

# Multivariate and Time Series Models for Circular Data with Applications to Protein Conformational Angles

Gareth Hughes

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

The University of Leeds  
Department of Statistics

January 2007

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. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

# Abstract

Over the past few years, the research area of bioinformatics has been the subject of increasing attention. At its core remains what is often termed the “holy grail” of structural biology: the prediction of a protein’s three dimensional structure from its associated amino acid sequence. Bioinformatics is an interdisciplinary field of research incorporating, among others, biochemists, statisticians and computer scientists. The international collaboration involved is exemplified by the Leeds Annual Statistical Research (LASR) workshops ([www.maths.leeds.ac.uk/statistics/workshop](http://www.maths.leeds.ac.uk/statistics/workshop)), yearly meetings of such researchers in which new ideas are shared.

The research contained in this thesis focuses on the development of circular multivariate and time series models and their application to the three-dimensional conformation of proteins. The aims of the work presented herein are therefore twofold: we strive to advance the scope of circular statistics both theoretically and also in practice. In terms of the former, a multivariate circular distribution is proposed and investigated and circular time series models are developed and explored. With respect to the latter, all the models considered are applied to the conformational angles of particular proteins, and their suitability discussed in detail.

Although the thesis focuses primarily on circular data, due to the close relationship between highly concentrated circular data and data on the real line, aspects of the work in this thesis inevitably concentrate on the latter as a limiting case of the former. In these cases, the results obtained for the linear case are clearly of interest in their own right, as well as being important to the circular cases to which they relate.

# Acknowledgements

My two lead supervisors, Kanti Mardia and Charles Taylor, have provided invaluable support throughout the course of my PhD studies. They have made themselves available and offered assistance even at the busiest of times. Their contributions to both the work in this thesis and to my experiences in Leeds are greatly appreciated. I also wish to thank John Kent (Department of Statistics, University of Leeds) for his helpful comments at annual reviews.

As mentioned in the abstract, bioinformatics is an interdisciplinary research area. This thesis has benefited from the valuable contributions of Thomas Hamelryck (The Bioinformatics Centre, University of Copenhagen), Walter Gilks (Centre for Statistical Bioinformatics, Department of Statistics, University of Leeds) and David Westhead (School of Biochemistry and Microbiology, University of Leeds).

Last, but not least, I would like to thank my girlfriend Rita for providing continued encouragement, my fellow PhD students for making my time at Leeds an enjoyable one, and my family for their support.

This PhD was funded by an EPSRC studentship.





# Contents

Abstract . . . . .	ii
Acknowledgements . . . . .	iii
Contents . . . . .	v
List of figures . . . . .	xi
List of tables . . . . .	xvii
 <b>1 Introduction</b>	 <b>1</b>
1.1 Motivation and background . . . . .	1
1.2 Introduction to protein structure . . . . .	2
1.2.1 The protein backbone . . . . .	2
1.2.2 The protein side chain . . . . .	6
1.3 Univariate circular statistics and the von Mises distribution . . . . .	7
1.3.1 Univariate circular statistics . . . . .	7
1.4 Bivariate von Mises models . . . . .	11
1.5 Outline of thesis . . . . .	15
1.5.1 Outline of Part I . . . . .	16

1.5.2	Outline of Part II . . . . .	17
<b>I</b>	<b>Multivariate von Mises models with application to protein conformational angles</b>	<b>19</b>
<b>2</b>	<b>A multivariate von Mises distribution</b>	<b>21</b>
2.1	Introduction . . . . .	21
2.2	The model . . . . .	22
2.3	Model properties . . . . .	23
2.4	Simulating data . . . . .	26
2.5	Conclusions . . . . .	30
<b>3</b>	<b>Efficiency of the pseudolikelihood for multivariate normal distributions</b>	<b>33</b>
3.1	Introduction . . . . .	33
3.2	Efficiency of the pseudolikelihood . . . . .	35
3.3	Multivariate normal with unknown $\sigma^2$ and $\rho$ . . . . .	36
3.3.1	Full pseudolikelihood . . . . .	37
3.3.2	Partial pseudolikelihood . . . . .	40
3.4	Loss of efficiency with known unit variances . . . . .	42
3.5	Efficiency for the case of unequal variances and correlations . . . . .	43
3.5.1	Efficiency for the bivariate normal distribution . . . . .	45
3.5.2	Efficiency for the trivariate normal distribution . . . . .	47

3.6	Comparison of estimating equations . . . . .	48
3.7	Conclusions . . . . .	51
<b>4</b>	<b>Parameter estimation and inference for the multivariate von Mises distribution</b>	<b>53</b>
4.1	Introduction . . . . .	53
4.2	Efficiency of the pseudolikelihood for the bivariate von Mises distribution	55
4.2.1	Fisher information for the full likelihood approach . . . . .	55
4.2.2	Information matrix for the pseudolikelihood approach . . . . .	57
4.2.3	Efficiency of the pseudolikelihood . . . . .	58
4.2.4	Simulation . . . . .	59
4.3	Data simulation and parameter estimation for the trivariate von Mises distribution . . . . .	61
4.3.1	Data simulation and pseudolikelihood estimation . . . . .	63
4.3.2	Pseudolikelihood versus full likelihood . . . . .	67
4.4	Application to protein data . . . . .	69
4.4.1	Gamma turn data . . . . .	70
4.4.2	Serine and valine data . . . . .	77
4.5	Conclusions . . . . .	89

## **II Time series models with application to protein conformational**

<b>angles</b>	<b>91</b>
<b>5 Directional time series models</b>	<b>93</b>
5.1 Introduction . . . . .	93
5.2 Directional time series and exploratory analysis . . . . .	94
5.3 The von Mises autoregressive process . . . . .	95
5.4 Linked processes . . . . .	97
5.5 Wrapped processes . . . . .	98
5.6 Projected processes . . . . .	100
5.7 Choosing between the models . . . . .	101
5.8 Conclusions . . . . .	101
<b>6 Sine and Cosine time series models</b>	<b>103</b>
6.1 Introduction . . . . .	103
6.2 The Sine time series model . . . . .	105
6.2.1 The model . . . . .	105
6.2.2 The deterministic component . . . . .	107
6.2.3 Data simulation . . . . .	110
6.2.4 Likelihood function and parameter estimation . . . . .	113
6.2.5 Application to protein data . . . . .	114
6.2.6 Conclusions . . . . .	121
6.3 The Cosine time series model . . . . .	122

6.3.1	The model . . . . .	122
6.3.2	Relationship with the Cosine model . . . . .	123
6.3.3	Model properties . . . . .	126
6.3.4	Data simulation . . . . .	129
6.3.5	Likelihood function and parameter estimation . . . . .	133
6.3.6	Application to protein data . . . . .	136
6.3.7	Conclusions . . . . .	140
6.4	Further work . . . . .	140
<b>7</b>	<b>A Möbius time series model</b>	<b>143</b>
7.1	Introduction . . . . .	143
7.2	The model . . . . .	144
7.3	Likelihood function . . . . .	146
7.4	Special values of $\omega$ . . . . .	147
7.5	The deterministic component of the model . . . . .	149
7.6	Simulating data . . . . .	150
7.7	Checking time-reversibility . . . . .	153
7.8	The equilibrium distribution . . . . .	154
7.8.1	Numerical approach . . . . .	156
7.8.2	Analytic approach . . . . .	164
7.9	Estimating model parameters . . . . .	167

7.9.1	Grid search and likelihood plots . . . . .	167
7.9.2	Gradient descent . . . . .	170
7.10	Application to protein data . . . . .	171
7.11	Conclusions . . . . .	178
<b>8</b>	<b>Conclusions</b>	<b>181</b>
8.1	Multivariate and time series models for circular data . . . . .	182
8.2	Modeling protein data with the multivariate and time series models . . . .	184
	<b>Bibliography</b>	<b>185</b>

# List of figures

1.1	The carboxyl group (COOH) of amino acid 1 condenses with the amine group (NH <sub>2</sub> ) of amino acid 2, eliminating H <sub>2</sub> O and forming a peptide bond between the C atom of amino acid 1 and the N atom of amino acid 2. The process continues to produce a polypeptide. . . . .	3
1.2	Dihedral angle $\theta$ defined in terms of four atoms $A_1, A_2, A_3$ (directly behind $A_2$ ) and $A_4$ . The angle $\theta$ is that between bonds $b_1$ and $b_3$ . . . . .	4
1.3	Ramachandran plot for the protein triosephosphate isomerase. Regions labelled $\beta$ , $\alpha_R$ and $\alpha_L$ indicate those in which the $(\phi, \psi)$ angles of $\beta$ strands, right-handed $\alpha$ helices and left-handed $\alpha$ helices can be observed. . . . .	6
1.4	Comparison of von Mises densities (solid lines) and normal densities (dashed lines) for $\kappa = 0.5, 1, 2, 4$ . . . . .	12
2.1	Univariate and bivariate marginals for trivariate von Mises distribution with $(\kappa_1, \kappa_2, \kappa_3, \lambda_{12}, \lambda_{13}, \lambda_{23})$ given above each plot. . . . .	27
2.2	Univariate and bivariate marginals for trivariate von Mises distribution with $(\kappa_1, \kappa_2, \kappa_3, \lambda_{12}, \lambda_{13}, \lambda_{23})$ given above each plot. . . . .	27
2.3	Target (black) and candidate (red) densities for four parameter configurations, and the efficiency of data simulation for the bivariate von Mises distribution. . . . .	30

3.1	Efficiency of FPL and PPL for multivariate normal distribution (with known unit variances) as a function of $\rho \in [0, 1]$ and for $p = 3, 5, 8, 10$ (top to bottom). . . . .	44
3.2	Efficiency of FPL and PPL for multivariate normal distribution (with known unit variances) as a function of $\rho \in (-1/(p-1), 0)$ and for $p = 3, 5, 8, 10$ (left to right). . . . .	44
4.1	Efficiency of pseudolikelihood for bivariate von Mises distribution with $\kappa = 1, 2, \dots, 15, \lambda = 0.5, 1, 2, \dots, 15$ . . . . .	60
4.2	Efficiency of pseudolikelihood for bivariate von Mises distribution with $\kappa = 1, 2, \dots, 10,  \lambda  = 0.5, 1, 2, 4, 6, 8, 10$ . . . . .	60
4.3	Estimates of bias and variance for pseudolikelihood and full likelihood estimators for the bivariate von Mises distribution. Based on 10 000 simulations for each configuration and samples of sizes 20, 50 and 100. . .	62
4.4	Matrix plot of $\phi$ angles of gamma turn data, with circular plots on main diagonal, pairwise plots on upper panels and correlations on lower panels	71
4.5	Matrix plot of $\psi$ angles of gamma turn data, with circular plots on main diagonal, pairwise plots on upper panels and correlations on lower panels	71
4.6	Circular plot of 497 simulated $\phi$ and $\psi$ triplets with true parameter values equal to the MLEs for the original gamma turn data. . . . .	73
4.7	Correlation plot of raw serine data, with circular plots on main diagonal, pairwise plots on upper panels and correlations on lower panels. . . . .	79
4.8	Correlation plot of raw valine data, with circular plots on main diagonal, pairwise plots on upper panels and correlations on lower panels. . . . .	79



4.9	Correlation plot of selected serine data, with circular plots on main diagonal, pairwise plots on upper panels and correlations on lower panels.	80
4.10	Correlation plot of selected valine data, with circular plots on main diagonal, pairwise plots on upper panels and correlations on lower panels.	80
4.11	Correlation plot of simulated serine data, with circular plots on main diagonal, pairwise plots on upper panels and correlations on lower panels.	84
4.12	Correlation plot of simulated valine data, with circular plots on main diagonal, pairwise plots on upper panels and correlations on lower panels.	84
6.1	Plots of $\mu_t$ versus $\theta_{t-1}$ for parameter configurations $(\lambda, \kappa) = (\pm 4, 1)$ , $(\pm 7, 8)$ , $(\pm 4, 8)$ and $(\pm 1, 8)$ for the Sine time series model. . . . .	108
6.2	Simulated data for the Sine time series model. Values of $(\lambda, \kappa)$ are given in brackets above each plot. . . . .	111
6.3	Profiles of loglikelihood function for simulated data sets in the vicinity of the true parameter values. True values of $(\lambda, \kappa)$ are given in brackets above each plot. . . . .	115
6.4	$\phi$ and $\psi$ conformational angles for a protein data set, and simulated $\phi$ and $\psi$ values based on maximum likelihood estimates for the true data. . . . .	117
6.5	Plots of $\theta_t$ versus $\theta_{t-1}$ , a smoothed version (red line) and $\hat{\mu}_t \pm 1/\sqrt{\hat{\kappa}_t}$ , where $\theta$ is one of $\phi, \phi^{\text{sim}}, \psi$ or $\psi^{\text{sim}}$ . . . . .	120
6.6	Marginal kde (solid) and pdf (dashed) based on data and MLEs respectively.	120
6.7	Plots of $\Pi(\theta)$ versus $\theta$ for $a = -2, b = 3$ (unimodal–solid line) and $a = -4, b = 3$ (bimodal–dashed line) . . . . .	125
6.8	Plots of $\mu_t(\pm 1/\sqrt{\kappa_t})$ versus $\theta_{t-1}$ for (a) $a = b$ , (b) $a = 2.1, b = 2$ , (c) $a = 3, b = 3.5$ and (d) $a = 3, b = 20$ . . . . .	126

6.9	Simulated unimodal Cosine time series data for the values of $(a, b)$ given above each plot. . . . .	130
6.10	Simulated bimodal Cosine time series data. Left: Time series plots. Middle: $\theta_t$ for $t$ even (solid) and $t$ odd (dashed). Right: $\mu_t \pm 1/\sqrt{\kappa_t}$ for the given $(a, b)$ values . . . . .	132
6.11	Profile loglikelihood plots. (A) and (D): $l(\boldsymbol{\theta}; a, b_0, \mu_0)$ for $a \in [a_0 - 2, a_0 + 2]$ ; (B) and (E): $l(\boldsymbol{\theta}; a_0, b, \mu_0)$ for $b \in [0, 3]$ ; (C) and (F) $l(\boldsymbol{\theta}; a, b, \mu_0)$ for $a \in [a_0 - 2, a_0 + 2]$ and $b \in [0, 3]$ . . . . .	135
6.12	$\phi$ and $\psi$ conformational angles for a protein data set, and simulated $\phi$ and $\psi$ values based on maximum likelihood estimates for the true data using the Cosine time series model. . . . .	137
6.13	Plots of $\theta_t$ versus $\theta_{t-1}$ , a smoothed version (red line) and $\hat{\mu}_t \pm 1/\sqrt{\hat{\kappa}_t}$ , where $\theta$ is one of $\phi, \phi^{\text{sim}}, \psi$ or $\psi^{\text{sim}}$ . . . . .	138
6.14	Marginal kde (solid) and pdf (dashed) based on data and MLEs respectively.	138
7.1	Plots of $\theta_t$ versus $\theta_{t-1}$ based on Equation (7.4) and for $\alpha = 0$ (both plots), $\omega = 0.2$ (left) and $\omega = 0.6$ (right). . . . .	150
7.2	Plots of simulated data for $\alpha = 0$ and those values of $\omega$ and $\kappa$ given above each plot. . . . .	152
7.3	Plots of simulated data for $\alpha = \pi/2$ , $\kappa = 4$ and those values of $\omega$ above each plot. . . . .	153
7.4	Plots of $LHS - RHS$ of Equation (7.19) versus $\boldsymbol{\theta} = (-\pi, \pi)$ for $(\omega, \kappa) = (0.3, 6)$ (left) and $(\omega, \kappa) = (0.6, 3)$ (right). Plot is for that value of $j$ leading to the largest absolute value on the $y$ -axis of the plot. . . . .	155

- 7.5 Comparison of numerical solution to the equilibrium distribution (solid line) based on initial parameter values  $\omega = 0.3$ ,  $\kappa_1 = 4$  and the von Mises density (dashed line) with  $\mu = 0$  and  $\kappa = 4$ . . . . . 157
- 7.6 Comparisons of numerical solutions (NS) to the equilibrium distribution and kde's based on samples from the equilibrium distribution, based on initial parameters ( $\omega = 0.6$ ,  $\kappa_1 = 3$ ) (left) and ( $\omega = -0.1$ ,  $\kappa_1 = 5$ ) (right). . . . . 159
- 7.7 Polar plot based on the right hand plot of Figure 7.6 comparing the numerical solution (NS) to the equilibrium distribution, based on initial values  $\omega = -0.1$ ,  $\kappa_1 = 5$ , with a kernel density estimate (KDE) based on a sample from the equilibrium distribution with these parameter values. . . 160
- 7.8 Left: comparison of numerical solution (NS) for equilibrium distribution based on the values ( $\omega = 0.6$ ,  $\kappa_1 = 4$ ) and the von Mises pdf (VM) with parameters ( $\mu = 0$ ,  $\kappa_2 = 2.4$ ). Right: Plots of  $\kappa_1$  versus  $\kappa_2^*$  for  $\omega = 0, 0.2, 0.4, 0.6, 0.8$  (top line to bottom line). . . . . 161
- 7.9 Left: Residuals versus fitted values for regression model  $\hat{\kappa}_2^* = 0.997 - 1.048\kappa_1\omega^2$ , with a symbol for each  $\kappa = 1, 3, 5, 7, 9$ . Right: Residuals versus fitted values for five models of the form  $\kappa_2^* = \alpha\kappa_1 + \beta\kappa_1\omega^2$ , with  $\kappa_1 = 1, 3, 5, 7, 9$  (increasing from left to right). . . . . 163
- 7.10 Left: Image plot of Equation (7.33) for  $\kappa_1 = 1, \dots, 10$  and  $\omega = 0, 0.1, \dots, 0.8$ . Right: Lines representing  $\kappa_2 = \kappa_1(1 - \omega^2)$  for  $\kappa_1 = 1, \dots, 10$  (x-axis) and  $\omega = 0, 0.2, 0.4, 0.6, 0.8$  (top line to bottom line), with corresponding  $\kappa_2^*$  superimposed as points. . . . . 166
- 7.11 Likelihood plots based on simulated data set 1, in which  $\alpha_0 = \pi/4$ ,  $\omega_0 = 0.6$ ,  $\kappa = 4$ ,  $n = 200$ . (a) Profile in  $\omega$  with  $\alpha = \alpha_0$ . (b) Profile in  $\alpha$  with  $\omega = \omega_0$ . (c) Profile in  $\alpha$  with  $\omega = -\omega_0$ . (d) Contour likelihood plot. . . . 168

- 7.12 Likelihood plots based on simulated data set 2, in which  $\alpha_0 = 2\pi/3$ ,  $\omega_0 = -0.4$ ,  $\kappa = 4$ ,  $n = 500$ . (a) Profile in  $\omega$  with  $\alpha = \alpha_0$ . (b) Profile in  $\alpha$  with  $\omega = \omega_0$ . (c) Profile in  $\alpha$  with  $\omega = -\omega_0$ . (d) Contour likelihood plot. . . . 169
- 7.13 Circular and time series plots of the  $\phi$  and  $\psi$  angles of a polypeptide chain, and of data simulated using the parameter estimates obtained by fitting the Möbius model to the protein data. . . . . 175
- 7.14 Profile loglikelihood plots for protein and simulated data. Profiles in each of  $\alpha$  and  $\omega$  taken at the maximum likelihood estimate of the other. . . . 176
- 7.15 Plots of  $\theta_t$  versus  $\theta_{t-1}$ , a smoothed version (red line) and  $\hat{\mu}_t \pm 1/\sqrt{\hat{\kappa}_t}$ , where  $\theta$  is one of  $\phi$ ,  $\phi^{\text{sim}}$ ,  $\psi$  or  $\psi^{\text{sim}}$ . . . . . 177
- 7.16 Numerically calculated equilibrium distribution based on parameter estimates (dashed lines) and kernel density estimates based on data (solid lines). . . . . 178

# List of tables

1.1	Specification of the dihedral angles $\phi_i$ , $\psi_i$ and $\omega_i$ in terms of the angle $\theta$ and atoms $A_1$ to $A_4$ of Figure 1.2. Atom $A_3$ is directly behind $A_2$ . . . . .	5
1.2	Necessary and sufficient conditions for bimodality of joint and marginal densities for the Sine and Cosine models. . . . .	14
4.1	Mean values and standard deviations of nlm estimates for generated trivariate von Mises data. . . . .	65
4.2	Mean values and standard deviations of nlm estimates for generated trivariate von Mises data with $\boldsymbol{\mu} = (2, 3, -1)$ . . . . .	65
4.3	Mean values and standard deviations of nlm estimates for generated trivariate von Mises data, once marginal circular means have been subtracted from the data columns. . . . .	66
4.4	Mean values and standard deviations of nlm estimates for generated trivariate von Mises data with $\kappa_1 = 1$ , $\kappa_2 = 1.5$ , $\kappa_3 = 0.5$ and $\lambda_{12} = \lambda_{13} = \lambda_{23} = 1$ . . .	67
4.5	Mean values and standard deviations of nlm estimates for generated trivariate von Mises data with $\kappa_1 = 0.5$ , $\kappa_2 = 0.75$ , $\kappa_3 = 0.25$ and $\lambda_{12} = \lambda_{13} = \lambda_{23} = 1$ .	67
4.6	Mean values and approximate standard errors of pseudolikelihood and full likelihood estimates for trivariate von Mises data. . . . .	68

4.7	Mean values and approximate standard errors of pseudolikelihood and full likelihood estimates for trivariate von Mises data. . . . .	68
4.8	Marginal MLEs of $\mu$ and $\kappa$ for gamma turn data. . . . .	72
4.9	Maximum likelihood estimates (MLEs) and pseudolikelihood estimates (PLEs) (and their standard errors) for $\phi$ and $\psi$ angles of Gamma turn data and for data simulated using these estimates. . . . .	72
4.10	Loglikelihood and log-pseudolikelihood values evaluated at MLEs and PLEs, and test statistics based on these values. . . . .	77
4.11	Univariate von Mises MLEs of $\mu$ and $\kappa$ for serine and valine data . . . . .	81
4.12	Maximum likelihood estimates (MLEs) and pseudolikelihood estimates (PLEs) (and their standard errors) of parameters for serine and valine data, and for data simulated using these estimates. . . . .	82
4.13	Estimates of $\kappa$ and $\lambda$ parameters for serine and valine data based on moment estimator of covariance matrix of the normal distribution. . . . .	85
4.14	Loglikelihood values and test statistics based on four different approaches. . . . .	87
6.1	Mean values (and standard deviations) of estimates for simulated Sine time series data. . . . .	113
6.2	Estimates (and standard errors) of Sine time series model parameters for protein and simulated data. . . . .	116
6.3	Probabilities, for various $a$ and $b$ values, that any particular $\theta$ value will fall in the tail beyond the turning points. . . . .	129

6.4	Mean values (and standard deviations) of estimates of $a$ , $b$ and $\mu$ for 10 combinations of $(a_0, b_0)$ and with $\mu_0 = 0$ , each based on 100 simulated data sets of 200 observations. Estimates obtained using the <code>nlm</code> function in R. . . . .	134
6.5	Estimates (and approximate standard errors) of Cosine time series parameters for protein and simulated data. . . . .	136
7.1	Mean values (and one minus mean resultant length) of estimates of $\alpha$ for 8 combinations of $(\alpha_0, \kappa_0, \omega_0)$ , each based on 100 simulated data sets of 400 observations. . . . .	172
7.2	Mean values (and standard deviations) of estimates of $\omega$ for 8 combinations of $(\alpha_0, \kappa_0, \omega_0)$ , each based on 100 simulated data sets of 400 observations. . . . .	172
7.3	Mean values (and standard deviations) of estimates of $\kappa$ for 8 combinations of $(\alpha_0, \kappa_0, \omega_0)$ , each based on 100 simulated data sets of 400 observations. . . . .	172
7.4	Maximum likelihood estimates (and estimated standard errors) for protein data and for data simulated based on these estimates. Also given are mean directions $\bar{\theta}$ for each data set, and AR(1) parameter estimates of $Y_t = \tan(\theta_t - \hat{\alpha})/2$ , where $\theta$ is to be replaced by the variable of interest. . . .	173





# Chapter 1

## Introduction

### 1.1 Motivation and background

The research presented in this thesis is motivated by the increasingly popular area of structural bioinformatics, and in particular by the challenge of understanding the three dimensional structure of proteins. The so-called “protein folding problem”, that of predicting the three dimensional structure of a protein given its amino acid sequence, is one of the most fundamental unsolved challenges in structural biology. Also of great interest in the study of proteins is the relationship between their structure and function. The link between the amino acid sequence and the function of a protein is therefore its three dimensional structure, and it is a better understanding of this structure that motivates the research presented herein.

