham family. *Journal of Communications, Statistics, Simulation and Computation*, 16:885–898, 1987.

[99] A. T. A. Wood. Estimation of the concentration parameters of the Fisher matrix disribution on SO(3) and the Bingham distribution on S^q, $q > 2$. *Australian Journal of Statistics*, 35:69–79, 1993.

[100] A. T. A. Wood. Simulation of the von Mises-Fisher distribution. *Journal of Communications, Statistics, Simulation and Computation*, 23:157–164, 1994.

[101] L. Yuan and J. D. Kalbeisch. On the Bessel distribution and related problems. *Annals of the Institute of Statistical Mathematics*, 52:438–447, 2000.