The work is partitioned into two separate yet connected statistical methodologies. The first, discussed in Part I of the thesis, is the formulation and application of multivariate circular models; the second, the focus of Part II, is the development and implementation of time series models for circular data. The application of each is to the conformational angles of protein data. We therefore unify in the present chapter the topics that recur

throughout the thesis. In particular, we give an introduction to protein structure in Section 1.2, outline certain aspects of univariate circular statistics (Section 1.3) and, in Section 1.4, give a summary of two bivariate circular distributions known as the Sine model and the Cosine model that will be revisited throughout the thesis. Section 1.5 gives a more detailed breakdown of the structure of the thesis.

## 1.2 Introduction to protein structure

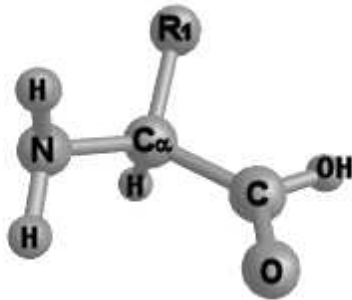
We here give a brief outline of the fundamental aspects of protein structure that will be important in the application to protein data of the statistical models formulated in the thesis. There are a number of excellent texts giving comprehensive introductions to protein structure, just two of which are Branden and Tooze (1998) and Lesk (2000).

There are twenty commonly occurring amino acids, each with a central carbon atom ( $C^\alpha$ ) and, attached to it, a hydrogen atom, an amine group of atoms, a carboxyl group of atoms and a side chain. It is the latter that enables us to distinguish between amino acids. The carboxyl group of one amino acid condenses with the amine group of another, creating a peptide bond between the two amino acids (see Figure 1.1). This process continues, giving a structural sequence called a polypeptide chain. One or more polypeptide chains then bind together to form a protein. Two aspects of this protein will be of particular interest in the current research; the backbone of the protein and the side chains of the amino acids.

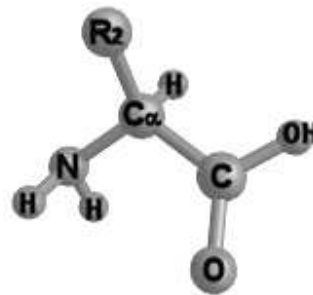
### 1.2.1 The protein backbone

The backbone of a polypeptide chain comprises a sequence of atoms

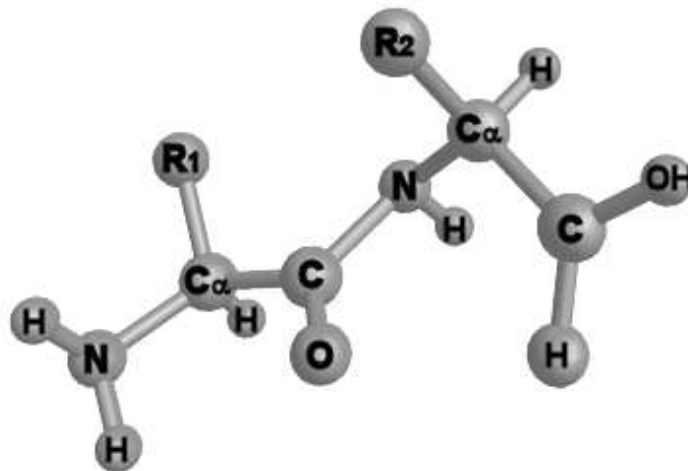
$$N_1 - C_1^\alpha - C_1 - N_2 - C_2^\alpha - C_2 - \dots - N_p - C_p^\alpha - C_p,$$



(a) Amino acid 1.



(b) Amino acid 2.



(c) Amino acids 1 and 2 joined by a peptide bond.

Figure 1.1: The carboxyl group (COOH) of amino acid 1 condenses with the amine group (NH<sub>2</sub>) of amino acid 2, eliminating H<sub>2</sub>O and forming a peptide bond between the C atom of amino acid 1 and the N atom of amino acid 2. The process continues to produce a polypeptide.

in which the lengths of the bonds between any two successive atoms and the angle between any three successive atoms are, to a good approximation, constant (Lesk, 2000). The degrees of freedom of the polypeptide therefore involve angles within chains of 4 atoms. Since there are 3 different atoms in the backbone (N, C and  $C^\alpha$ ), there are 3 angles to be considered. The situation is summarised in Figure 1.2. We label the atoms  $A_1$ ,  $A_2$ ,  $A_3$  and  $A_4$  and the bonds  $b_1$ ,  $b_2$  and  $b_3$ . Looking at  $A_2$  in such a way that  $A_3$  and  $b_2$  are blocked from view by  $A_2$ , and  $A_1$  is directly below  $A_2$  (see Figure 1.2), the dihedral angle in each case is that between bonds  $b_1$  and  $b_3$ . A zero direction is observed when  $b_1$  eclipses  $b_3$ , whilst a counter clockwise direction is taken as positive. Angles are measured between  $-\pi$  and  $\pi$ .

Table 1.1 defines the dihedral angles  $\phi_i$ ,  $\psi_i$  and  $\omega_i$  in terms of the atoms  $A_1$  to  $A_4$ . For example, when  $A_1 = C_{i-1}$ ,  $A_2 = N_i$ ,  $A_3 = C_i^\alpha$  and  $A_4 = C_i$  then the angle  $\theta$  depicted in Figure 1.2 represents  $\phi_i$ .

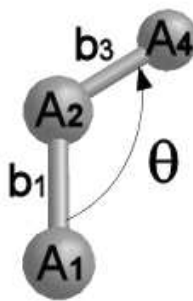


Figure 1.2: Dihedral angle  $\theta$  defined in terms of four atoms  $A_1$ ,  $A_2$ ,  $A_3$  (directly behind  $A_2$ ) and  $A_4$ . The angle  $\theta$  is that between bonds  $b_1$  and  $b_3$ .

Note from the above that  $\phi_1$  and  $\psi_p$  are undefined. The angle  $\omega$  is restricted to be about zero. Most combinations of  $\phi$  and  $\psi$  angles in proteins are not observed in practice as

$\theta$	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>4</sub>
$\phi_i$	C <sub>i-1</sub>	N <sub>i</sub>	C <sub>i</sub> <sup>α</sup>	C <sub>i</sub>
$\psi_i$	N <sub>i</sub>	C <sub>i</sub> <sup>α</sup>	C <sub>i</sub>	N <sub>i+1</sub>
$\omega_i$	C <sub>i-1</sub> <sup>α</sup>	C <sub>i-1</sub>	N <sub>i</sub>	C <sub>i</sub> <sup>α</sup>

Table 1.1: Specification of the dihedral angles  $\phi_i$ ,  $\psi_i$  and  $\omega_i$  in terms of the angle  $\theta$  and atoms A<sub>1</sub> to A<sub>4</sub> of Figure 1.2. Atom A<sub>3</sub> is directly behind A<sub>2</sub>.

they would result in steric collisions between backbone and side chain atoms.

A Ramachandran plot (Ramachandran et al., 1963) is commonly used to plot  $\psi$  versus  $\phi$  for amino acids or proteins, and can also be used to highlight the  $\phi$  and  $\psi$  combinations observable for the components of a protein called alpha helices and beta strands. Figure 1.3 shows a Ramachandran plot for the protein triosephosphate isomerase. As can be seen from the plot, the  $\phi$ ,  $\psi$  pairs tend to separate into clusters. In general the possible  $\phi$ ,  $\psi$  values for  $\beta$  strands, right-handed  $\alpha$  helices and left-handed  $\alpha$  helices for any protein cluster around the regions on Figure 1.3 labelled  $\beta$ ,  $\alpha_R$  and  $\alpha_L$  respectively. The amino acid glycine is an exception to the rule: its side chain is a single hydrogen atom, meaning that a wider range of  $\phi$ ,  $\psi$  values are possible without causing steric collisions. Most proteins contain only small quantities of glycine.

Recent examples of the modeling of protein conformational angles include Singh et al. (2002), who fit the bivariate von Mises model called the Sine model to angles of the amino acid proline, and Mardia et al. (2007b), who discuss more broadly the Cosine and Sine models and model the  $\phi$  and  $\psi$  angles of the proteins malate dehydrogenase and myoglobin using mixture models.

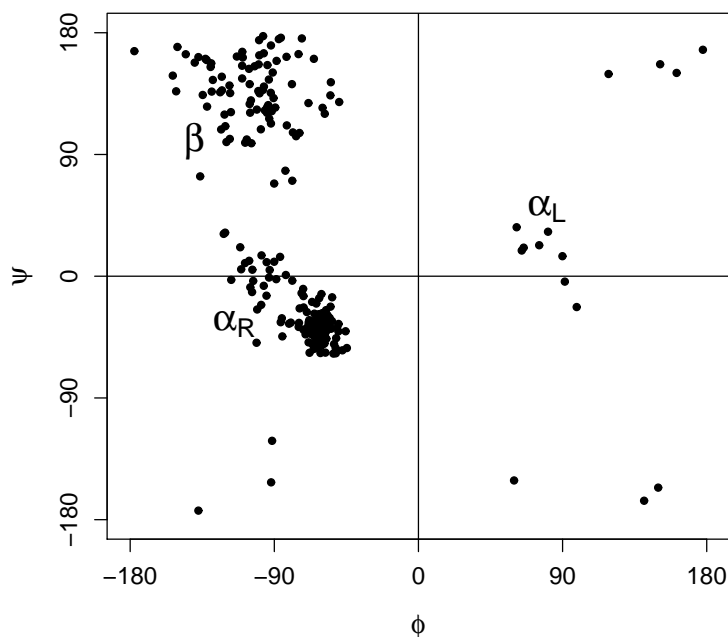


Figure 1.3: Ramachandran plot for the protein triosephosphate isomerase. Regions labelled  $\beta$ ,  $\alpha_R$  and  $\alpha_L$  indicate those in which the  $(\phi, \psi)$  angles of  $\beta$  strands, right-handed  $\alpha$  helices and left-handed  $\alpha$  helices can be observed.

### 1.2.2 The protein side chain

The carbon atoms on the side chain of an amino acid are labelled according to the letters of the Greek alphabet. The letter  $\alpha$  is given to the central carbon atom of the amino acid ( $C^\alpha$ ). Each carbon atom on the side chain is then allocated a letter sequentially,  $C^\beta, C^\gamma, \dots$ , and the dihedral angles involving these side chain carbon atoms are denoted  $\chi_1, \chi_2, \dots$ . Each amino acid therefore has associated with it a certain number of  $\chi$  values, depending on the number of carbon atoms in its side chain.  $\chi_1$  for example can be defined in terms of Figure 1.2 by taking  $A_1 = N$ ,  $A_2 = C^\alpha$ ,  $A_3 = C^\beta$  and  $A_4 = C^\gamma$ . The angle  $\theta$  in the figure then represents  $\chi_1$ . Two amino acids that each have a single side chain angle  $\chi_1$  are serine and valine.

There has been a vast amount of research investigating the relationship between the side chain  $\chi$  angles and the  $\phi$  and  $\psi$  angles of amino acids and proteins. Dunbrack and Cohen (1997), for example, present a Bayesian analysis of side chain conformations in order to predict  $\chi$  values from the associated  $\phi$  and  $\psi$  angles.

## 1.3 Univariate circular statistics and the von Mises distribution

Circular statistics occur in very many and diverse settings. In nature, the homing or migration direction of birds are but two examples. In medicine and biology, features that occur periodically, such as month of onset of an illness, or circadian rhythms, can be regarded as measurements on the circle. In psychology the perception of direction and in astronomy the relative orbit of planets are two further examples. A contemporary example that is of particular significance with the vast interest in bioinformatics, and that will play a central role in this thesis, is the three-dimensional structure of proteins, defined in terms of dihedral angles.

In order to both highlight some of the aspects of statistical modeling of circular data and to lay the foundation for the multivariate and time series models discussed in the thesis, we here give an introduction to univariate circular statistics and the von Mises distribution. For a more detailed outline than is given here and a more general discussion of directional statistics see, for example, Mardia and Jupp (1999).

### 1.3.1 Univariate circular statistics

Consider a circular data set  $\theta_1, \theta_2, \dots, \theta_n$  where  $\theta_i \in [-\pi, \pi)$ ,  $i = 1, 2, \dots, n$  (in which interval we shall take circular variables to be throughout the thesis). Each  $\theta_i$  has associated

with it a unit vector  $\mathbf{x}_i = (\cos \theta_i, \sin \theta_i)$ ,  $i = 1, \dots, n$ . The mean direction  $\bar{\theta}$  of  $\theta_1, \dots, \theta_n$  is defined as the direction of the resultant of the vectors  $\mathbf{x}_1, \dots, \mathbf{x}_n$ . Defining

$$C = \sum_{i=1}^n \cos \theta_i \quad \text{and} \quad S = \sum_{i=1}^n \sin \theta_i$$

we see that the length of this resultant is

$$R = \sqrt{C^2 + S^2}$$

and that  $\bar{\theta}$  is the solution of the equations

$$C = R \cos \bar{\theta}, \quad S = R \sin \bar{\theta}.$$

Taking  $\bar{\theta}$  to be in the interval  $[-\pi, \pi)$ , this solution is given (assuming  $S \neq 0$ ,  $C \neq 0$ ) by the function

$$\bar{\theta} = \tan^{-1}(S, C) = \begin{cases} \tan^{-1}(S/C) & \text{if } C > 0 \\ \tan^{-1}(S/C) + \pi \operatorname{sgn}(S) & \text{if } C < 0 \end{cases},$$

where  $\operatorname{sgn}(\cdot)$  is the sign function. If  $C = 0$ ,  $S \neq 0$  then  $\bar{\theta} = \frac{\pi}{2} \operatorname{sgn}(S)$  whilst if  $S = 0$ ,  $C \neq 0$  then  $\bar{\theta} = \frac{\pi}{2} [\operatorname{sgn}(C) - 1]$ . We therefore see that the function  $\tan^{-1}(a, b) \in [-\pi, \pi)$  is the angle between the positive  $x$ -axis and the vector  $(b, a)$ .

The mean resultant length is given by  $\bar{R} = R/n$ , and  $\bar{R}$  and  $\bar{\theta}$  are the polar coordinates of the sample first trigonometric moment

$$\mathbf{m}'_1 = \bar{C} + i\bar{S} = \bar{R}e^{i\bar{\theta}} = \hat{\rho}e^{i\hat{\mu}},$$

where  $\bar{C} = C/n$ ,  $\bar{S} = S/n$  and  $\rho$  and  $\mu$  are the polar coordinates of the population first trigonometric moment. The sample  $p$ th trigonometric moment is defined analogously as

$$\mathbf{m}'_p = a_p + ib_p = \bar{R}_p e^{im_p}, \quad (1.1)$$

say, where

$$a_p = \frac{1}{n} \sum_{i=1}^n \cos p\theta_i, \quad b_p = \frac{1}{n} \sum_{i=1}^n \sin p\theta_i, \quad (1.2)$$



$\overline{R}_p$  is the length of the  $p$ th mean resultant and  $m_p$  is its direction. Clearly  $a_1 = \overline{C}$ ,  $b_1 = \overline{S}$ ,  $\overline{R}_1 = \overline{R}$  and  $m_1 = \overline{\theta}$ . From (1.1) we have

$$a_p = \overline{R}_p \cos m_p, \quad b_p = \overline{R}_p \sin m_p. \quad (1.3)$$

The  $p$ th trigonometric moment about the sample mean direction  $\overline{\theta}$  is defined as

$$\mathbf{m}_p = \overline{a}_p + i\overline{b}_p, \quad (1.4)$$

where

$$\overline{a}_p = \frac{1}{n} \sum_{i=1}^n \cos p(\theta_i - \overline{\theta}), \quad \overline{b}_p = \frac{1}{n} \sum_{i=1}^n \sin p(\theta_i - \overline{\theta}). \quad (1.5)$$

Expanding the expressions in (1.5) and using the relations (1.2) and (1.3) for  $a_p$  and  $b_p$  gives

$$\overline{a}_p = \overline{R}_p \cos(m_p - p\overline{\theta}), \quad \overline{b}_p = \overline{R}_p \sin(m_p - p\overline{\theta}), \quad (1.6)$$

so that, from (1.4),

$$\mathbf{m}_p = \overline{R}_p e^{i(m_p - p\overline{\theta})}.$$

We now consider the population equivalents of the sample trigonometric moments. Unlike distributions on the line, circular distributions are uniquely defined by their moments. In particular, the characteristic function  $\phi$  evaluated at integer  $p$  is also the  $p$ th trigonometric moment  $\mu'_p$  of  $\theta$ :

$$\mu'_p = \phi_p = E(e^{ip\theta}) = \alpha_p + i\beta_p = \rho_p e^{i\mu_p}, \quad (1.7)$$

say, where

$$\alpha_p = E(\cos p\theta), \quad \beta_p = E(\sin p\theta).$$

The  $p$ th central trigonometric moment is defined as

$$\mu_p = E \{ e^{ip(\theta - \mu_1)} \} = \overline{\alpha}_p + i\overline{\beta}_p.$$

Using (1.7) we have

$$\mu_p = \rho_p e^{i(\mu_p - p\mu_1)},$$

giving

$$\overline{\alpha}_p = \rho_p \cos(\mu_p - p\mu_1), \quad \overline{\beta}_p = \rho_p \sin(\mu_p - p\mu_1).$$

The most widely used circular distribution is the von Mises distribution, which enjoys a similar status in circular statistics to that of the normal distribution for statistics on the line. The connection between the two is one that is exploited throughout the thesis, and is worth highlighting in the simplest, univariate case.

**Definition 1.3.1** *The random angle  $\theta$  is said to follow a von Mises distribution with mean direction  $\mu$  and “concentration”  $\kappa \geq 0$  if its density (with respect to the uniform measure on the circle) is*

$$f(\theta) = \{2\pi I_0(\kappa)\}^{-1} \exp\{\kappa \cos(\theta - \mu)\}, \quad (1.8)$$

where  $I_\nu(\cdot)$  is the modified Bessel function of the first kind and order  $\nu$ . In this case we write  $\theta \sim M(\mu, \kappa)$ .

Alternatively, we can define the von Mises distribution in terms of the vector  $\mathbf{x}$ :

**Definition 1.3.2** *The unit random vector  $\mathbf{x} = (\cos \theta, \sin \theta)^T$  has a von Mises distribution with mean direction  $\mu$  and “concentration”  $\kappa \geq 0$  if its density (with respect to the uniform measure on the circle) is*

$$f(\mathbf{x}) = [2\pi I_0(\kappa)]^{-1} \exp(\mathbf{x}^T \boldsymbol{\alpha}), \quad (1.9)$$

where the concentration vector  $\boldsymbol{\alpha} = \kappa(\cos \mu, \sin \mu)^T$  has length  $\kappa$  and angle  $\mu$ , and  $I_\nu(\cdot)$  is the modified Bessel function of the first kind and order  $\nu$ . In this case we write  $\mathbf{x} \sim VM(\boldsymbol{\alpha})$  (in order to distinguish between this definition and that given in Definition 1.3.1).

From Equation (1.8) and since  $dI_0(y)/dy = I_1(y)$  it is easily seen that the maximum likelihood estimates of  $\mu$  and  $\kappa$  are given by

$$\hat{\mu} = \overline{\theta} \quad \text{and} \quad A_1(\hat{\kappa}) = \overline{R}$$

respectively, where  $A_1(\cdot) = I_1(\cdot)/I_0(\cdot)$ .

Under high concentration (ie. if  $\kappa$  is large), we have  $\theta - \mu \approx 0$  so that  $\cos(\theta - \mu) \approx 1 - (\theta - \mu)^2/2$  using a second order Taylor series expansion. From Equation (1.8) we therefore have

$$\begin{aligned} f(\theta) &\approx [2\pi I_0(\kappa)]^{-1} \exp \left\{ \kappa - \kappa(\theta - \mu)^2/2 \right\} \\ &\propto \exp \left\{ -\kappa(\theta - \mu)^2/2 \right\}, \end{aligned}$$

which is proportional to the density of a normal distribution with mean  $\mu$  and variance  $1/\kappa$ . In other words, if  $\theta \sim M(\mu, \kappa)$  and  $\kappa$  is large, then  $\theta \sim N(\mu, 1/\kappa)$  approximately. Figure 1.4 compares the probability density functions for  $M(\mu, \kappa)$  and  $N(\mu, 1/\kappa)$  for  $\mu = 0$  and  $\kappa = 0.5, 1, 2$  and  $4$ . As can be seen from the plots, the approximation of the latter to the former improves as  $\kappa$  increases, and for  $\kappa = 4$  the approximation can be seen to be reasonably good.

There has been much work on the possible extension of the univariate von Mises distribution to higher dimensions. In the next section we outline some of these, focusing on two bivariate von Mises distributions referred to as the Sine model and the Cosine model, which will be central to much of the research in the thesis.

## 1.4 Bivariate von Mises models

Mardia (1975) proposed a bivariate distribution for two circular variables  $\theta_1$  and  $\theta_2$  which, with zero directional mean vector, has probability density function proportional to

$$\exp \left[ \kappa_1 \cos \theta_1 + \kappa_2 \cos \theta_2 + \{ \cos \theta_1, \sin \theta_1 \} \mathbf{A} \{ \cos \theta_2, \sin \theta_2 \}^T \right] \quad (1.10)$$

where  $\mathbf{A}$  is a  $2 \times 2$  matrix. Rivest (1987) considered a submodel of (1.10), the density of which is proportional to

$$\exp \left[ \kappa_1 \cos \theta_1 + \kappa_2 \cos \theta_2 + \alpha \cos \theta_1 \cos \theta_2 + \beta \sin \theta_1 \sin \theta_2 \right]. \quad (1.11)$$

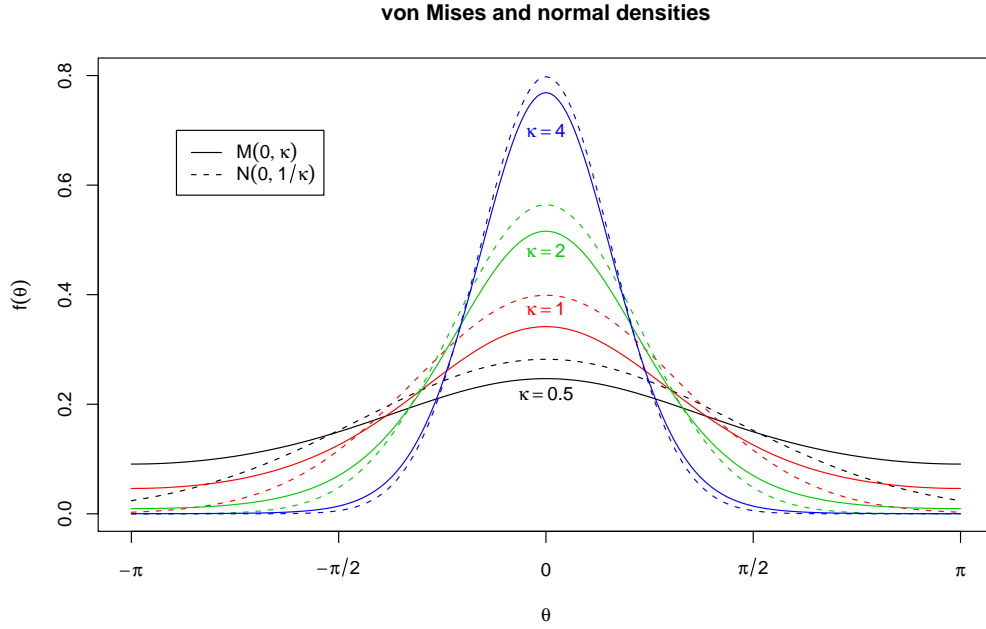


Figure 1.4: Comparison of von Mises densities (solid lines) and normal densities (dashed lines) for  $\kappa = 0.5, 1, 2, 4$ .

Including mean direction parameters, this model has six parameters. In order to obtain a distribution that is approximately normal under high concentration, we require further restrictions on  $\alpha$  and  $\beta$ , since the bivariate normal distribution has five parameters.

With the further restriction  $\alpha = 0$  and with  $\beta = \lambda$  we have the so-called Sine model, studied in detail by Singh et al. (2002) and Mardia et al. (2007b), the density of which, including mean directions, is given by

$$f_s(\theta_1, \theta_2) = C_s \exp \left\{ \kappa_1 \cos(\theta_1 - \mu_1) + \kappa_2 \cos(\theta_2 - \mu_2) + \lambda \sin(\theta_1 - \mu_1) \sin(\theta_2 - \mu_2) \right\}, \quad (1.12)$$

for  $-\pi \leq \theta_1, \theta_2 < \pi$ , where  $\kappa_1, \kappa_2 \geq 0$ ,  $-\infty < \lambda < \infty$ ,  $-\pi \leq \mu_1, \mu_2 < \pi$  and

$$C_s^{-1} = 4\pi^2 \sum_{m=0}^{\infty} \binom{2m}{m} \left( \frac{\lambda^2}{4\kappa_1\kappa_2} \right)^m I_m(\kappa_1) I_m(\kappa_2).$$

Alternatively, taking  $\alpha = \beta = -\kappa_3$  gives the density of the so-called Cosine model, studied in depth by Mardia et al. (2007b):

$$f_c(\theta_1, \theta_2) = C_c \exp \left\{ \kappa_1 \cos(\theta_1 - \mu_1) + \kappa_2 \cos(\theta_2 - \mu_2) - \kappa_3 \cos[(\theta_1 - \mu_1) - (\theta_2 - \mu_2)] \right\}, \quad (1.13)$$

for  $-\pi \leq \theta_1, \theta_2 < \pi$ , where  $\kappa_1 \geq \kappa_3 \geq 0$ ,  $\kappa_2 \geq \kappa_3 \geq 0$ ,  $-\infty < \lambda < \infty$ ,  $-\pi \leq \mu_1, \mu_2 < \pi$  and

$$C_c^{-1} = 4\pi^2 \left\{ I_0(\kappa_1)I_0(\kappa_2)I_0(\kappa_3) + 2 \sum_{m=1}^{\infty} I_m(\kappa_1)I_m(\kappa_2)I_m(\kappa_3) \right\}.$$

The joint distribution of  $\theta_1$  and  $\theta_2$  is approximately normal based on both  $f_s(\theta_1, \theta_2)$  and  $f_c(\theta_1, \theta_2)$  for highly concentrated data, so long as the joint circular distribution in question is unimodal, the conditions for which are discussed below. For the models given by Equations (1.12) and (1.13), the parameters  $\lambda$  and  $\kappa_3$  account for the circular dependency between  $\theta_1$  and  $\theta_2$ .

The marginal and conditional densities of the Sine and Cosine models are derived in Singh et al. (2002) and Mardia et al. (2007b), and we here quote these densities for future reference. For the Sine model, the marginal density of  $\theta_1$  is given by

$$f_s(\theta_1) = 2\pi C_s I_0(a_1) e^{\kappa_1 \cos(\theta_1 - \mu_1)} \quad (1.14)$$

where  $a_1 = \{\kappa_2^2 + \lambda^2 \sin^2(\theta_1 - \mu_1)\}^{1/2}$ . For the Cosine model, the corresponding result is

$$f_c(\theta_1) = 2\pi C_c I_0(\kappa_{23}) e^{\kappa_1 \cos(\theta_1 - \mu_1)}$$

where  $\kappa_{23}^2 = \kappa_2^2 + \kappa_3^2 - 2\kappa_2\kappa_3 \cos(\theta_1 - \mu_1)$ .

The conditional density of  $\theta_2$  given  $\theta_1$  for the Sine model is

$$\begin{aligned} f_s(\theta_2|\theta_1) &= [2\pi I_0(a_1)]^{-1} e^{a_1 \cos(\theta_2 - \mu_2 - b_1)} \\ &= [2\pi I_0(a_1)]^{-1} e^{\kappa_2 \cos(\theta_2 - \mu_2) + \lambda \sin(\theta_1 - \mu_1) \sin(\theta_2 - \mu_2)} \end{aligned} \quad (1.15)$$

where  $\tan b_1 = \frac{\lambda}{\kappa} \sin(\theta_1 - \mu_1)$ . For the Cosine model, we have

$$f_c(\theta_2|\theta_1) = [2\pi I_0(\kappa_{23})]^{-1} e^{\kappa_{23} \cos(\theta_2 - \mu_2 - c_1)} \quad (1.16)$$

$$= [2\pi I_0(\kappa_{23})]^{-1} e^{\kappa_{23} \cos(\theta_2 - \mu_2) - \kappa_3 \cos[(\theta_1 - \mu_1) - (\theta_2 - \mu_2)]} \quad (1.17)$$

where  $\tan c_1 = -\kappa_3 \sin(\theta_1 - \mu_1) / [\kappa_2 - \kappa_3 \cos(\theta_1 - \mu_1)]$ . We therefore see that for both the Sine model and the Cosine model, the conditional density of  $\theta_2$  given  $\theta_1$  is univariate von Mises. The same is true of  $\theta_1$  given  $\theta_2$ , for which we have similar expressions to those given above.

Singh et al. (2002) considered the shape of the marginal distribution  $f_s(\theta_1)$  for the Sine model, whilst Mardia et al. (2007b) explored the shapes of the marginal and joint densities  $f_c(\theta_1)$  and  $f_c(\theta_1, \theta_2)$  for the Cosine model, and the shape of the joint density  $f_s(\theta_1, \theta_2)$  for the Sine model. In fact all of these densities are symmetric and either unimodal or bimodal. Table 1.2 gives the condition, for each, under which the density is bimodal.

	Sine model	Cosine model
Joint	$\kappa_1 \kappa_2 < \lambda^2$	$\kappa_3 > \kappa_1 \kappa_2 / (\kappa_1 + \kappa_2)$
Marginal	$A_1(\kappa_2) > \kappa_1 \kappa_2 / \lambda^2$	$A_1( \kappa_1 - \kappa_3 ) > \frac{ \kappa_1 - \kappa_3  \kappa_2}{\kappa_1 \kappa_3}$

Table 1.2: Necessary and sufficient conditions for bimodality of joint and marginal densities for the Sine and Cosine models.

For the marginal densities  $f_s(\theta_1)$  and  $f_c(\theta_1)$ , the unique mode is at  $\mu_1$  if the density is unimodal. If the density is bimodal, then the modes are located at  $\mu_1 \pm \theta_1^*$ , where  $\theta_1^*$  is defined as follows. For the Sine model,  $\theta_1^*$  is the solution of

$$\cos(\theta_1^* - \mu_1) A_1(a_1^*) / a_1^* = \kappa_1 / \lambda^2, \quad (1.18)$$

where  $a_1^* = \{\kappa_2^2 + \lambda^2 \sin^2(\theta_1^* - \mu_1)\}^{1/2}$ . For the Cosine model,  $\theta_1^*$  is the solution to

$$\kappa_2 \kappa_3 A_1(\kappa_{23}^*) / \kappa_{23}^* - \kappa_1 = 0, \quad (1.19)$$

where  $\kappa_{23}^* = \{\kappa_2^2 + \kappa_3^2 - 2\kappa_2\kappa_3 \cos(\theta_1^* - \mu_1)\}^{1/2}$ .

When the joint densities are unimodal the Sine and Cosine models can, under high concentration, be approximated by normal distributions with common mean vector  $(\mu_1, \mu_2)$  and inverse covariance matrices given respectively by

$$\Sigma_s^{-1} = \begin{bmatrix} \kappa_1 & -\lambda \\ -\lambda & \kappa_2 \end{bmatrix} \quad \text{and} \quad \Sigma_c^{-1} = \begin{bmatrix} \kappa_1 - \kappa_3 & \kappa_3 \\ \kappa_3 & \kappa_2 - \kappa_3 \end{bmatrix}.$$

The correlations between  $\theta_1$  and  $\theta_2$  for the Sine and Cosine models based on these approximations are

$$\rho_s = \frac{\lambda}{\sqrt{\kappa_1\kappa_2}} \quad \text{and} \quad \rho_c = \frac{-\kappa_3}{\sqrt{(\kappa_1 - \kappa_3)(\kappa_2 - \kappa_3)}}$$

respectively. Mardia et al. (2007b) compared  $\rho_s$  and  $\rho_c$  with the correlations of  $(\cos \theta_1, \cos \theta_2)$  and  $(\sin \theta_1, \sin \theta_2)$  for various large values of  $\kappa_1$  and  $\kappa_2$ , concluding that  $\kappa_3$  appears to “track” the value of  $\rho_c$  well over the entire range  $(-1, 1)$  of the latter, whereas  $\lambda$  sometimes only tracks  $\rho_s$  over the interval  $(0, 0.4)$ . Their numerical comparisons of moments also show the Cosine model to outperform the Sine model in terms of the ability to capture the correlation between  $\theta_1$  and  $\theta_2$ . On the other hand, the Sine model is easier to extend to more dimensions, and it is this model that will be extended in Chapter 2 to a multivariate setting. Mardia et al. (2007b) also give a brief comparison of the conditional densities of  $\theta_2$  given  $\theta_1$ , and this comparison will be further explored in Chapter 6. Having outlined the recurring topics of Part I and Part II of the thesis, we next give a more detailed breakdown of the thesis structure.

## 1.5 Outline of thesis

As previously mentioned, the thesis is divided into two parts: in Part I (comprising Chapters 2–4) we focus on a multivariate von Mises distribution with applications to

protein data, whilst Part II (Chapters 5–6) considers circular time series models with applications to protein data. A more detailed description of the chapters in Parts I and II is now given.

### 1.5.1 Outline of Part I

In Chapter 2 we introduce a multivariate von Mises distribution, which is a natural extension of the Sine model (Equation (1.12)). The model properties are discussed, including the normal approximation under high concentration and derivation of the conditional densities. Data simulation is also discussed for the bivariate and trivariate models, the former using a rejection algorithm and the latter a Gibbs sampling approach.

Chapter 3 introduces a method of estimation called pseudolikelihood estimation, which provides an alternative to maximum likelihood estimation when the normalizing constant of a density is unknown, as is the case for the multivariate von Mises distribution of dimension greater than two. A review of existing work on the pseudolikelihood is given, and a method of obtaining its efficiency is presented. Two different pseudolikelihoods are defined in terms of products of conditional densities. The efficiency of both is discussed, in this chapter, with reference to the multivariate normal distribution, which approximates the multivariate von Mises distribution under high concentration. The efficiency is shown to be unity for all bivariate and trivariate situations considered, except in the bivariate case when  $\sigma_1$  and  $\sigma_2$  are known.

In Chapter 4 we discuss parameter estimation and inference for the multivariate von Mises distribution. In particular, the efficiency of the pseudolikelihood is calculated numerically for the bivariate case, and simulations support the conclusions reached. For the trivariate case, efficiency is explored through simulation of data and subsequent parameter estimation. A trivariate von Mises distribution is used to model two separate protein data sets, and maximum likelihood estimation is compared with pseudolikelihood



estimation for these data. Inference is discussed in terms of likelihood ratio testing for both data sets, which testing includes the use of the pseudolikelihood and normal approximations.

### 1.5.2 Outline of Part II

Chapter 5 gives a review of circular time series models. Exploratory analysis is discussed in terms of the  $k$ -lag circular autocorrelation coefficient. A circular correlogram is defined and a method of assessing the significance of the coefficients is given. Four approaches to time series modeling of circular data are discussed, namely the von Mises AR process (defined in terms of conditional von Mises densities), processes based on link functions, wrapped AR processes and projected processes.

The use of conditional von Mises densities for the specification of AR(1) models is discussed in Chapters 6 and 7. In Chapter 6 the focus is on the use of the conditional densities of the Sine and Cosine models discussed in Section 1.4, whilst Chapter 7 adapts a circular regression model due to Downs and Mardia (2002) to a time series setting, and we will refer to the resulting model as the Möbius model. In each of Chapters 6 and 7, the discussion of models proceeds as follows. The deterministic and random components of the model are discussed, and the behaviour of simulated data considered in terms of the observations made. The equilibrium distribution of the processes is given explicitly for the Cosine and Sine time series models, whilst that for the Möbius model is considered numerically. The likelihood function and parameter estimation are considered, and the models are fitted to protein data. An analysis of goodness-of-fit follows, including the simulation of data based on maximum likelihood estimates and comparison of the resulting data with the original data. For each model, the potential for further work and ways in which the model can be extended are also discussed.

In Chapter 8 we summarize the conclusions of the thesis and assess its achievements.



## **Part I**

# **Multivariate von Mises models with application to protein conformational angles**



## Chapter 2

# A multivariate von Mises distribution

## 2.1 Introduction

In this chapter we present a multivariate circular distribution that is a natural extension of the univariate von Mises distribution and which, for highly concentrated data, follows an approximate multivariate normal distribution. The model is also a multivariate generalization of the Sine model outlined in Section 1.4. This chapter focuses on model properties and data simulation, whilst the ensuing two chapters (Chapters 3 and 4) consider parameter estimation and inference for the model (respectively), both in terms of simulated and actual protein data.

There are very few multivariate circular distributions in the literature. A wrapped multivariate normal distribution was discussed by Baba (1981). The univariate and bivariate marginals of this distribution are also wrapped normal, whereas the multivariate von Mises distribution to be discussed does not have von Mises marginals. However, maximum likelihood estimation is not computationally feasible for the wrapped model. A  $p$ -variate extension of the bivariate model of Equation (1.10) was given by Mardia and

Patrangenaru (2005) as

$$\exp \left\{ \sum a_s \cos \theta_s + \sum b_s \sin \theta_s + \sum a_{st} \cos \theta_s \cos \theta_t + \sum b_{st} \cos \theta_s \sin \theta_t + \sum c_{st} \sin \theta_s \sin \theta_t \right\} \quad (2.1)$$

where  $a_{ss} = b_{ss} = c_{ss} = 0$ ,  $b_{st} \neq b_{ts}$  and  $s, t = 1, \dots, p$ . In this chapter we present an extension of the Sine model (1.12) to  $p$  dimensions. The model is also a natural extension of the univariate von Mises distribution incorporating circular dependency between variables and approximates a multivariate normal distribution when the range of observations is small. The model is defined in Section 2.2, and its properties discussed in Section 2.3, in which section conditional, joint and marginal distributions are also derived. Methods for simulating bivariate and trivariate data from the model are presented in Section 2.4, and can be extended to higher dimensions.

## 2.2 The model

We define the probability density function of the multivariate von Mises distribution for  $\Theta = (\Theta_1, \Theta_2, \dots, \Theta_p)$  as follows:

$$f(\Theta) = C_p^{-1}(\boldsymbol{\kappa}, \boldsymbol{\Lambda}) \exp \left\{ \boldsymbol{\kappa}^T c(\boldsymbol{\theta}, \boldsymbol{\mu}) + \frac{1}{2} s(\boldsymbol{\theta}, \boldsymbol{\mu})^T \boldsymbol{\Lambda} s(\boldsymbol{\theta}, \boldsymbol{\mu}) \right\}, \quad (2.2)$$

where  $-\pi < \theta_j \leq \pi$ ,  $-\pi < \mu_j \leq \pi$ ,  $\kappa_j \geq 0$ ,  $-\infty < \lambda_{jl} < \infty$ ,

$$c(\boldsymbol{\theta}, \boldsymbol{\mu})^T = (\cos(\theta_1 - \mu_1), \cos(\theta_2 - \mu_2), \dots, \cos(\theta_p - \mu_p)),$$

$$s(\boldsymbol{\theta}, \boldsymbol{\mu})^T = (\sin(\theta_1 - \mu_1), \sin(\theta_2 - \mu_2), \dots, \sin(\theta_p - \mu_p)),$$

$$\boldsymbol{\mu}^T = (\mu_1, \mu_2, \dots, \mu_p), \quad \boldsymbol{\kappa}^T = (\kappa_1, \kappa_2, \dots, \kappa_p),$$

$$[\boldsymbol{\Lambda}]_{jl} = \lambda_{jl} = \lambda_{lj}, \quad \lambda_{jj} = 0,$$

and  $C_p^{-1}(\boldsymbol{\kappa}, \boldsymbol{\Lambda})$  is a normalizing constant. We denote the  $p$ -variate von Mises distribution by  $\Theta \sim M_p(\boldsymbol{\mu}, \boldsymbol{\kappa}, \boldsymbol{\Lambda})$ . For  $p = 1$ , (2.2) reduces to the univariate von Mises distribution,

and for  $p = 2$  to the bivariate distribution studied by Singh et al. (2002). The general  $p$ -variate model was proposed by Singh and Mardia (2004). The work herein expands on their preliminary work, and is summarised in Singh et al. (2007).

## 2.3 Model properties

In this section we derive certain properties of the multivariate von Mises distribution: it is shown that when the circular variables are highly concentrated they follow an approximate multivariate normal distribution; the univariate conditional distributions are shown to be von Mises distributed, and the shape of univariate and bivariate marginals for the trivariate model are explored. Without loss of generality, it will be assumed throughout this section that  $\boldsymbol{\mu} = \mathbf{0}$ .

When the fluctuations in the variables  $\Theta_1, \Theta_2, \dots, \Theta_p$  are small, we have

$$\cos \theta_j \approx 1 - \theta_j^2/2, \quad \sin \theta_j \approx \theta_j; \quad j = 1, 2, \dots, p.$$

If we substitute these expressions back into (2.2), then the part of the exponential involving  $\boldsymbol{\theta}$  is

$$\exp \left\{ -\frac{1}{2} \boldsymbol{\theta}^T \mathbf{K} \boldsymbol{\theta} + \frac{1}{2} \boldsymbol{\theta}^T \boldsymbol{\Lambda} \boldsymbol{\theta} \right\} \quad (2.3)$$

where  $\mathbf{K} = \text{diag}(\boldsymbol{\kappa})$ . (2.3) is proportional to the probability density function of the multivariate normal distribution with inverse covariance matrix  $\boldsymbol{\Sigma}^{-1} = \mathbf{K} - \boldsymbol{\Lambda}$ , ie.  $(\boldsymbol{\Sigma}^{-1})_{jj} = \kappa_j$ ,  $(\boldsymbol{\Sigma}^{-1})_{jl} = -\lambda_{jl}$ ,  $j \neq l$ . We therefore see that, under high concentration, the multivariate von Mises distribution follows an approximate multivariate normal distribution.

We now derive the univariate conditional distributions associated with (2.2).

**Theorem 2.3.1** *Let  $\boldsymbol{\Theta}$  have the  $p$ -variate von Mises distribution (2.2). Then the conditional distribution of  $\Theta_p$  given  $\theta_1, \dots, \theta_{p-1}$  is univariate von Mises with parameters*

given in (2.7).

### Proof

For  $p > 1$  write (2.2) as

$$f(\theta_1, \dots, \theta_p) = C_p^{-1} \exp \left\{ \sum_{j=1}^p \kappa_j \cos \theta_j + \sum_{j=1}^{p-1} \sum_{l=j+1}^p \lambda_{jl} \sin \theta_j \sin \theta_l \right\}. \quad (2.4)$$

Then the conditional density of  $\Theta_p$  given  $\theta_1, \dots, \theta_{p-1}$  is proportional to the expression obtained upon collecting the terms in the exponent of (2.4) involving  $\theta_p$ . Hence we have

$$f(\theta_p | \theta_1, \dots, \theta_{p-1}) \propto \exp \left\{ \kappa_p \cos \theta_p + \sum_{j=1}^{p-1} \lambda_{jp} \sin \theta_j \sin \theta_p \right\}. \quad (2.5)$$

Now define  $\kappa_{p \cdot 1, \dots, p-1}$  and  $\mu_{p \cdot 1, \dots, p-1}$  (to emphasize conditioning on  $\theta_1, \dots, \theta_{p-1}$ ) by

$$\begin{aligned} \kappa_p &= \kappa_{p \cdot 1, \dots, p-1} \cos \mu_{p \cdot 1, \dots, p-1} \\ \sum_{j=1}^{p-1} \lambda_{jp} \sin \theta_j &= \kappa_{p \cdot 1, \dots, p-1} \sin \mu_{p \cdot 1, \dots, p-1} \end{aligned} \quad (2.6)$$

and substitute these expressions into (2.5). Then we have

$$\begin{aligned} f(\theta_p | \theta_1, \dots, \theta_{p-1}) &\propto \exp \{ \kappa_{p \cdot 1, \dots, p-1} [\cos \theta_p \cos \mu_{p \cdot 1, \dots, p-1} + \sin \theta_p \sin \mu_{p \cdot 1, \dots, p-1}] \} \\ &= \exp \{ \kappa_{p \cdot 1, \dots, p-1} \cos (\theta_p - \mu_{p \cdot 1, \dots, p-1}) \}. \end{aligned} \quad (2.7)$$

(2.7) is proportional to the pdf of a univariate von Mises distribution with, from Equation (2.6), mean direction and concentration parameter given respectively by the equations

$$\begin{aligned} \tan \mu_{p \cdot 1, \dots, p-1} &= \sum_{j=1}^{p-1} \lambda_{jp} \sin \theta_j / \kappa_p \\ \kappa_{p \cdot 1, \dots, p-1} &= \left\{ \kappa_p^2 + \left( \sum_{j=1}^{p-1} \lambda_{jp} \sin \theta_j \right)^2 \right\}^{1/2} \end{aligned}$$

Thus

$$f(\theta_p | \theta_1, \dots, \theta_{p-1}) = [2\pi I_0(\kappa_{p \cdot 1, \dots, p-1})]^{-1} \exp \{ \kappa_{p \cdot 1, \dots, p-1} \cos (\theta_p - \mu_{p \cdot 1, \dots, p-1}) \}. \quad (2.8)$$

□



**Corollary 2.3.2** *The marginal pdf  $f_{\text{marg}} = f(\theta_1, \dots, \theta_{p-1})$  of  $(\Theta_1, \Theta_2, \dots, \Theta_{p-1})$  is given by*

$$f_{\text{marg}} = 2\pi C_p^{-1} I_0(\kappa_{p \cdot 1, \dots, p-1}) \exp \left\{ \sum_{j=1}^{p-1} \kappa_j \cos \theta_j + \sum_{j=0}^{p-2} \sum_{l=j+1}^{p-1} \lambda_{jl} \sin \theta_j \sin \theta_l \right\}. \quad (2.9)$$

**Proof**

Equation (2.9) is given by the ratio  $f(\theta_1, \dots, \theta_p)/f(\theta_p|\theta_1, \dots, \theta_{p-1})$ .  $\square$

More generally, we can obtain the conditional distribution of  $\Theta_1, \Theta_2, \dots, \Theta_r$  given  $\theta_{r+1}, \theta_{r+2}, \dots, \theta_p$  using the same approach as in the proof of Theorem 2.3.1. We obtain  $f_{\text{cond}} = f(\Theta_1, \dots, \Theta_r|\theta_{r+1}, \dots, \theta_p)$  as

$$f_{\text{cond}} \propto \exp \left\{ \sum_{j=1}^r \kappa_j \cos \theta_j + \sum_{j=1}^{r-1} \sum_{l=j+1}^r \lambda_{jl} \sin \theta_j \sin \theta_l + \sum_{j=1}^r c_j \sin \theta_j \right\},$$

where  $c_j$  is constant with respect to  $\theta_1, \dots, \theta_r$ , for  $j = 1, \dots, r$ . Writing  $\kappa_j = a_j \cos \nu_j$  and  $c_j = a_j \sin \nu_j$  gives that

$$f_{\text{cond}} \propto \exp \left\{ \sum_{j=1}^r a_j \cos(\theta_j - \nu_j) + \sum_{j=1}^{r-1} \sum_{l=j+1}^r \lambda_{jl} \sin \theta_j \sin \theta_l \right\}.$$

Next writing  $\phi_j = \theta_j - \nu_j$ , so that  $\theta_j = \phi_j + \nu_j$ , gives

$$f_{\text{cond}} \propto \exp \left\{ \sum_{j=1}^r a_j \cos \phi_j + \sum_{j=1}^{r-1} \sum_{l=j+1}^r \lambda_{jl} \sin(\phi_j + \nu_j) \sin(\phi_l + \nu_l) \right\}.$$

Expanding each sine term and their product gives, for constants  $a_{jl}$ ,  $b_{jl}$ ,  $c_{jl}$  and  $d_{jl}$ ,

$$f_{\text{cond}} \propto \exp \left\{ \sum_{j=1}^r a_j \cos \phi_j + \sum_{j=1}^{r-1} \sum_{l=j+1}^r [a_{jl} \cos \phi_j \cos \phi_l + b_{jl} \sin \phi_j \sin \phi_l + c_{jl} \cos \phi_j \sin \phi_l + d_{jl} \sin \phi_j \cos \phi_l] \right\}. \quad (2.10)$$

Equation (2.10) is the density of Mardia and Patrangenaru (2005) – given by Equation (2.1) – with  $b_s = 0$  for all  $s$ .

We now consider the shapes of the univariate and bivariate marginal distributions of the *trivariate* von Mises distribution, ie. (2.2) with  $p = 3$ . As discussed in Section 1.4, Singh et al. (2002) showed that for  $p = 2$  the univariate marginals are symmetric and unimodal or bimodal, and derived the conditions under which each occur, whilst Mardia et al. (2007b) did the same for the joint bivariate distribution.

When  $\Theta$  is distributed according to (2.2) it is clear that  $\Theta$  and  $-\Theta$  have the same distribution. This implies that all marginal distributions are symmetric. For the trivariate von Mises density, the bivariate marginal density of  $\Theta_1$  and  $\Theta_2$  is obtained by substituting  $p = 3$  into (2.9). It does not appear possible to obtain an analytic expression for the univariate marginal of  $\Theta_1$  from this bivariate density, since it involves the integral of a Bessel function applied to a function of  $\theta_1$  and  $\theta_2$ . In Figures 2.1 and 2.2 we plot the univariate and bivariate marginal densities for the trivariate von Mises distribution, the former obtained by integrating the latter numerically and the normalizing constant for the trivariate distribution obtained by numerically integrating (2.2) with  $p = 3$ . The plots correspond to the four parameter configurations  $(\kappa_1, \kappa_2, \kappa_3, \lambda_{12}, \lambda_{13}, \lambda_{23}) = (2, 3, 1, 2, 2, 2), (0.5, 0.75, 0.25, 2, 3, 4), (2, 2, 2, 20, 30, 40)$  and  $(2, 2, 2, 0.1, 0.1, 0.1)$ .

We see from the marginal plots that the univariate and bivariate densities are symmetric and either unimodal or bimodal. For many other parameter conformations we observe the same, and it appears that all univariate and bivariate marginals associated with the trivariate density have either a single or two modes and are symmetric about  $\mu = 0$ .

## 2.4 Simulating data

In this section we outline two possible methods for simulating data from the bivariate and trivariate von Mises distributions. Simulation from the bivariate model was discussed by Subramaniam (2005) and Mardia et al. (2007b). The first method to be considered is the acceptance–rejection method and the second a Gibbs sampler. For the former

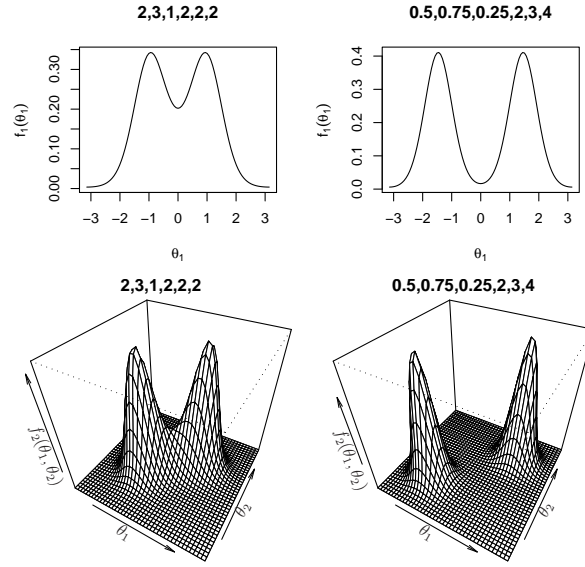


Figure 2.1: Univariate and bivariate marginals for trivariate von Mises distribution with  $(\kappa_1, \kappa_2, \kappa_3, \lambda_{12}, \lambda_{13}, \lambda_{23})$  given above each plot.

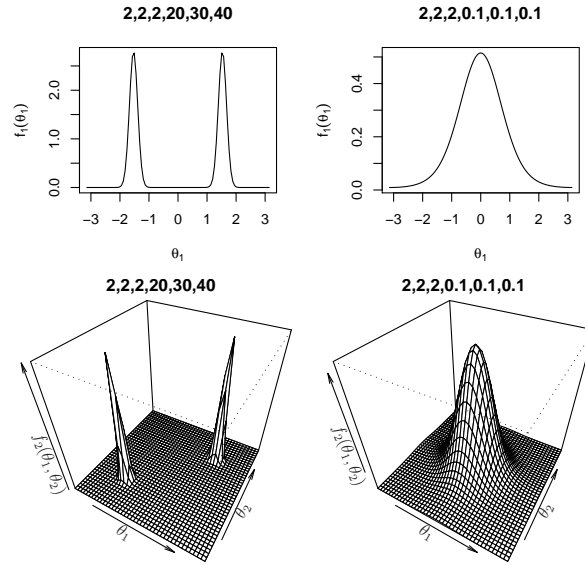


Figure 2.2: Univariate and bivariate marginals for trivariate von Mises distribution with  $(\kappa_1, \kappa_2, \kappa_3, \lambda_{12}, \lambda_{13}, \lambda_{23})$  given above each plot.

we describe the algorithm for the bivariate case. Denote by  $f_2(\theta_1, \theta_2; \kappa_1, \kappa_2, \lambda_{12})$  the bivariate von Mises density and by  $f_1(\theta_1; \kappa_1, \kappa_2, \lambda_{12})$  the marginal density of  $\Theta_1$  in this case, obtained by substituting  $p = 2$  into (2.4) and (2.9) respectively. Singh et al. (2002) compared the marginal of  $\Theta_1$  with a univariate von Mises distribution and showed that, for an appropriately chosen  $\kappa$  in the latter, the two have asymptotically the same normal density, and are therefore “similar”. As outlined in Section 1.4,  $f_1(\theta_1; \kappa_1, \kappa_2, \lambda_{12})$  is either unimodal or bimodal. An appropriate candidate density from which to sample is therefore a univariate von Mises distribution or a mixture of univariate von Mises distributions, depending on whether  $f_1(\theta_1; \kappa_1, \kappa_2, \lambda_{12})$  is unimodal or bimodal. We therefore define the candidate density by

$$g(\theta_1; \kappa, \nu) = \frac{1}{2} [2\pi I_0(\kappa)]^{-1} \{e^{\kappa \cos(\theta_1 + \nu)} + e^{\kappa \cos(\theta_1 - \nu)}\},$$

where  $\nu = 0$  if  $f_1(\theta_1; \kappa_1, \kappa_2, \lambda_{12})$  is unimodal and is given by  $\theta^*$  from Equation (1.19) if it is bimodal. Then  $f_1(\theta_1; \kappa_1, \kappa_2, \lambda_{12}) \leq Mg(\theta_1; \kappa, \nu) = h(\theta_1; \kappa, \nu)$  for some constant  $M > 1$  and for all  $\theta_1 \in [-\pi, \pi)$ . The algorithm for simulating a pair of variates  $(\theta_1, \theta_2)$  proceeds as follows:

1. For given  $\kappa_1, \kappa_2$  and  $\lambda_{12}$ , minimise the function

$$\max_{-\pi < \theta_1 \leq \pi} \left\{ \frac{f_1(\theta_1; \kappa_1, \kappa_2, \lambda_{12})}{g(\theta_1; \kappa, \nu)} \right\}$$

with respect to  $\kappa$ . Take  $M$  to be this minimum (thus making  $M$  as close to unity as possible subject to the constraint  $f_1(\theta_1; \kappa_1, \kappa_2, \lambda_{12})/g(\theta_1; \kappa, \nu) \leq M$ ), and denote by  $\kappa^*$  the value of  $\kappa$  giving this minimum.

2. Generate a random variate  $\theta_1^*$  from the candidate density  $g(\theta_1; \kappa^*, \nu)$  and a uniformly distributed random variate  $u \in [0, 1]$ .
3. If  $uMg(\theta_1^*; \kappa^*, \nu) \leq f_1(\theta_1^*; \kappa_1, \kappa_2, \lambda_{12})$  then accept  $\theta_1^*$  as a random variate from  $f_1(\theta_1; \kappa_1, \kappa_2, \lambda_{12})$ , else return to step 2.

4. Generate a random variate  $\theta_2^*$  using the conditional distribution of  $\Theta_2$  given  $\theta_1 = \theta_1^*$ , which is given by Equation (2.8) with  $p = 2$  and is univariate von Mises.

The minimisation in step 1 of the algorithm is performed using the `optim` function in R. In principle, steps 1, 2 and 3 of the above algorithm could be adapted to simulate random variates for the univariate marginals of the *trivariate* von Mises distribution. In practice, however, since we do not have an analytic expression for these densities, the process would be computationally very expensive.

Figure 2.3 compares the target density  $f_1(\theta_1; \kappa_1, \kappa_2, \lambda_{12})$  with  $M$  multiplied by the candidate density  $g(\theta_1, \kappa^*, \nu)$  for four parameter configurations. Also shown on the plots is the efficiency  $1/M$  of the simulation in each case. It can be seen from the plots, and is observed more generally, that efficiency appears to be improve as  $\kappa$  increases and as  $\lambda_{12}$  decreases (in absolute value – efficiency is the same for  $\lambda_{12}$  and  $-\lambda_{12}$ ).

An alternative approach to simulating variates from the bivariate and trivariate von Mises distributions is to use the Gibbs sampling approach. For the trivariate case, this appears to be the only feasible means of generating variates, due to the computational expense of the rejection sampling mentioned above. We first generate a vector of  $\theta_1$  values from a univariate von Mises distribution with concentration parameter  $\kappa_1$  and a vector of  $\theta_2$  values from a univariate von Mises distribution with concentration parameter  $\kappa_2$ . Using the conditional distribution of  $\Theta_3$  given  $(\theta_1, \theta_2)$  we generate a vector of  $\theta_3$  values conditional on the other two vectors. We then cycle through the  $\theta_1$ ,  $\theta_2$  and  $\theta_3$  vectors, replacing the values in the vector of interest at each stage with values generated conditionally on the values in the other two vectors. If we update each vector  $r$  times, then the point at which we have a sample from the trivariate von Mises distribution clearly depends on  $r$ . An appropriate choice of  $r$  is discussed further in Chapter 4.

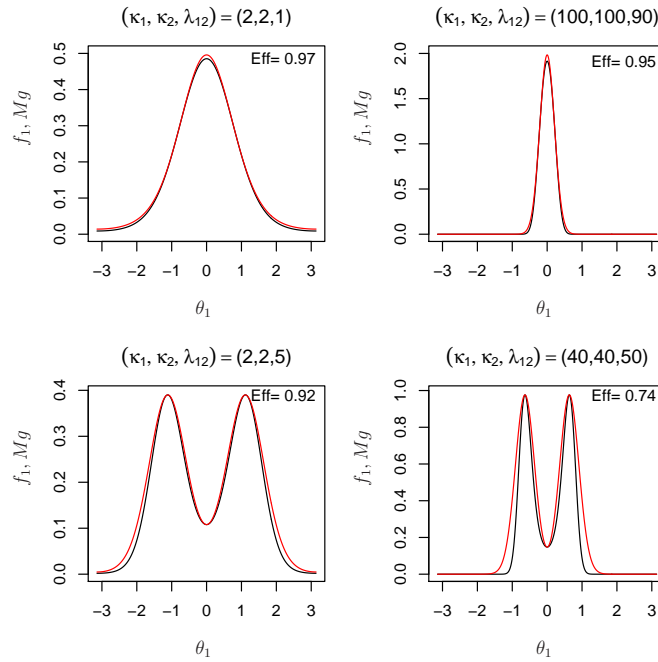


Figure 2.3: Target (black) and candidate (red) densities for four parameter configurations, and the efficiency of data simulation for the bivariate von Mises distribution.

As an alternative initialization of the  $\theta_1$  and  $\theta_2$  vectors, we could use the rejection simulation algorithm above to generate variates from the *bivariate* model, thereby incorporating a  $\lambda$  parameter.

## 2.5 Conclusions

We have presented in this chapter a multivariate circular distribution that is a natural extension of both the univariate von Mises distribution and the bivariate distribution known as the Sine model presented in Section 1.4. Many of the properties derived and observed are themselves natural extensions of the properties of the bivariate model derived by Singh et al. (2002) and Mardia et al. (2007b).

In discussion of the model properties it has been seen that the marginal distributions associated with the multivariate von Mises distribution are not themselves von Mises, but that all univariate conditional densities do follow von Mises distributions.

For the trivariate von Mises distribution it has been observed, for all parameter configurations considered, that all univariate and bivariate marginal densities are symmetric and either unimodal or bimodal, and it is conjectured that this is the case more generally.

Under high concentration (which requires unimodality of the joint density) it has been shown that the multivariate von Mises distribution can be approximated by a multivariate normal distribution, and the specific form of this normal distribution has been derived.

Two potential methods for data simulation have been presented. The first, an acceptance–rejection method, is suitable for the bivariate case (when the normalizing constant is known); the second, a Gibbs sampling approach, is more appropriate for higher dimensions, and will be used in subsequent chapters for simulation from the trivariate von Mises distribution.

Before considering parameter estimation for the multivariate von Mises distribution, we outline in the next chapter a general method of estimation useful when the normalizing constant is unknown, and discuss its efficiency for the multivariate normal distribution.





## Chapter 3

# Efficiency of the pseudolikelihood for multivariate normal distributions

### 3.1 Introduction

In certain circumstances, inference based on the likelihood function can be hindered by, for example, computational complexity or an unknown normalizing constant. In such cases it is necessary to seek an alternative method of estimation. Two pseudolikelihoods, each based on conditional distributions, are assessed in terms of their efficiency relative to the full likelihood for the multivariate normal distribution. By comparing information matrices, it is shown that both the pseudolikelihoods are fully efficient for the multivariate normal distribution with all variances equal to  $\sigma^2$  and all correlations equal to  $\rho$ . Loss of efficiency is shown for the estimator of equal  $\rho$  in the case of known  $\sigma^2$ . Both pseudolikelihoods are also shown to be fully efficient for the bivariate normal distribution with parameters  $\sigma_1$ ,  $\sigma_2$  and  $\rho$ . The derivation provides an outline of the procedure used to show that both pseudolikelihoods are also efficient for the trivariate normal distribution (with known mean vector). For the latter, information matrix elements are given.

An alternative approach to showing full efficiency is outlined for the five-parameter bivariate normal distribution, for which estimating equations for the full and pseudolikelihood approaches are compared.

The results for the normal distribution are of interest in their own right, but also give an insight into a limiting case of the multivariate von Mises distribution which, in the previous chapter, was shown to approximate a multivariate normal distribution when the variables are highly concentrated.

The two pseudolikelihoods to be studied are defined as follows. For a  $p$ -dimensional vector random variable  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_p)^T$  with joint probability density function  $f(\mathbf{y}; \mathbf{q})$ , where  $\mathbf{q}$  is an unknown parameter vector of length  $r$ , we define the full pseudolikelihood (FPL), based on a random sample of  $n$  observations of  $\mathbf{Y}$ , as

$$\text{FPL}(\mathbf{Y}; \mathbf{q}) = \prod_{j=1}^p \prod_{i=1}^n g_1(Y_{ji} | \text{rest}; \mathbf{q}) \quad (3.1)$$

where  $g_1(Y_j | \text{rest}; \mathbf{q})$  is the conditional distribution of  $Y_j$  given  $(Y_1, \dots, Y_{j-1}, Y_{j+1}, \dots, Y_p)$ . The partial pseudolikelihood (PPL) is defined as

$$\text{PPL}(\mathbf{Y}; \mathbf{q}) = \prod_{j \neq k}^p \prod_{i=1}^n g_2(Y_{ji} | Y_{ki}; \mathbf{q}). \quad (3.2)$$

Maximum pseudolikelihood estimates are then obtained by maximising (3.1) and (3.2) with respect to the parameters in  $\mathbf{q}$ .

Both pseudolikelihoods are consistent and asymptotically normal under the usual regularity conditions, as discussed by Arnold and Strauss (1991). Attention herein therefore focuses solely on efficiency. Cox and Reid (2004) discuss a class of pseudolikelihoods of which PPL is a special case. They discuss consistency of estimators for a single parameter and for parameter vectors for a  $p$ -dimensional random variable as  $n$  increases and for fixed  $n$  as  $p$  increases. FPL is the pseudolikelihood studied by Besag (1975, 1977). Arnold and Strauss (1991) discuss a general pseudolikelihood

of which both FPL and PPL are special cases. They show the pseudolikelihood estimator to be consistent and asymptotically normal under the standard regularity conditions, and consider its efficiency in special cases for exponential and beta conditional distributions, among others. They also briefly consider alternative specifications of the pseudolikelihood with trivariate or higher dimension. Arnold et al. (2001) give a comprehensive review of the role of conditionally specified distributions, including the pseudolikelihood. Wood (1993) uses the same approach as we use here in order to calculate the efficiency of a pseudolikelihood for the Bingham distribution, in which ML estimation is hindered by the normalizing constant. As we do for the bivariate von Mises distribution in the next chapter, he shows that efficiency tends to unity as parameters approach certain limits.

In Section 3.2 we give an outline of how to obtain the efficiency of the pseudolikelihood. Section 3.3 gives the derivation for the multivariate normal case with equal variances and equal correlations, whilst Section 3.4 describes the loss of efficiency in the estimator for  $\rho$  when the variances are known. Full efficiency is derived for the bivariate and trivariate cases in Section 3.5 with the only restriction on parameters being that the means are assumed to be known. Finally, the alternative approach of comparing estimating equations is outlined in Section 3.6, in which we show the pseudolikelihood to be fully efficient for the five parameter bivariate normal distribution.

## 3.2 Efficiency of the pseudolikelihood

The efficiency of one estimation method relative to another will be defined in terms of the ratio of determinants of information matrices for the two methods given the underlying density  $f(\mathbf{y}; \mathbf{q})$ . We therefore first calculate the information matrix  $\mathbf{I}$  for each method. Following Kent (1982), define the score vector by  $\mathbf{U}_f(\mathbf{y}; \mathbf{q}) = \partial \log f(\mathbf{y}; \mathbf{q}) / \partial \mathbf{q}$ . The

Fisher information matrix can be defined equivalently as

$$\begin{aligned}\mathbf{I}_f(\mathbf{q}) &= \int \mathbf{U}_f(\mathbf{y}; \mathbf{q}) \mathbf{U}_f(\mathbf{y}; \mathbf{q})^T f(\mathbf{y}; \mathbf{q}) d\mathbf{y} \quad \text{and} \\ \mathbf{I}_f(\mathbf{q}) &= - \int \partial \mathbf{U}_f(\mathbf{y}; \mathbf{q}) / \partial \mathbf{q}^T f(\mathbf{y}; \mathbf{q}) d\mathbf{y}.\end{aligned}$$

Now suppose we wish to calculate the information based on a density  $g(\mathbf{y}; \mathbf{q})$ , when the underlying density is  $f(\mathbf{y}; \mathbf{q})$ . To this end, define the squared score matrix by

$$\mathbf{J}_g(\mathbf{q}) = \int \mathbf{U}_g(\mathbf{y}; \mathbf{q}) \mathbf{U}_g(\mathbf{y}; \mathbf{q})^T f(\mathbf{y}; \mathbf{q}) d\mathbf{y},$$

and the expected score derivative matrix by

$$\mathbf{H}_g(\mathbf{q}) = - \int \partial \mathbf{U}_g(\mathbf{y}; \mathbf{q}) / \partial \mathbf{q}^T f(\mathbf{y}; \mathbf{q}) d\mathbf{y}.$$

The information matrix based on  $g$  when the underlying density is  $f$ , is then

$$\mathbf{I}_g(\mathbf{q}) = \mathbf{H}_g(\mathbf{q}) \mathbf{J}_g(\mathbf{q})^{-1} \mathbf{H}_g(\mathbf{q}), \quad (3.3)$$

Note that, if  $f = g$ , we have  $\mathbf{I}_g = \mathbf{H}_g = \mathbf{J}_g$ .

The efficiency of estimation based on  $g$  relative to estimation based on  $f$  can then be obtained as (see, for example, Davison (2003, p. 113))

$$\left\{ \frac{|\mathbf{I}_g(\mathbf{q})|}{|\mathbf{I}_f(\mathbf{q})|} \right\}^{1/r} \quad (3.4)$$

where  $r$  is the dimension of  $\mathbf{q}$ .

### 3.3 Multivariate normal with unknown $\sigma^2$ and $\rho$

We consider first the case in which  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_p)^T$  follows a multivariate normal distribution with  $\text{var}(Y_j) = \sigma^2$ ,  $\text{corr}(Y_j, Y_k) = \rho\sigma^2$  and  $E(Y_j) = 0$ , for all  $1 \leq j \neq k \leq p$ .

In this case the vector  $\mathbf{q}$  of unknown parameters is  $\mathbf{q} = (\sigma, \rho)$  and  $r = 2$ .

### 3.3.1 Full pseudolikelihood

For FPL (Equation (3.1)) we have

$$Y_j | \text{rest} \sim N_p \left( \frac{\rho}{1 + (p-2)\rho} \left[ \sum_{r=1}^p y_r - y_j \right], \sigma^2 \left[ 1 - \frac{(p-1)\rho^2}{1 + (p-2)\rho} \right] \right), \quad (3.5)$$

$j = 1, 2, \dots, p$ . Denote by  $g_{\text{FPL}}$  the contribution to FPL (ignoring constant terms) from a single observation  $\mathbf{y}$  of  $\mathbf{Y}$ . Based on (3.5) we have

$$\log g_{\text{FPL}} = -\frac{p}{2} \log \left\{ \sigma^2 \left[ 1 - \frac{(p-1)\rho^2}{1 + (p-2)\rho} \right] \right\} - \frac{\sum_{j=1}^p \left[ y_j - \frac{\rho(\sum_{r=1}^p y_r - y_j)}{1 + (p-2)\rho} \right]^2}{2\sigma^2 \left[ 1 - \frac{(p-1)\rho^2}{1 + (p-2)\rho} \right]}. \quad (3.6)$$

Writing  $\sum_{r=1}^p y_r = p\bar{y}$ , multiplying the numerator and denominator of the second term in (3.6) by  $a^2$ , where  $a = [1 + (p-2)\rho]$ , and expanding the square, gives the numerator of this second term as

$$b^2 \sum_{j=1}^p y_j^2 + p(c^2 - 2bc)\bar{y}^2, \quad (3.7)$$

where  $b = 1 + (p-1)\rho$  and  $c = p\rho$ . Now write  $\sum_j y_j^2 = \sum_j (y_j - \bar{y})^2 + p\bar{y}^2$  and define  $SS_W = \sum_j (y_j - \bar{y})^2$  and  $SS_B = \bar{y}^2$ . Then (3.7) becomes

$$b^2 SS_W + p(1 - \rho)^2 SS_B,$$

where the coefficient of  $SS_B$  is  $p(b^2 + c^2 - 2bc) = p(b - c)^2 = p(1 - \rho)^2$ .  $\log g_{\text{FPL}}$  is therefore

$$-\frac{p}{2} \log \left\{ \sigma^2 \left[ 1 - \frac{(p-1)\rho^2}{a} \right] \right\} - \frac{b^2 SS_W + p(1 - \rho)^2 SS_B}{2\sigma^2 \{a^2 - (p-1)\rho^2 a\}}. \quad (3.8)$$

Due to the complicated nature of some of the expressions, most of the derivatives in this chapter are checked using Maple. Differentiating (3.8) with respect to  $\sigma$  gives

$$\frac{\partial \log g_{\text{FPL}}}{\partial \sigma} = -\frac{p}{\sigma} + \frac{b^2 SS_W + p(1 - \rho)^2 SS_B}{\sigma^3 \{a^2 - (p-1)\rho^2 a\}}, \quad (3.9)$$

whilst the first derivative with respect to  $\rho$  can be calculated as

$$\begin{aligned} \frac{\partial \log g_{\text{FPL}}}{\partial \rho} &= \frac{p\sigma^2 [2\rho(p-1)a - \rho^2(p-1)(p-2)] - 2[b(p-1)SS_W - p(1-\rho)SS_B]}{2\sigma^2 [a^2 - (p-1)\rho^2 a]} \\ &+ \frac{[b^2 SS_W + p(1-\rho)^2 SS_B] [2a(p-2) - (p-1)\rho(3a-1)]}{2\sigma^2 [a^2 - (p-1)\rho^2 a]^2}. \end{aligned} \quad (3.10)$$

The second derivatives of (3.8) can be found as follows:

$$\frac{\partial^2 \log g_{\text{FPL}}}{\partial \sigma^2} = \frac{p}{\sigma^2} - 3 \frac{b^2 SS_W + p(1-\rho)^2 SS_B}{\sigma^4 \{a^2 - (p-1)\rho^2 a\}}; \quad (3.11)$$

$$\frac{\partial^2 \log g_{\text{FPL}}}{\partial \rho^2} = ASS_W + BSS_B + C \quad (3.12)$$

where

$$\begin{aligned} A &= \frac{(b+2)(p-2)^2 \rho^2 + 6(p-2)\rho + 4-p}{\sigma^2 [(p-1)\rho^2 - (p-3)\rho - 1]^3}, \\ B &= \frac{p[(p-1)^2(p-2)^2(\rho^3 - 3\rho^2) - 6(p-1)^2(p-2)\rho - (p-1)(3p-4)]}{\sigma^2 [(p-1)(p-2)\rho^2 + (2p-3)\rho + 1]^3}, \\ C &= \frac{p(p-1)[(p-1)(p-2)^2 \rho^4 + 4(p-1)(p-2)\rho^3 + 2(p-1)\rho^2 + 2a]}{2[(p-1)\rho^2 - (p-2)\rho - 1]^2 a^2}. \end{aligned}$$

Finally,

$$\frac{\partial^2 \log g_{\text{FPL}}}{\partial \rho \partial \sigma} = \frac{(p-2)(b+1)\rho + 2}{\sigma^3(1-\rho)^2 a^2} SS_W + \frac{p(p-1)[(p-2)(\rho^2 - 2\rho) - 2]}{\sigma^3 a^2 b^2} SS_B. \quad (3.13)$$

In order to obtain the  $\mathbf{H}_{g_{\text{FPL}}}$  and  $\mathbf{J}_{g_{\text{FPL}}}$  matrices for FPL we require expressions for the expectations of  $SS_B$ ,  $SS_W$  and their product and squares. We first show that  $SS_B$  and  $SS_W$  are independent, enabling us to calculate the expected value of their product as the product of their expected values.

**Theorem 3.3.1** *Let  $\mathbf{Y}$  have a  $p$ -variate normal distribution with all correlations equal to  $\rho$ , all variances equal to  $\sigma^2$  and all means equal to zero. Then the expressions  $SS_W = \sum_j (y_j - \bar{y})^2$  and  $SS_B = \bar{y}^2$ , where  $\bar{y} = \sum_{j=1}^p y_j/p$ , are independent.*

**Proof**

$SS_B$  and  $SS_W$  are independent if  $(y_j - \bar{y})$  and  $\bar{y}$  are independent for all  $j = 1, 2, \dots, p$ . Now  $(y_j - \bar{y})$  and  $\bar{y}$  are independent if  $\text{Cov}\{(y_j - \bar{y}), \bar{y}\} = 0$ , and since  $E(\bar{y})E(y_j - \bar{y}) = 0$  it suffices to show that  $E\{(y_j - \bar{y})\bar{y}\} = 0$ , ie. that  $E(y_j \bar{y}) = E(\bar{y}^2)$ . Now

$$E(y_j \bar{y}) = \frac{E(y_j^2) + (p-1)E(y_j y_k)}{p} = \frac{\sigma^2 [1 + (p-1)\rho]}{p},$$

$k \neq j$ , whilst

$$E(\bar{y}^2) = \frac{pE(y_j^2) + p(p-1)E(y_j y_k)}{p^2} = \frac{\sigma^2 [1 + (p-1)\rho]}{p}.$$

Thus  $E(y_j \bar{y}) = E(\bar{y}^2)$  and  $SS_W$  and  $SS_B$  are independent.  $\square$

**Remark:** The independence of  $SS_W$  and  $SS_B$  requires that all correlations are equal, all variances are equal and all means are equal (but not necessarily zero).

By collecting terms of the form  $y_i y_j$  for  $SS_B$  and  $SS_W$ , and of the form  $y_i y_j y_k y_l$  for their squares, where any, all or none of  $i, j, k, l$  could be equal, and by using Theorem 3.3.1, we find that

$$\begin{aligned} E(SS_B) &= \frac{b\sigma^2}{p}; & E(SS_W) &= (p-b)\sigma^2; & E(SS_B SS_W) &= \frac{b(p-b)\sigma^4}{p}; \\ E(SS_B^2) &= \frac{3b^2\sigma^4}{p^2}; & E(SS_W^2) &= (p^2-1)(1-\rho)^2\sigma^4. \end{aligned} \quad (3.14)$$

By substituting the expressions from (3.14) into the second derivatives (3.11)–(3.13) and the product and squares of the first derivatives (3.9) and (3.10) we obtain the symmetric elements of the matrices  $\mathbf{H}_{g\text{FPL}}$  and  $\mathbf{J}_{g\text{FPL}}$ :

$$\begin{aligned} [\mathbf{H}_{g\text{FPL}}]_{11} &= \frac{2p}{\sigma^2} \\ [\mathbf{H}_{g\text{FPL}}]_{12} &= -\frac{p(p-1)(a+1)\rho}{ab(1-\rho)\sigma} \\ [\mathbf{H}_{g\text{FPL}}]_{22} &= \frac{p(p-1)\{(p-1)(p-2)(a+3)\rho^3 + 2[a + (p-1)\rho^2]\}}{2a^2b^2(1-\rho)^2} \\ [\mathbf{J}_{g\text{FPL}}]_{11} &= \frac{2p\{[1 + (p-1)(p-2)]\rho^2 + 2(p-2)\rho + 1\}}{a^2\sigma^2} \\ [\mathbf{J}_{g\text{FPL}}]_{12} &= -\frac{p(p-1)\rho\{(p-2)\{[1 + (p-1)(p-2)]\rho^3 + (4a+3)\rho\} + 4\}}{a^3b(1-\rho)\sigma} \\ [\mathbf{J}_{g\text{FPL}}]_{22} &= \frac{p(p-1)\{(p-1)(p-2)^2\{[1 + (p-1)(p-2)]\rho^6 + (6a+11)\rho^4 + D\}\}}{2a^4b^2(1-\rho)^2\sigma^2} \end{aligned} \quad (3.15)$$

where

$$D = 4(p-1)(p-2)(1+4\rho)\rho^2 + 4(2a + \rho^2 - 1).$$

We can therefore calculate the information matrix  $\mathbf{I}_{g_{\text{FPL}}} = \mathbf{H}_{g_{\text{FPL}}} \mathbf{J}_{g_{\text{FPL}}}^{-1} \mathbf{H}_{g_{\text{FPL}}}$  as defined in Section 3.2. We obtain

$$\mathbf{I}_{g_{\text{FPL}}} = \begin{bmatrix} \frac{2p}{\sigma^2} & -\frac{p(p-1)\rho}{b\sigma(1-\rho)} \\ -\frac{p(p-1)\rho}{b\sigma(1-\rho)} & \frac{p(p-1)[1+(p-1)\rho^2]}{2b^2(1-\rho)^2} \end{bmatrix}. \quad (3.16)$$

The matrix in (3.16) can be compared with the Fisher Information matrix based on the full likelihood, the  $(i, j)$ th element of which can be written as (see, for example, Porat and Friedlander (1986))

$$[\mathbf{I}_f]_{ij} = \left[ \frac{\partial \boldsymbol{\mu}}{\partial \mathbf{q}_i} \right]^T \boldsymbol{\Sigma}^{-1} \left[ \frac{\partial \boldsymbol{\mu}}{\partial \mathbf{q}_j} \right] + \frac{1}{2} \text{tr} \left\{ \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \mathbf{q}_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \mathbf{q}_j} \right\}; \quad (3.17)$$

$i, j = 1, 2, \dots, p$ , where  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  are the mean vector and covariance matrix respectively. In fact (3.16) and (3.17) are equal when  $\boldsymbol{\mu} = \mathbf{0}$ , all diagonal elements of  $\boldsymbol{\Sigma}$  equal  $\sigma^2$  and all other elements are equal to  $\rho\sigma^2$ , showing that the full pseudolikelihood is fully efficient for the multivariate normal distribution with unknown parameters  $\rho$  and  $\sigma$ . We next consider the partial pseudolikelihood in the same parameter setting.

### 3.3.2 Partial pseudolikelihood

For the partial pseudolikelihood (PPL) (Equation (3.2)) we have

$$Y_j | Y_k \sim N(\rho y_k, \sigma^2[1 - \rho^2]), \quad 1 \leq j \neq k \leq p.$$

The contribution  $\log g_{\text{PPL}}$  to  $\log \text{PPL}$  (ignoring constant terms) from a single observation of  $\mathbf{Y}$  is therefore

$$\log g_{\text{PPL}} = -p(p-1) \log \sigma - \frac{p(p-1)}{2} \log(1 - \rho^2) - \frac{\sum_{j=1}^p \sum_{j \neq k=1}^p (y_j - \rho y_k)^2}{2\sigma^2(1 - \rho^2)}. \quad (3.18)$$



By writing the numerator of the final term in (3.18) as

$$\begin{aligned} & (p-1) \sum_{j=1}^p y_j^2 + \rho^2(p-1) \sum_{j=1}^p y_j^2 - 2\rho \left[ \left( \sum_{j=1}^p y_j \right)^2 - \sum_{j=1}^p y_j^2 \right] \\ &= [(p-1)(1+\rho^2) + 2\rho] \sum_{j=1}^p y_j^2 - 2\rho \left( \sum_{j=1}^p y_j \right)^2 \end{aligned}$$

we can write  $\log g_{\text{PPL}}$  as

$$-p(p-1) \log \sigma - \frac{p(p-1)}{2} \log(1-\rho^2) - \frac{[(p-1)(1+\rho^2) + 2\rho] SS_1 - 2\rho SS_2}{2\sigma^2(1-\rho^2)} \quad (3.19)$$

where  $SS_1 = \sum_{j=1}^p y_j^2$  and  $SS_2 = (\sum_{j=1}^p y_j)^2$ . Differentiating (3.19) with respect to  $\sigma$  gives

$$\frac{\partial \log g_{\text{PPL}}}{\partial \sigma} = \frac{[(p-1)(1+\rho^2) + 2\rho] SS_1 - 2\rho SS_2}{\sigma^3(1-\rho^2)} - \frac{p(p-1)}{\sigma}$$

whilst the first derivative with respect to  $\rho$  is

$$\frac{\partial \log g_{\text{PPL}}}{\partial \rho} = \frac{(1+\rho^2) SS_2 - [2b - (1-\rho^2)] SS_1 + p(p-1)\rho(1-\rho^2)\sigma^2}{\sigma^2(1-\rho^2)^2}.$$

The second derivatives can be calculated as

$$\begin{aligned} \frac{\partial^2 \log g_{\text{PPL}}}{\partial \sigma^2} &= \frac{p(p-1)}{\sigma^2} - \frac{3\{[(p-1)(1+\rho^2) + 2\rho] SS_1 - 2\rho SS_2\}}{\sigma^4(1-\rho^2)}, \\ \frac{\partial^2 \log g_{\text{PPL}}}{\partial \rho^2} &= \frac{2\rho(\rho^2+3) SS_2 - \{6b\rho + 2(p-1) + 2\rho^3\} SS_1 + p(p-1)(1-\rho^4)\sigma^2}{\sigma^2(1-\rho^2)^3} \end{aligned}$$

and

$$\frac{\partial^2 \log g_{\text{PPL}}}{\partial \rho \partial \sigma} = \frac{2\{[2(p-1)\rho + (1+\rho^2)] SS_1 - (1+\rho^2) SS_2\}}{\sigma^3(1-\rho^2)^2}.$$

Using the approach described in Section 3.3.1 we obtain the expected values of  $SS_1$ ,  $SS_2$  and their product and squares as

$$E(SS_1) = p\sigma^2; \quad E(SS_2) = bp\sigma^2; \quad E(SS_1 SS_2) = bp(2b+p)\sigma^4;$$

$$E(SS_1^2) = p\sigma^4 [2\rho^2(p-1) + 2 + p]; \quad E(SS_2^2) = 3b^2 p^2 \sigma^4$$

where  $b = 1 + (p - 1)\rho$  as in Section 3.3.1. The elements of the symmetric matrices  $\mathbf{H}_{g_{\text{FPL}}}$  and  $\mathbf{J}_{g_{\text{FPL}}}$  can then be calculated as:

$$\begin{aligned} [\mathbf{H}_{g_{\text{PPL}}}]_{11} &= \frac{2p(p-1)}{\sigma^2} \\ [\mathbf{H}_{g_{\text{PPL}}}]_{12} &= -\frac{2p(p-1)\rho}{\sigma(1-\rho^2)} \\ [\mathbf{H}_{g_{\text{PPL}}}]_{22} &= \frac{p(p-1)(1+\rho^2)}{(1-\rho^2)^2} \\ [\mathbf{J}_{g_{\text{PPL}}}]_{11} &= \frac{2p(p-1)\{(p-1)[(p-1)\rho^4 - 2(p-3)\rho^3 + 2\rho + 1] + (p^2 - 5p + 8)\rho^2\}}{\sigma^2(1+\rho)^2} \\ [\mathbf{J}_{g_{\text{PPL}}}]_{12} &= \frac{2p(p-1)\rho\{(p-1)(p-2)\rho^3 - 2[p^2 - 3(p-1)]\rho^2 + (p-1)(p-6)\rho - 2\}}{\sigma(1+\rho)^3(1-\rho)} \\ [\mathbf{J}_{g_{\text{PPL}}}]_{22} &= \frac{2p(p-1)\{[p^2 - 3(p-1)]\rho^4 - 2(p-1)(p-3)\rho^3 + E\}}{(1+\rho)^4(1-\rho)^2}, \end{aligned}$$

where  $E = (p^2 - 3p + 4)\rho^2 + 2b - 1$ .

Calculating the information matrix  $\mathbf{I}_{g_{\text{PPL}}} = \mathbf{H}_{g_{\text{PPL}}} \mathbf{J}_{g_{\text{PPL}}}^{-1} \mathbf{H}_{g_{\text{PPL}}}$  yields the same matrix as in Equation (3.16) for FPL. We therefore conclude that PPL is also fully efficient for the multivariate normal distribution being studied.

### 3.4 Loss of efficiency with known unit variances

We now consider the case in which  $\mathbf{Y}$  follows a multivariate normal distribution with (known) unit variances and unknown equal correlations  $\rho$ . Cox and Reid (2004) calculate the efficiency of the partial pseudolikelihood in this context, and their results will be compared with the efficiency of the full pseudolikelihood. We proceed as in the case of unknown  $\sigma$ . The terms of interest  $\log g_{\text{FPL}}$  and its derivatives with respect to  $\rho$ , the expected values of the sums of squares and the values  $H_{g_{\text{FPL}}}$  and  $J_{g_{\text{FPL}}}$  can all be obtained by substituting  $\sigma^2 = 1$  into the relevant expressions in Section 3.3.1. In particular, the information  $I_{g_{\text{FPL}}} = H_{g_{\text{FPL}}}^2 / J_{g_{\text{FPL}}}$  from a single observation in the current setting

is obtained by substituting  $\sigma^2 = 1$  into  $[\mathbf{H}_{g_{\text{FPL}}}]_{22}$  from Equation (3.15) to obtain  $H_{g_{\text{FPL}}}$ , and similarly for  $J_{g_{\text{FPL}}}$ . The information  $I_f$  based on the full likelihood is simply the second diagonal element of the matrix in Equation (3.16), and the efficiency of FPL is the ratio of  $I_{g_{\text{FPL}}}$  to  $I_f$ .

Figure 3.1 shows the efficiency of both FPL and PPL (the latter as obtained by Cox and Reid (2004)) for  $\rho \in [0, 1]$ . The efficiency of both is unity for  $p = 2$  and for any  $p$  if  $\rho$  is 0 or 1. Figure 3.1 displays the efficiency for  $p = 3, 5, 8$  and 10, increasing  $p$  resulting in a loss of efficiency. It can be seen that in general the efficiency of FPL is greater than that of PPL, although for  $p = 3$  they are almost identical.

In the current context, the covariance matrix is positive definite for  $\rho \in (-1/(p-1), 1)$ . Figure 3.2 shows the efficiency of each pseudolikelihood for  $\rho \in (-1/(p-1), 0)$  and for the same values of  $p$  as before. In this interval it is clear that FPL outperforms PPL. As can be seen from the plots, the efficiency when  $\rho = -1/(p-1)$  is unity for FPL and zero for PPL. The exception is when  $p = 2$ , in which case the two pseudolikelihoods are equal with full efficiency for all  $\rho \in [-1, 1]$ .

### 3.5 Efficiency for the case of unequal variances and correlations

In this section we consider the efficiency for the bivariate and trivariate normal distributions when the variances and correlations are not restricted to being equal. By deriving the information matrices for the pseudolikelihood and full likelihood, we show the former to be fully efficient for the bivariate case. The same approach can be adopted in order to show that both pseudolikelihoods are fully efficient for the trivariate normal distribution. For this case, we give information matrix elements and highlight symmetries that could potentially be exploited in order to show the pseudolikelihoods to be fully

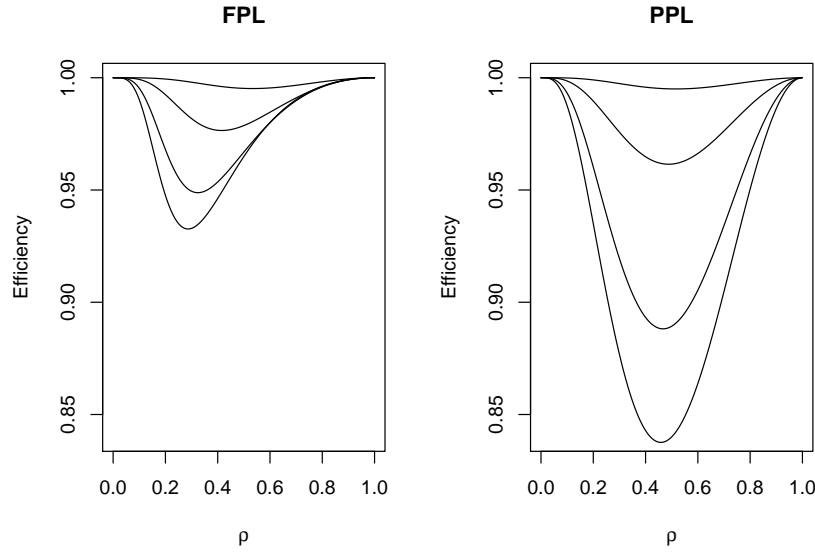


Figure 3.1: Efficiency of FPL and PPL for multivariate normal distribution (with known unit variances) as a function of  $\rho \in [0, 1]$  and for  $p = 3, 5, 8, 10$  (top to bottom).

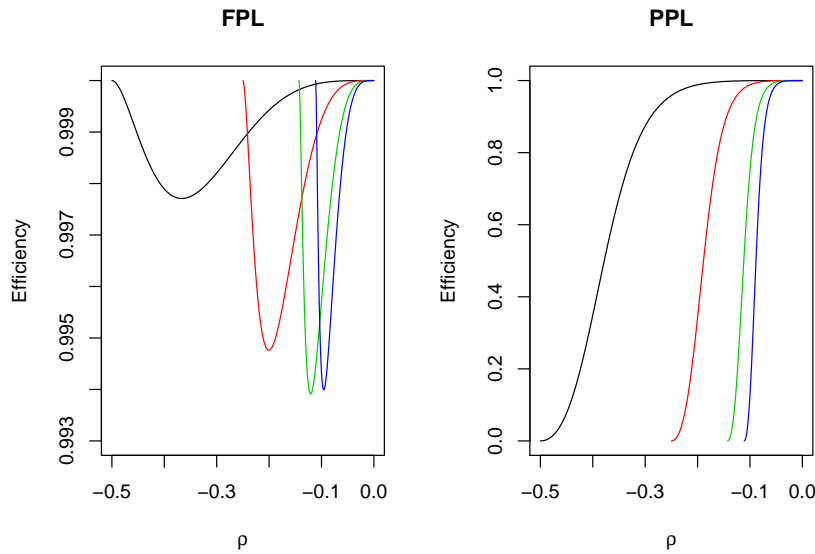


Figure 3.2: Efficiency of FPL and PPL for multivariate normal distribution (with known unit variances) as a function of  $\rho \in (-1/(p-1), 0)$  and for  $p = 3, 5, 8, 10$  (left to right).

efficient for the general  $p$ -variate normal distribution.

### 3.5.1 Efficiency for the bivariate normal distribution

We follow the approach outlined in Section 3.2 and applied in Sections 3.3 and 3.4 in order to obtain the efficiency of the pseudolikelihood for the bivariate normal distribution with parameters  $\sigma_1$ ,  $\sigma_2$  and  $\rho$ .

The log density of the bivariate normal distribution with zero mean vector is given (ignoring constant terms) by

$$\log f(y_1, y_2) = -\log \sigma_1 - \log \sigma_2 - \frac{1}{2} \log(1 - \rho^2) - \frac{y_1^2 \sigma_2^2 + y_2^2 \sigma_1^2 - 2\rho y_1 y_2 \sigma_1 \sigma_2}{2\sigma_1^2 \sigma_2^2 (1 - \rho^2)}.$$

For direct comparison the log product of conditional densities (omitting constant terms) used for the pseudolikelihood is

$$\log g(y_1, y_2) = -\log \sigma_1 - \log \sigma_2 - \log(1 - \rho^2) - \frac{(y_1^2 \sigma_2^2 + y_2^2 \sigma_1^2)(1 + \rho^2) - 4\rho y_1 y_2 \sigma_1 \sigma_2}{2\sigma_1^2 \sigma_2^2 (1 - \rho^2)}.$$

The information matrix based on maximum likelihood estimation is calculated from Equation (3.17) as

$$\mathbf{I}_{ML} = \begin{bmatrix} \frac{2-\rho^2}{\sigma_1^2(1-\rho^2)} & \frac{-\rho^2}{\sigma_1\sigma_2(1-\rho^2)} & \frac{-\rho}{\sigma_1(1-\rho^2)} \\ \frac{-\rho^2}{\sigma_1\sigma_2(1-\rho^2)} & \frac{2-\rho^2}{\sigma_2^2(1-\rho^2)} & \frac{-\rho}{\sigma_2(1-\rho^2)} \\ \frac{-\rho}{\sigma_1(1-\rho^2)} & \frac{-\rho}{\sigma_2(1-\rho^2)} & \frac{1+\rho^2}{(1-\rho^2)^2} \end{bmatrix}. \quad (3.20)$$

We now derive the information matrix based on the pseudolikelihood approach. The derivatives required can be calculated and simplified to give the following expressions, for  $i, j = 1, 2; i \neq j$ :

$$\frac{\partial \log g}{\partial \sigma_i} = \frac{y_i^2 \sigma_j (1 + \rho^2) - \sigma_i^2 \sigma_j (1 - \rho^2) - 2\rho y_i y_j \sigma_i}{\sigma_i^3 \sigma_j (1 - \rho^2)}, \quad (3.21)$$

$$\frac{\partial \log g}{\partial \rho} = \frac{2[y_1 y_2 \sigma_1 \sigma_2 (1 + \rho^2) + \sigma_1^2 \sigma_2^2 \rho (1 - \rho^2) - \rho (y_1^2 \sigma_2^2 + y_2^2 \sigma_1^2)]}{\sigma_1^2 \sigma_2^2 (1 - \rho^2)^2}, \quad (3.22)$$

$$\frac{\partial^2 \log g}{\partial \sigma_i^2} = \frac{\sigma_i^2 \sigma_j (1 - \rho^2) - 3y_i^2 \sigma_j (1 + \rho^2) + 4\rho y_i y_j \sigma_i}{\sigma_i^4 \sigma_j (1 - \rho^2)}, \quad (3.23)$$

$$\frac{\partial^2 \log g}{\partial \rho^2} = \frac{2[2\rho y_1 y_2 \sigma_1 \sigma_2 (\rho^2 + 3) + \sigma_1^2 \sigma_2^2 (1 - \rho^4) - (y_1^2 \sigma_2^2 + y_2^2 \sigma_1^2)(1 + 3\rho^2)]}{\sigma_1^2 \sigma_2^2 (1 - \rho^2)^3}, \quad (3.24)$$

$$\frac{\partial^2 \log g}{\partial \sigma_i \partial \sigma_j} = \frac{2\rho y_i y_j}{\sigma_i^2 \sigma_j^2 (1 - \rho^2)}, \quad (3.25)$$

$$\frac{\partial^2 \log g}{\partial \sigma_i \partial \rho} = \frac{2y_i [2\rho y_i \sigma_j - y_j \sigma_i (1 + \rho^2)]}{\sigma_i^3 \sigma_j (1 - \rho^2)^2}. \quad (3.26)$$

Taking expectations of the second derivatives, and expectations of the products of the first derivatives in Equations (3.21)–(3.26), enables us to calculate elements of the  $\mathbf{H}_{PL}$  and  $\mathbf{J}_{PL}$  matrices needed to obtain the information matrix for the pseudolikelihood. Omitting much algebra, we obtain

$$\mathbf{H}_{PL} = \begin{bmatrix} \frac{2}{\sigma_1^2(1-\rho^2)} & \frac{-2\rho^2}{\sigma_1\sigma_2(1-\rho^2)} & \frac{-2\rho}{\sigma_1(1-\rho^2)} \\ \frac{-2\rho^2}{\sigma_1\sigma_2(1-\rho^2)} & \frac{2}{\sigma_2^2(1-\rho^2)} & \frac{-2\rho}{\sigma_2(1-\rho^2)} \\ \frac{-2\rho}{\sigma_1(1-\rho^2)} & \frac{-2\rho}{\sigma_2(1-\rho^2)} & \frac{2(1+\rho^2)}{(1-\rho^2)^2} \end{bmatrix} \quad (3.27)$$

and

$$\mathbf{J}_{PL} = \begin{bmatrix} \frac{2(1+\rho^2)}{\sigma_1^2(1-\rho^2)} & \frac{-2\rho^2(1+\rho^2)}{\sigma_1\sigma_2(1-\rho^2)} & \frac{-4\rho}{\sigma_1(1-\rho^2)} \\ \frac{-2\rho^2(1+\rho^2)}{\sigma_1\sigma_2(1-\rho^2)} & \frac{2(1+\rho^2)}{\sigma_2^2(1-\rho^2)} & \frac{-4\rho}{\sigma_2(1-\rho^2)} \\ \frac{-4\rho}{\sigma_1(1-\rho^2)} & \frac{-4\rho}{\sigma_2(1-\rho^2)} & \frac{4(1+\rho^2)}{(1-\rho^2)^2} \end{bmatrix}. \quad (3.28)$$

We obtain the inverse of  $\mathbf{J}_{PL}$  as

$$\mathbf{J}_{PL}^{-1} = \begin{bmatrix} \frac{\sigma_1^2(1+3\rho^4)}{2(1+\rho^2)^2(1-\rho^2)^2} & \frac{\rho^2\sigma_1\sigma_2(3+\rho^4)}{2(1+\rho^2)^2(1-\rho^2)^2} & \frac{\rho\sigma_1}{2(1-\rho^2)} \\ \frac{\rho^2\sigma_1\sigma_2(3+\rho^4)}{2(1+\rho^2)^2(1-\rho^2)^2} & \frac{\sigma_2^2(1+3\rho^4)}{2(1+\rho^2)^2(1-\rho^2)^2} & \frac{\rho\sigma_2}{2(1-\rho^2)} \\ \frac{\rho\sigma_1}{2(1+\rho^2)} & \frac{\rho\sigma_2}{2(1-\rho^2)} & \frac{1+\rho^2}{4} \end{bmatrix}. \quad (3.29)$$

Calculating the information matrix  $I_{PL}$  based on the pseudolikelihood approach yields the matrix in Equation (3.20), the information matrix obtained using the maximum likelihood approach. We therefore conclude that the pseudolikelihood is fully efficient for the bivariate normal distribution with parameters  $\sigma_1$ ,  $\sigma_2$  and  $\rho$ .

### 3.5.2 Efficiency for the trivariate normal distribution

Using the approach adopted throughout the current chapter, it is possible to show that both the full and partial pseudolikelihoods are fully efficient for the trivariate normal distribution with known means and unknown parameters  $\sigma_1$ ,  $\sigma_2$ ,  $\sigma_3$ ,  $\rho_{12}$ ,  $\rho_{13}$  and  $\rho_{23}$ . The calculations involved were done in Maple, and most of the formulae are too complicated to include. We therefore include only the elements of the information matrix  $\mathbf{I}$  obtained from both the full likelihood and pseudolikelihood approaches.

For conciseness, we shall introduce the notations  $\Pi\rho = \rho_{12}\rho_{13}\rho_{23}$  and  $\Sigma\rho^2 = \rho_{12}^2 + \rho_{13}^2 + \rho_{23}^2$ .

We have, for  $i, j, k = 1, 2, 3$  and  $i \neq j \neq k$ ,

$$\begin{aligned}
I_{\sigma_i \sigma_i} &= \frac{2 - \rho_{jk}^2 - \Sigma \rho^2 + 2\Pi \rho}{\sigma_1^2(1 - \Sigma \rho^2 + 2\Pi \rho)} \\
I_{\sigma_i \sigma_j} &= \frac{\Pi \rho - \rho_{ij}^2}{\sigma_i \sigma_j (1 - \Sigma \rho^2 + 2\Pi \rho)} \\
I_{\rho_{ij} \rho_{ij}} &= \frac{1 + \rho_{ij}^2 - \rho_{ik}^2 - \rho_{jk}^2 - 2\rho_{ik}^2 \rho_{jk}^2 - 2\Pi \rho}{(1 - \Sigma \rho^2 + 2\Pi \rho)^2} \\
I_{\rho_{ij} \rho_{ik}} &= \frac{2\rho_{ij} \rho_{ik} - \rho_{ij}^2 \rho_{jk} - \rho_{ik}^2 \rho_{jk} - \rho_{jk}(1 - \rho_{jk}^2)}{(1 - \Sigma \rho^2 + 2\Pi \rho)^2} \\
I_{\sigma_i \rho_{ij}} &= \frac{\rho_{ik} \rho_{jk} - \rho_{ij}}{\sigma_1(1 - \Sigma \rho^2 + 2\Pi \rho)} \\
I_{\sigma_i \rho_{jk}} &= 0.
\end{aligned}$$

The expressions for the  $\mathbf{H}$  and  $\mathbf{J}$  matrices based on a pseudolikelihood approach are more complicated, without being particularly enlightening, and are therefore not given.

### 3.6 Comparison of estimating equations

An alternative way of assessing the pseudolikelihood approach is to compare the estimates themselves with those based on a full likelihood approach. Here, we consider the five-parameter bivariate normal distribution and, by comparing the estimating equations for the two approaches, show that the MLE's are also solutions to the estimating equations for the pseudolikelihood approach. Since we are here concerned with the bivariate case, the two pseudolikelihoods FPL and PPL are equivalent, and in this section we will simply use the notation 'PL'.

Suppose  $Y_1$  and  $Y_2$  follow a (five-parameter) bivariate Normal distribution. Then the conditional distributions of  $Y_1|Y_2$  and  $Y_2|Y_1$  are

$$(Y_1|Y_2) \sim N\left(\mu_1 + \rho \frac{\sigma_1}{\sigma_2}(y_2 - \mu_2), \sigma_1^2(1 - \rho^2)\right)$$



and

$$(Y_2|Y_1) \sim N\left(\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(y_1 - \mu_1), \sigma_2^2(1 - \rho^2)\right)$$

respectively. Given a sample of size  $n$ , the maximum likelihood estimates based on the full likelihood are

$$\begin{aligned} \hat{\mu}_1 = \bar{y}_1 = \sum y_{1i}/n; \quad \hat{\mu}_2 = \bar{y}_2 = \sum y_{2i}/n; \quad \hat{\rho}\hat{\sigma}_1\hat{\sigma}_2 = S_{12} = \sum \frac{(y_{1i} - \mu_1)(y_{2i} - \mu_2)}{n} \\ \hat{\sigma}_1^2 = S_1^2 = \sum (y_{1i} - \mu_1)^2/n; \quad \hat{\sigma}_2^2 = S_2^2 = \sum (y_{2i} - \mu_2)^2/n. \end{aligned} \quad (3.30)$$

Consider now the pseudolikelihood

$$PL = \prod_{i=1}^n f(y_{1i}|y_{2i})f(y_{2i}|y_{1i}) \propto [\sigma_1^2(1 - \rho^2)]^{-n/2} [\sigma_2^2(1 - \rho^2)]^{-n/2} \exp\{E\} \quad (3.31)$$

where

$$E = -\frac{1}{2} \sum_{i=1}^n \left[ \frac{\left\{ (y_{1i} - \mu_1) - \rho \frac{\sigma_1}{\sigma_2}(y_{2i} - \mu_2) \right\}^2}{\sigma_1^2(1 - \rho^2)} + \frac{\left\{ (y_{2i} - \mu_2) - \rho \frac{\sigma_2}{\sigma_1}(y_{1i} - \mu_1) \right\}^2}{\sigma_2^2(1 - \rho^2)} \right].$$

Taking the logarithm of (3.31) after rearranging its exponent gives

$$Pl = \log(PL) = -\frac{n}{2} \log \sigma_1^2 - \frac{n}{2} \log \sigma_2^2 - n \log(1 - \rho^2) - \frac{S}{2\sigma_1^2\sigma_2^2(1 - \rho^2)} \quad (3.32)$$

where

$$\begin{aligned} S = \sum_{i=1}^n \{ \sigma_2^2(y_{1i} - \mu_1)^2(1 + \rho^2) + \sigma_1^2(y_{2i} - \mu_2)^2(1 + \rho^2) - \\ 4\rho\sigma_1\sigma_2(y_{1i} - \mu_1)(y_{2i} - \mu_2) \}. \end{aligned}$$

Writing

$$\begin{aligned} \sum_{i=1}^n (y_{1i} - \mu_1)^2 &= \sum_{i=1}^n [(y_{1i} - \bar{y}_1) + (\bar{y}_1 - \mu_1)]^2 \\ &= \sum_{i=1}^n (y_{1i} - \bar{y}_1)^2 + n(\bar{y}_1 - \mu_1)^2 \end{aligned}$$

and similarly for  $\sum (y_{2i} - \mu_2)^2$  and  $\sum (y_{1i} - \mu_1)(y_{2i} - \mu_2)$ , we see that (3.32) is maximised with respect to  $\mu_1$  and  $\mu_2$  when

$$m = \sigma_2^2 (1 + \rho^2) (\bar{y}_1 - \mu_1)^2 + \sigma_1^2 (1 + \rho^2) (\bar{y}_2 - \mu_2)^2 - 4\rho\sigma_1\sigma_2(\bar{y}_1 - \mu_1)(\bar{y}_2 - \mu_2) \quad (3.33)$$

is minimised with respect to these two parameters. We get  $\mu_1 = \bar{y}_1$  and  $\mu_2 = \bar{y}_2$ , so that the maximum pseudo likelihood estimates of  $\mu_1$  and  $\mu_2$  are equal to the full maximum likelihood estimates.

Substituting the estimates of  $\mu_1$  and  $\mu_2$  back into (3.32) gives

$$-\frac{n}{2} \log \sigma_1^2 - \frac{n}{2} \log \sigma_2^2 - n \log (1 - \rho^2) - F \quad (3.34)$$

where  $S_1^2$ ,  $S_2^2$  and  $S_{12}$  are as defined in (3.30) and

$$F = \frac{n [\sigma_1^2 S_2^2 (1 + \rho^2) + \sigma_2^2 S_1^2 (1 + \rho^2) - 4\rho\sigma_1\sigma_2 S_{12}]}{2\sigma_1^2 \sigma_2^2 (1 - \rho^2)}.$$

Differentiating (3.34) with respect to  $\sigma_1^2$ , setting it equal to zero and multiplying throughout by  $-2/n$  gives

$$\begin{aligned} \frac{1}{\sigma_1^2} + \left\{ \left[ S_2^2 (1 + \rho^2) - 2\rho \frac{\sigma_2^2}{\sigma_1} S_{12} \right] [\sigma_1^2 \sigma_2^2 (1 - \rho^2)] - \right. \\ \left. - [\sigma_2^2 (1 - \rho^2)] [\sigma_1^2 S_2^2 (1 + \rho^2) + \sigma_2^2 S_1^2 (1 + \rho^2) - 4\rho\sigma_1\sigma_2 S_{12}] \right\} / \sigma_1^4 \sigma_2^2 (1 - \rho^2). \end{aligned}$$

Multiplying throughout by  $\sigma_1^3 \sigma_2^2 (1 - \rho^2)$  and rearranging gives

$$\sigma_1 (1 - \rho^2) \sigma_2^2 = (1 + \rho^2) \frac{\sigma_2^2}{\sigma_1} S_1^2 - 2\rho\sigma_2 S_{12}. \quad (3.35)$$

By symmetry, the corresponding equation for  $\sigma_2$  is

$$\sigma_2 (1 - \rho^2) \sigma_1^2 = (1 + \rho^2) \frac{\sigma_1^2}{\sigma_2} S_2^2 - 2\rho\sigma_1 S_{12}. \quad (3.36)$$

Differentiating (3.34) with respect to  $\rho$ , setting the derivative equal to zero and multiplying throughout by  $\sigma_1^2 \sigma_2^2 (1 - \rho^2)^2$  gives

$$\begin{aligned} 2\rho (1 - \rho^2) \sigma_1^2 \sigma_2^2 = \rho [(1 + \rho^2) (\sigma_1^2 S_2^2 + \sigma_2^2 S_1^2) - 4\rho\sigma_1\sigma_2 S_{12}] + \\ (1 - \rho^2) (\rho\sigma_1^2 S_2^2 + \rho\sigma_2^2 S_1^2 - 2\sigma_1\sigma_2 S_{12}). \end{aligned} \quad (3.37)$$

Substituting the maximum likelihood estimates  $\hat{\sigma}_1^2$ ,  $\hat{\sigma}_2^2$  and  $\hat{\rho}$  from (3.30) into (3.35), (3.36) and (3.37) verifies that the maximum likelihood estimates also give solutions to the estimating equations for the pseudo likelihood, as the left hand sides and right hand sides are indeed equal upon making these substitutions. This indicates that the pseudolikelihood estimates are fully efficient.

### 3.7 Conclusions

By considering multivariate normal distributions with various restrictions on parameters, we have shown in this chapter that (except for the case of known  $\sigma$ ) both the full and partial pseudolikelihood are fully efficient when all variances are equal, all correlations are equal and means are known.

More generally, as would be expected given the results for the bivariate and trivariate cases described in this chapter, it turns out that both pseudolikelihoods are fully efficient for the general  $p$ -variate normal distribution (Mardia et al., 2007a).

The approach considered also provides an insight into the efficiency of pseudolikelihood estimators for the multivariate von Mises distribution when the circular variables are highly concentrated. In the next chapter, we will discuss parameter estimation for the multivariate von Mises distribution, calculating the efficiency of estimators in the bivariate case and comparing maximum likelihood estimates with pseudolikelihood estimates in the trivariate case, both in terms of accuracy and in terms of computational expense.



## Chapter 4

# Parameter estimation and inference for the multivariate von Mises distribution

### 4.1 Introduction

In this chapter we discuss parameter estimation and inference for the multivariate von Mises distribution introduced in Chapter 2. This builds on the preliminary results of Singh and Mardia (2004). We discuss the full pseudolikelihood in the context of the bivariate and trivariate von Mises distributions, deriving first the efficiency for the bivariate case in Section 4.2. Simulation from the bivariate and trivariate models was discussed in Chapter 2, and in Section 4.2.4 we simulate data from the bivariate model that support the efficiency calculations and show the bias of the MLEs and MPLEs to be very similar. In Section 4.3, the pseudolikelihood is investigated for the trivariate case by considering properties of the pseudolikelihood estimates themselves and comparing them with maximum likelihood estimates, for various parameter configurations and based on simulated data. The Gibbs sampler data simulation method outlined in Section 2.4 is also scrutinized. Finally, in Section 4.4, we apply the trivariate von Mises distribution

to two types of protein data set, one of them protein folds and the other amino acid conformations. We also investigate hypothesis testing for the trivariate model in the context of these data.

As defined in Chapter 3, the full pseudolikelihood, based on a random sample of  $n$  observations of  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_p)^T$ , is given by

$$\text{FPL}(\mathbf{Y}; \mathbf{q}) = \prod_{j=1}^p \prod_{i=1}^n g_j(Y_{ji} | \text{rest}; \mathbf{q}) \quad (4.1)$$

where  $g_j(Y_j | \text{rest}; \mathbf{q})$  is the conditional distribution of  $Y_j$  given  $(Y_1, \dots, Y_{j-1}, Y_{j+1}, \dots, Y_p)$  and  $\mathbf{q}$  is an unknown parameter vector of length  $r$ . For the  $p$ -variate von Mises distribution introduced in Chapter 2, we have the vector  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^T$  of observations. Under high concentration, this follows an approximate  $p$ -variate normal distribution, in which case the methods of Chapter 3 are appropriate. The full pseudolikelihood for the  $p$ -variate von Mises case is given by

$$\text{FPL} = (2\pi)^{-pn} \prod_{j=1}^p \prod_{i=1}^n \left[ I_0(\kappa_{j\cdot\text{rest}}^{(i)}) \right]^{-1} \exp \left\{ \kappa_{j\cdot\text{rest}}^{(i)} \cos(\theta_{ji} - \mu_{j\cdot\text{rest}}^{(i)}) \right\}$$

where

$$\begin{aligned} \mu_{j\cdot\text{rest}}^{(i)} &= \mu_j + \tan^{-1} \left\{ \left[ \sum_{l \neq j} \lambda_{jl} \sin(\theta_{li} - \mu_l) \right] / \kappa_j \right\}, \\ \kappa_{j\cdot\text{rest}}^{(i)} &= \left\{ \kappa_j^2 + \left[ \sum_{l \neq j} \lambda_{jl} \sin(\theta_{li} - \mu_l) \right]^2 \right\}^{1/2} \end{aligned}$$

and  $\mu_{j\cdot\text{rest}}$  and  $\kappa_{j\cdot\text{rest}}$  are the mean direction and concentration parameter respectively of the conditional distribution of  $\Theta_j$  given all other  $\theta$  values, which was shown by Theorem 2.3.1 in Chapter 2 to be univariate von Mises. The vector  $\mathbf{q}$  of unknown observations in this case is  $\mathbf{q} = (\kappa_1, \kappa_2, \kappa_3, \lambda_{12}, \lambda_{13}, \lambda_{23}, \mu_1, \mu_2, \mu_3)$  with length  $r = 9$ .

## 4.2 Efficiency of the pseudolikelihood for the bivariate von Mises distribution

In this section we use a numerical approach in order to assess the efficiency of the pseudolikelihood for the bivariate von Mises distribution with unknown  $\kappa_1 = \kappa_2 = \kappa$ , say, known  $\mu_1 = \mu_2 = 0$  and unknown parameter  $\lambda$ . The approach outlined in Chapter 3.2, based on comparison of information matrices, will be used. Since we are dealing with the bivariate case, the full pseudolikelihood FPL and the pairwise pseudolikelihood PPL are equivalent. The joint probability density function  $f(\theta_1, \theta_2)$  of the bivariate von Mises distribution is given by

$$C^{-1} \exp \{ \kappa_1 \cos(\theta_1 - \mu_1) + \kappa_2 \cos(\theta_2 - \mu_2) + \lambda \sin(\theta_1 - \mu_1) \sin(\theta_2 - \mu_2) \}, \quad (4.2)$$

where the normalizing constant's inverse  $C = C(\kappa_1, \kappa_2, \lambda)$  was given by Singh et al. (2002) as

$$C = 4\pi^2 \sum_{m=0}^{\infty} \binom{2m}{m} \left( \frac{\lambda}{2} \right)^{2m} \kappa_1^{-m} I_m(\kappa_1) \kappa_2^{-m} I_m(\kappa_2). \quad (4.3)$$

and  $I_\nu(\cdot)$  is the modified Bessel function of the first kind and order  $\nu$ .

The following identity, adapted from Abramowitz and Stegun (1972), will be used to obtain derivatives of Bessel functions:

$$\frac{d}{dx} \left[ \frac{I_\nu(x)}{x^\nu} \right] = \frac{I_{\nu+1}(x)}{x^\nu}. \quad (4.4)$$

### 4.2.1 Fisher information for the full likelihood approach

We consider the joint density ((4.2) and (4.3)) with  $\kappa_1 = \kappa_2 = \kappa$  and  $\mu_1 = \mu_2 = 0$ . When calculating the Fisher information matrix from the joint pdf of the distribution, the matrices  $\mathbf{J}_f$  and  $\mathbf{H}_f$  are equal. We have

$$\log f(\theta_1, \theta_2) = \kappa \cos \theta_1 + \kappa \cos \theta_2 + \lambda \sin \theta_1 \sin \theta_2 - \log C,$$

and since the second derivatives of this expression with respect to  $\kappa$  and  $\lambda$  are independent of  $\theta_1$  and  $\theta_2$ , the elements of  $\mathbf{H}_f$  can be found directly as

$$\begin{aligned} [\mathbf{H}_f]_{11} &= -\frac{\partial^2 \log f}{\partial \kappa^2} = C^{-1} \frac{\partial^2 C}{\partial \kappa^2} - C^{-2} \left( \frac{\partial C}{\partial \kappa} \right)^2 \\ [\mathbf{H}_f]_{12} &= [\mathbf{H}_f]_{21} = -\frac{\partial^2 \log f}{\partial \kappa \partial \lambda} = C^{-1} \frac{\partial^2 C}{\partial \kappa \partial \lambda} - C^{-2} \frac{\partial C}{\partial \kappa} \frac{\partial C}{\partial \lambda} \\ [\mathbf{H}_f]_{22} &= -\frac{\partial^2 \log f}{\partial \lambda^2} = C^{-1} \frac{\partial^2 C}{\partial \lambda^2} - C^{-2} \left( \frac{\partial C}{\partial \lambda} \right)^2. \end{aligned} \quad (4.5)$$

Using the identity in Equation (4.4) we now calculate the derivatives in (4.5). We get

$$\frac{\partial C}{\partial \kappa} = 8\pi^2 \sum_{m=0}^{\infty} \binom{2m}{m} \left( \frac{\lambda}{2} \right)^{2m} [\kappa^{-m} I_m(\kappa)] [\kappa^{-m} I_{m+1}(\kappa)]. \quad (4.6)$$

Writing  $\kappa^{-m} = \kappa \kappa^{-(m+1)}$ , and using (4.4) and the chain rule, we find

$$\begin{aligned} \frac{\partial^2 C}{\partial \kappa^2} &= 8\pi^2 \sum_{m=0}^{\infty} \binom{2m}{m} \left( \frac{\lambda}{2} \right)^{2m} \left\{ \kappa^{-m} I_m(\kappa) \left[ \kappa \kappa^{-(m+1)} I_{m+2}(\kappa) + \right. \right. \\ &\quad \left. \left. + \kappa^{-(m+1)} I_{m+1}(\kappa) \right] + [\kappa^{-m} I_{m+1}(\kappa)]^2 \right\}, \end{aligned}$$

whilst differentiating (4.6) with respect to  $\lambda$  gives

$$\frac{\partial^2 C}{\partial \kappa \partial \lambda} = 8\pi^2 \sum_{m=0}^{\infty} \binom{2m}{m} 2m \lambda^{2m-1} 2^{-2m} [\kappa^{-m} I_m(\kappa)] [\kappa \kappa^{-(m+1)} I_{m+1}(\kappa)].$$

Finally,

$$\frac{\partial C}{\partial \lambda} = 4\pi^2 \sum_{m=0}^{\infty} \binom{2m}{m} 2m \lambda^{2m-1} 2^{-2m} [\kappa^{-m} I_m(\kappa)]^2$$

and

$$\frac{\partial^2 C}{\partial \lambda^2} = 4\pi^2 \sum_{m=0}^{\infty} \binom{2m}{m} 2m(2m-1) \lambda^{2m-2} 2^{-2m} [\kappa^{-m} I_m(\kappa)]^2. \quad (4.7)$$

Substituting  $C$  (with  $\kappa_1 = \kappa_2 = \kappa$ ) from Equation (4.3), and its derivatives from Equations (4.6)–(4.7), into Equation (4.5), gives the Fisher information matrix  $\mathbf{I}_f = \mathbf{H}_f = \mathbf{J}_f$  for the full likelihood approach. For comparison, we now consider the Fisher information matrix obtained for the pseudolikelihood approach.



### 4.2.2 Information matrix for the pseudolikelihood approach

Since the marginal density of  $\theta_1$  can be written as (see Equation (1.14))

$$f(\theta_1) = 2\pi C^{-1} I_0(a_1) e^{\kappa \cos \theta_1},$$

where  $C$  is as in Equation (4.3) and  $a_1 = \sqrt{(\kappa^2 + \lambda^2 \sin^2 \theta_1)}$ , and similarly for  $f(\theta_2)$ , the logarithm of the product of conditional densities  $\log[f(\theta_1|\theta_2)f(\theta_2|\theta_1)] = \log\{f(\theta_1, \theta_2)^2/[f(\theta_1)f(\theta_2)]\}$  for a pseudolikelihood approach can be written

$$\log g_{\text{PL}} = -2 \log 2\pi + \kappa \cos \theta_1 + \kappa \cos \theta_2 + 2\lambda \sin \theta_1 \sin \theta_2 - \log I_0(a_1) - \log I_0(a_2)$$

where  $a_2 = \sqrt{(\kappa^2 + \lambda^2 \sin^2 \theta_2)}$ .

Noting that  $\partial a_j / \partial \kappa = \kappa / a_j$  and  $\partial a_j / \partial \lambda = (\lambda \sin^2 \theta_j) / a_j$ ,  $j = 1, 2$ , and using Equation (4.4), we find

$$\frac{\partial g_{\text{PL}}}{\partial \kappa} = \cos \theta_1 + \cos \theta_2 - \kappa \frac{A_1(a_1)}{a_1} - \kappa \frac{A_1(a_2)}{a_2} \quad (4.8)$$

where  $A_\nu(\cdot) = I_\nu(\cdot) / I_0(\cdot)$ . Similarly

$$\frac{\partial g_{\text{PL}}}{\partial \lambda} = 2 \sin \theta_1 \sin \theta_2 - \lambda \sin^2 \theta_1 \frac{A_1(a_1)}{a_1} - \lambda \sin^2 \theta_2 \frac{A_1(a_2)}{a_2} \quad (4.9)$$

Now

$$\frac{\partial}{\partial \kappa} \left[ \kappa \frac{A_1(a_j)}{a_j} \right] = \frac{\kappa^2}{a_j^2} \left[ A_2(a_j) + \frac{A_1(a_j) a_j}{\kappa^2} - A_1(a_j)^2 \right].$$

Therefore

$$\frac{\partial^2 g_{\text{PL}}}{\partial \kappa^2} = \sum_{j=1}^2 \frac{\kappa^2}{a_j^2} \left[ A_1(a_j)^2 - A_2(a_j) - A_1(a_j) \frac{a_j}{\kappa^2} \right]. \quad (4.10)$$

Using a similar approach, or by consideration of the symmetry involved, we find

$$\frac{\partial^2 g_{\text{PL}}}{\partial \lambda^2} = \sum_{j=1}^2 \frac{\lambda^2 \sin^4 \theta_j}{a_j^2} \left[ A_1(a_j)^2 - A_2(a_j) - A_1(a_j) \frac{a_j}{\lambda^2 \sin^2 \theta_j} \right] \quad (4.11)$$

and

$$\frac{\partial^2 g_{\text{PL}}}{\partial \kappa \partial \lambda} = \sum_{j=1}^2 \frac{\kappa \lambda \sin^2 \theta_j}{a_j^2} [A_1(a_j)^2 - A_2(a_j)]. \quad (4.12)$$

In order to obtain the matrices  $\mathbf{H}_{g_{PL}}$  and  $\mathbf{J}_{g_{PL}}$  for the pseudolikelihood approach, we need to take expectations of the expressions (or products or squares of expressions in the case of  $\mathbf{J}_{g_{PL}}$ ) in Equations (4.8)–(4.12). To do this, we multiply the expression of interest by the joint pdf (4.2), and integrate the resulting expressions with respect to  $\theta_1$  and  $\theta_2$ , each integral over the interval  $(0, 2\pi)$ . As these double integrals are intractable analytically, numerical integration is performed using the `adapt` function in R for the desired  $\kappa$  and  $\lambda$  values. The information matrix for the pseudolikelihood approach is then obtained numerically using Equation (3.3). For any pair of parameter values  $(\kappa, \lambda)$ , we now have a means of calculating the information matrix for both the full likelihood approach and the pseudolikelihood approach, and can therefore calculate the relative efficiency of the two methods using Equation (3.4).

The terms in the infinite sums involved become small very quickly for  $m$  greater than about 3. For all  $(\kappa, \lambda)$  parameter configurations to be used in Section 4.2.3, we compared the ratio of the first seventy terms of Equation (4.3) to the result obtained by integrating numerically the exponential of the joint pdf (4.2). This ratio was found to be unity in each case. In order to compute the infinite sums, therefore, we take the first seventy terms in each case. For those parameter configurations to be considered, we observe the largest 70th term to be  $10^{-78}$ .

### 4.2.3 Efficiency of the pseudolikelihood

The efficiency of the pseudolikelihood, for any pair of values  $(\kappa, \lambda)$ , can be calculated from Equation (3.4). We compute the efficiency for all combinations of  $\kappa = 1, 2, \dots, 15$  and  $|\lambda| = 0.5$  and  $1, 2, \dots, 15$  (efficiency for  $\lambda = l$  is the same as for  $\lambda = -l$ ). If  $\kappa_1 = \kappa_2 = 0$  then the expression in Equation (4.3) for the normalizing constant is not suitable. In fact the normalizing constant in this case appears to involve beta functions. Since  $\kappa_1 = \kappa_2 = 0$  is something of a degenerate case, we will not consider  $\kappa = 0$  in the

following.

The results of the efficiency calculations are displayed in Figures 4.1 and 4.2. From Figure 4.1 we can more easily consider the functional relationship between efficiency and the parameter values for either  $\kappa$  or  $\lambda$  fixed and the other varying, whilst Figure 4.2 contains only selected values of  $|\lambda|$  (as shown on the legend), and gives a better idea of efficiency values.

For  $\lambda = 0$  the efficiency is unity, since in this case  $f(\theta_1, \theta_2) = f(\theta_1|\theta_2)f(\theta_2|\theta_1)$ . We observe that efficiency is high for small  $|\lambda|$ . It was shown in Section 3.3 that both pseudolikelihoods are fully efficient for the bivariate normal distribution with known  $\mu_1 = \mu_2 = 0$  and unknown parameters  $\sigma$  and  $\rho$ . Since the bivariate von Mises distribution tends to a normal distribution for large  $\kappa$  values (Singh et al., 2002), the efficiency should tend to unity in the current case as  $\kappa \rightarrow \infty$ . From Singh et al. (2002), for the bivariate case with  $\mu_1 = \mu_2 = 0$  and  $\kappa_1 = \kappa_2 = \kappa$ , the corresponding parameters of the normal distribution are

$$\sigma^2 = \frac{\kappa}{\kappa^2 - \lambda^2}, \quad \rho = \frac{\lambda}{\kappa}$$

In order that  $\sigma^2 > 0$  we require  $\kappa > |\lambda|$ . The aforementioned improvement in efficiency as  $\kappa$  increases should therefore only be expected to occur, for fixed  $\lambda$ , for those regions in Figure 4.2 in which  $\kappa > |\lambda|$ . This is indeed the case, and the efficiency is greater than 0.9 for all such  $\kappa, \lambda$  pairs in Figure 4.2. For  $\kappa < \lambda$ , the joint distribution of  $\theta_1$  and  $\theta_2$  is bimodal (Mardia et al., 2007b), and in this case the efficiency appears to be a quadratic function of  $\kappa$ , with greater efficiency for very small  $\kappa$  and  $\kappa$  close to  $\lambda$ .

#### 4.2.4 Simulation

We now investigate the properties of the pseudolikelihood estimator for the bivariate von Mises distribution by simulating 10 000 bivariate samples for each of four different parameter configurations and for each of three sample sizes. The rejection sampling

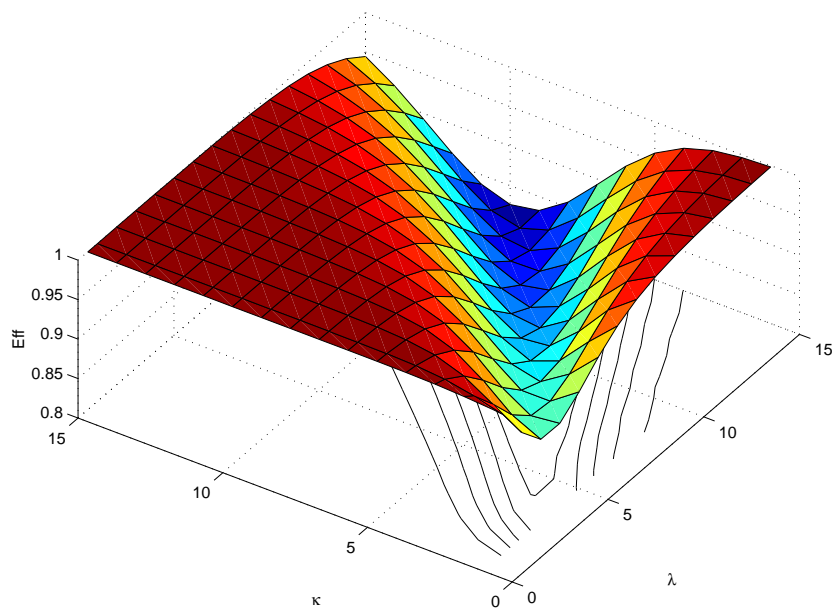


Figure 4.1: Efficiency of pseudolikelihood for bivariate von Mises distribution with  $\kappa = 1, 2, \dots, 15$ ,  $\lambda = 0.5, 1, 2, \dots, 15$ .

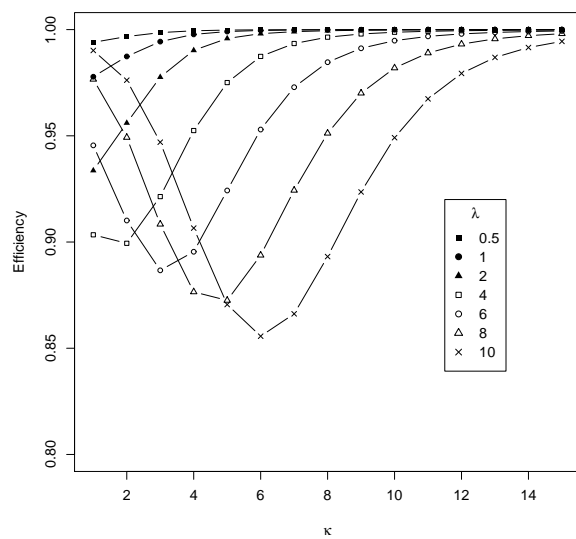


Figure 4.2: Efficiency of pseudolikelihood for bivariate von Mises distribution with  $\kappa = 1, 2, \dots, 10$ ,  $|\lambda| = 0.5, 1, 2, 4, 6, 8, 10$ .

algorithm described in Chapter 2.4 is used for the data simulation. The configurations  $(\kappa, \lambda) = (2, 0.5), (8, 4), (6, 6)$  and  $(2, 4)$  will be used for samples of sizes 20, 50 and 100. Figure 4.3 displays the estimated bias and variance plotted against sample size for each parameter configuration. The efficiency values for the pseudolikelihood for these parameter configurations, as calculated in Section 4.2.3 and displayed in Figures 4.1 and 4.2, are, in the order stated above and plotted, 0.997, 0.996, 0.953 and 0.899. Certainly the variance plots exemplify this ordering of efficiency, and variances of parameter estimates for the pseudolikelihood and full likelihood are closely comparable when the sample size is at least 50. Even in the worst case scenario – when  $(\kappa, \lambda) = (2, 4)$  (which corresponds to a bimodal density) with a sample size of 20 – the ratio of variances full:pseudo for the estimate of  $\lambda$  is 0.834, and clearly greater for  $\kappa$ . There is also a slightly greater bias for the pseudolikelihood with this parameter configuration, though in general the bias is very similar for the two methods. When  $(\kappa, \lambda) = (8, 4)$  there is even the slight suggestion of less bias for the pseudolikelihood estimates.

### 4.3 Data simulation and parameter estimation for the trivariate von Mises distribution

In this section we investigate data simulation and parameter estimation for the trivariate von Mises distribution. The Gibbs sampling technique described in Section 2.4 is used to simulate data, and the properties of pseudolikelihood estimates are investigated and compared with maximum likelihood estimates. The full pseudolikelihood for the trivariate von Mises distribution is given by

$$\begin{aligned}
 (2\pi)^{-3n} \prod_{i=1}^n [I_0(\kappa_{1,2,3}^{(i)}) I_0(\kappa_{2,1,3}^{(i)}) I_0(\kappa_{3,1,2}^{(i)})]^{-1} \exp \{ \kappa_{1,2,3}^{(i)} \cos(\theta_{1i} - \mu_{1,2,3}^{(i)}) + \\
 + \kappa_{2,1,3}^{(i)} \cos(\theta_{2i} - \mu_{2,1,3}^{(i)}) + \kappa_{3,1,2}^{(i)} \cos(\theta_{3i} - \mu_{3,1,2}^{(i)}) \}
 \end{aligned} \tag{4.13}$$

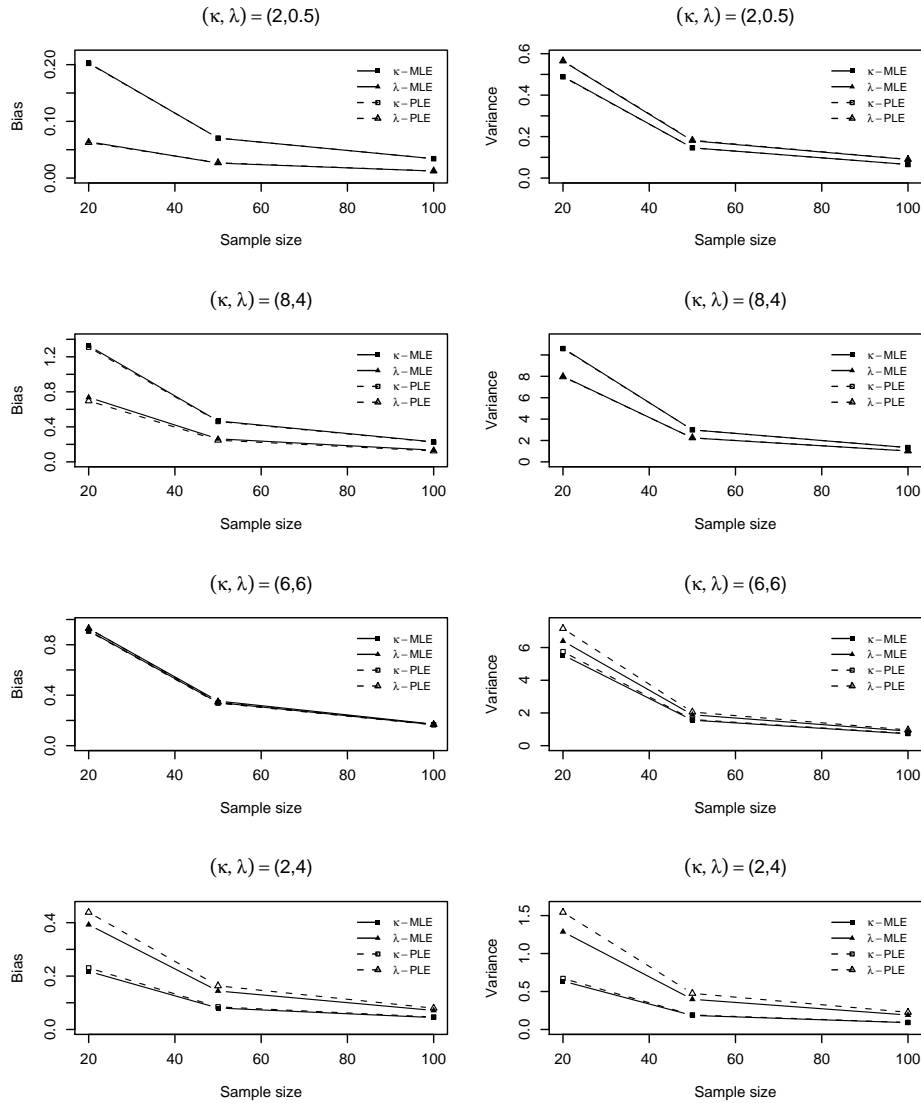


Figure 4.3: Estimates of bias and variance for pseudolikelihood and full likelihood estimators for the bivariate von Mises distribution. Based on 10 000 simulations for each configuration and samples of sizes 20, 50 and 100.

where

$$\begin{aligned}\mu_{1,2,3}^{(i)} &= \mu_1 + \tan^{-1}\{[\lambda_{12} \sin(\theta_2^{(i)} - \mu_2) + \lambda_{13} \sin(\theta_3^{(i)} - \mu_3)]/\kappa_1\} \\ \kappa_{1,2,3}^{(i)} &= \sqrt{\{\kappa_1^2 + [\lambda_{12} \sin(\theta_2^{(i)} - \mu_2) + \lambda_{13} \sin(\theta_3^{(i)} - \mu_3)]^2\}}\end{aligned}$$

and analogously for  $\mu_{2,1,3}^{(i)}$ ,  $\kappa_{2,1,3}^{(i)}$ ,  $\mu_{3,1,2}^{(i)}$  and  $\kappa_{3,1,2}^{(i)}$ .

### 4.3.1 Data simulation and pseudolikelihood estimation

For chosen parameters  $\kappa_1$ ,  $\kappa_2$ ,  $\kappa_3$ ,  $\lambda_{12}$ ,  $\lambda_{13}$  and  $\lambda_{23}$ , the following Gibbs sampling method will be used in order to simulate variates from the trivariate von Mises distribution. For the simulation method, the mean direction vector will be set to zero. For subsequent estimation, mean directions will be added to (zero–mean) simulated data.

- Firstly, two independent von Mises distributed vectors of dimension  $n_1$  are simulated with concentration parameters  $\kappa_1$  and  $\kappa_2$  respectively. These are the initial vectors representing  $\theta_1$  and  $\theta_2$  respectively.
- Next, using the parameters  $\kappa_3$ ,  $\lambda_{13}$  and  $\lambda_{23}$ , a vector of  $\theta_3$  values is simulated such that  $\theta_{3i}$  is von Mises with mean direction  $\mu_{3,1,2}^{(i)}$  and concentration parameter  $\kappa_{3,1,2}^{(i)}$ , as defined for Equation (4.13) and given the  $\theta_1$  and  $\theta_2$  vectors ( $i = 1, \dots, n_1$ ).
- The simulation proceeds by cycling through columns one at a time, each time replacing the values therein with new, von Mises distributed data generated conditionally on the current values in the other two columns. This loop is done  $n_2$  times.
- $n_3$  such data sets are generated giving, in total,  $n_1 n_3$  vectors  $(\theta_1, \theta_2, \theta_3)$ .

Parameters are estimated by minimising the negative logarithm of the pseudolikelihood (4.13). The parameter values chosen are  $(\kappa_1, \kappa_2, \kappa_3, \lambda_{12}, \lambda_{13}, \lambda_{23}) = (2, 3, 1, 2, 2, 2)$ . We

experiment with three different estimation methods, two values of  $n_1$  and two values of  $n_2$ . Tables 4.1–4.3 summarise the results:

#### **Table 4.1**

Table 4.1 gives means and standard deviations of pseudolikelihood estimates for simulated trivariate von Mises data with zero means. Panels (a), (b) and (c) show the effect of changing the values of  $n_1$  (number of data points for each  $\theta_j$ ) and  $n_2$  (number of times each vector is updated in the Gibbs sampler).

#### **Table 4.2**

As for Table 4.1 except that a mean vector  $(2, 3, 1)$  is added to the simulated data and estimated using nlm and the pseudolikelihood. In this case the means for  $\mu_1$ ,  $\mu_2$  and  $\mu_3$  in the table are the sample mean directions of the nlm estimates, whilst the figures in the (SD) column are  $1 - \bar{R}$ , where  $\bar{R}$  is the mean resultant length of the nlm estimates.

#### **Table 4.3**

As for Table 4.2 except that mean directions are estimated using the sample means of each variable in each data set, and nlm estimates are calculated for the remaining parameters using mean (direction) corrected data. *Between* tables but *within* panels (a), (b) and (c), the parameters are estimated using the same data.

### **Conclusions**

The results of the above procedure are a means of assessing the adequacy of a number of things: the simulation procedure and the number of iterations of the Gibbs sampler; nlm as an estimation algorithm, and the use of the pseudolikelihood as an estimation method. The starting values of parameters used in the nlm algorithm are  $(\kappa_1, \kappa_2, \kappa_3, \lambda_{12}, \lambda_{13}, \lambda_{23}, \mu_1, \mu_2, \mu_3) = (\hat{\kappa}_1, \hat{\kappa}_2, \hat{\kappa}_3, 0, 0, 0, \bar{\theta}_1, \bar{\theta}_2, \bar{\theta}_3)$ , where  $\hat{\kappa}_j$  is the univariate maximum likelihood estimate of  $\kappa_j$  based on the  $\theta_j$ 's,  $\bar{\theta}_j$  is the sample mean direction of the  $\theta_j$ 's for each data set and for  $j = 1, 2, 3$ . These latter are used only when



	(a)		(b)		(c)	
	$n_1 = 100$	$n_2 = 100$	$n_1 = 100$	$n_2 = 50$	$n_1 = 200$	$n_2 = 100$
	Mean	(SD)	Mean	(SD)	Mean	(SD)
$\kappa_1$	2.02	(0.27)	2.08	(0.33)	2.06	(0.22)
$\kappa_2$	3.15	(0.44)	3.07	(0.42)	3.02	(0.31)
$\kappa_3$	1.04	(0.24)	1.01	(0.22)	1.03	(0.13)
$\lambda_{12}$	2.01	(0.60)	2.03	(0.54)	2.03	(0.39)
$\lambda_{13}$	2.06	(0.40)	2.05	(0.46)	2.08	(0.35)
$\lambda_{23}$	2.13	(0.45)	2.07	(0.53)	2.02	(0.34)

Table 4.1: Mean values and standard deviations of nlm estimates for generated trivariate von Mises data.

	(a)		(b)		(c)	
	$n_1 = 100$	$n_2 = 100$	$n_1 = 100$	$n_2 = 50$	$n_1 = 200$	$n_2 = 100$
	Mean	(SD)	Mean	(SD)	Mean	(SD)
$\kappa_1$	2.03	(0.28)	2.09	(0.33)	2.07	(0.22)
$\kappa_2$	3.17	(0.45)	3.09	(0.43)	3.04	(0.31)
$\kappa_3$	1.05	(0.25)	1.01	(0.22)	1.04	(0.13)
$\lambda_{12}$	2.02	(0.60)	2.04	(0.55)	2.03	(0.39)
$\lambda_{13}$	2.08	(0.41)	2.08	(0.47)	2.10	(0.35)
$\lambda_{23}$	1.99	(0.47)	2.09	(0.54)	2.03	(0.35)
$\mu_1$	1.99	(0.002)	2.0	(0.002)	2.01	(0.001)
$\mu_2$	3.01	(0.002)	3.00	(0.002)	3.00	(0.001)
$\mu_3$	-1.00	(0.002)	-1.01	(0.002)	-1.00	(0.001)

Table 4.2: Mean values and standard deviations of nlm estimates for generated trivariate von Mises data with  $\boldsymbol{\mu} = (2, 3, -1)$ .

	(a)		(b)		(c)	
	$n_1 = 100$	$n_2 = 100$	$n_1 = 100$	$n_2 = 50$	$n_1 = 200$	$n_2 = 100$
	Mean	(SD)	Mean	(SD)	Mean	(SD)
$\kappa_1$	2.00	(0.28)	2.06	(0.32)	2.05	(0.22)
$\kappa_2$	3.11	(0.43)	3.02	(0.42)	3.00	(0.32)
$\kappa_3$	1.03	(0.24)	1.00	(0.22)	1.03	(0.12)
$\lambda_{12}$	2.13	(0.58)	2.15	(0.54)	2.07	(0.38)
$\lambda_{13}$	1.86	(0.42)	1.89	(0.50)	2.01	(0.35)
$\lambda_{23}$	1.98	(0.48)	1.90	(0.53)	1.95	(0.37)
$\mu_1$	2.00	(0.01)	2.01	(0.01)	2.02	(0.01)
$\mu_2$	3.01	(0.01)	3.01	(0.01)	3.00	(0.003)
$\mu_3$	-1.00	(0.04)	-0.99	(0.04)	-0.99	(0.02)

Table 4.3: Mean values and standard deviations of nlm estimates for generated trivariate von Mises data, once marginal circular means have been subtracted from the data columns.

we use `nlm` to estimate the mean directions.

A comparison of panels (a) and (b) of each table indicates that, at least with these parameter values, a value of  $n_2 = 50$  gives comparable accuracy to that achieved with  $n_2 = 100$ , and is enough to give reasonably accurate estimates. Comparing parts (a) and (c), we see that a value of  $n_1 = 200$  gives more accurate results than a value of  $n_1 = 100$  for fixed  $n_2, n_3$ . Overall, the accuracy of the estimates and the effect of changing  $n_1$  are encouraging signs that both the simulation method and the estimation method are effective even though, thus far, only one combination of parameter values has been tested. Comparison of Tables 4.1, 4.2 and 4.3 reveals little or no difference in the effectiveness of the estimation procedure for the three scenarios. In the case of Table 4.3, the sample mean directions serve as a good estimate of the  $\mu$  parameters, being very close to the true values  $(2, 3, -1)$ .

In order to investigate the effectiveness of the estimation procedure for small  $\kappa$  values, we first take  $\kappa_1 = 1$ ,  $\kappa_2 = 1.5$  and  $\kappa_3 = 0.5$ . With  $\lambda_{12} = \lambda_{13} = \lambda_{23} = 1$ , known  $\boldsymbol{\mu} = \mathbf{0}$  and using  $(n_1 = 100, n_2 = 50)$  the results shown in Table 4.4 are obtained. Clearly, the halving of the  $\kappa$  values has had no significant detrimental effect on the estimation.

	$\kappa_1$	$\kappa_2$	$\kappa_3$	$\lambda_{12}$	$\lambda_{13}$	$\lambda_{23}$
Mean	1.02	1.53	0.54	2.03	2.00	2.10
(SD)	(0.26)	(0.31)	(0.22)	(0.53)	(0.51)	(0.52)

Table 4.4: Mean values and standard deviations of nlm estimates for generated trivariate von Mises data with  $\kappa_1 = 1$ ,  $\kappa_2 = 1.5$ ,  $\kappa_3 = 0.5$  and  $\lambda_{12} = \lambda_{13} = \lambda_{23} = 1$

	$\kappa_1$	$\kappa_2$	$\kappa_3$	$\lambda_{12}$	$\lambda_{13}$	$\lambda_{23}$
Mean	0.53	0.80	0.28	2.10	2.02	2.06
(SD)	(0.24)	(0.21)	(0.17)	(0.52)	(0.53)	(0.54)

Table 4.5: Mean values and standard deviations of nlm estimates for generated trivariate von Mises data with  $\kappa_1 = 0.5$ ,  $\kappa_2 = 0.75$ ,  $\kappa_3 = 0.25$  and  $\lambda_{12} = \lambda_{13} = \lambda_{23} = 1$

Halving the true  $\kappa$  values again yields the results in Table 4.5. Again the estimates are adequate in terms of the mean values, although the standard deviation of the  $\kappa$  estimates increases relative to the size of the true parameter values as the true  $\kappa$  values become smaller.

### 4.3.2 Pseudolikelihood versus full likelihood

Having developed a data simulation method and verified that the pseudolikelihood provides a reasonable means of estimating parameters for the trivariate von Mises distribution, we next compare the properties of pseudolikelihood estimates with their full likelihood counterparts, for various parameter configurations. Tables 4.6 and 4.7 display maximum likelihood (ML) and pseudolikelihood (PL) estimates based on a single data

set for each of four different parameter configurations and using  $n_1 = 100$ ,  $n_2 = 50$  (see previous section). The first of these configurations corresponds to that studied in the previous section. The four configurations are also those four for which the univariate and bivariate marginals were plotted in Figures 2.1 and 2.2. The figures in brackets give approximate standard errors of estimates, and 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. Maximum likelihood estimates are obtained by incorporating a numerical integration into each stage of the `nlm` algorithm in order to evaluate the unknown normalizing constant.

	True	ML	(SD)	PL	(SD)	True	ML	(SD)	PL	(SD)
$\kappa_1$	2	2.23	(0.31)	2.23	(0.33)	0.5	0.46	(0.23)	0.46	(0.23)
$\kappa_2$	3	2.81	(0.38)	2.81	(0.36)	0.75	0.92	(0.26)	0.91	(0.26)
$\kappa_3$	1	1.09	(0.22)	1.17	(0.22)	0.25	0.01	(0.25)	0.01	(0.26)
$\lambda_{12}$	2	1.63	(0.51)	1.39	(0.48)	2.0	1.84	(0.71)	1.73	(0.64)
$\lambda_{13}$	2	2.09	(0.43)	2.37	(0.38)	3.0	3.14	(0.67)	3.25	(0.64)
$\lambda_{23}$	2	2.30	(0.46)	2.58	(0.46)	4.0	3.75	(0.67)	3.76	(0.64)

Table 4.6: Mean values and approximate standard errors of pseudolikelihood and full likelihood estimates for trivariate von Mises data.

	True	ML	(SD)	PL	(SD)	True	ML	(SD)	PL	(SD)
$\kappa_1$	2	2.72	(0.80)	2.73	(0.81)	2.0	2.52	(0.31)	2.52	(0.31)
$\kappa_2$	2	2.03	(0.76)	2.03	(0.77)	2.0	2.32	(0.28)	2.32	(0.28)
$\kappa_3$	2	2.34	(0.91)	2.33	(0.93)	2.0	2.12	(0.26)	2.12	(0.26)
$\lambda_{12}$	20	12.45	(6.65)	12.69	(7.45)	0.1	0.16	(0.32)	0.15	(0.26)
$\lambda_{13}$	30	38.60	(7.42)	38.51	(7.42)	0.1	-0.12	(0.31)	-0.12	(0.26)
$\lambda_{23}$	40	38.43	(3.90)	38.27	(7.42)	0.1	0.42	(0.31)	0.44	(0.27)

Table 4.7: Mean values and approximate standard errors of pseudolikelihood and full likelihood estimates for trivariate von Mises data.

Comparison of the MLEs and MPLEs themselves reveals very little difference in terms

of accuracy. The approximate standard errors are also closely comparable, with the odd exception, for example for the estimates of  $\lambda_{23}$  for parameter configuration three. For the  $\lambda$  estimates in parameter configurations one, two and four, the given standard errors are in general slightly smaller for the pseudolikelihood estimates than for the full likelihood estimates. It should be noted however that these are approximate and the difference is small.

As an indication of the relative computational expense of the two methods, the estimation of parameters using the pseudolikelihood took less than 1 second for each of the configurations in Table 4.7, whilst the corresponding figures for the full likelihood were 38 minutes and 15 minutes. These figures clearly indicate the need for an alternative to the full likelihood in the current situation, whilst the figures in the Tables of the current chapter show the pseudolikelihood to be a good candidate for this alternative. It should be expected that the computational expense of MLEs relative to PLEs is even greater for higher dimensional data.

We now use the trivariate von Mises distribution in order to model protein data sets, using both the full likelihood and the pseudolikelihood.

## 4.4 Application to protein data

In this section we consider two different types of protein data and report the results of fitting trivariate von Mises distributions to each type. The first data set to be considered comprises the conformational angles of Gamma turns, whilst the second contains main chain and side chain conformational angles for the amino acids serine and valine. Modeling of the former was also considered by Hughes et al. (2006).

### 4.4.1 Gamma turn data

**Definition:** A gamma turn is a three-residue chain defined by the existence of a hydrogen bond between CO of residue  $i$  and NH of residue  $i + 2$ . In addition, the  $\phi$  and  $\psi$  angles of residue  $i + 1$  fall in the ranges  $\phi_{i+1} \in [35^\circ, 115^\circ] = [0.61, 2.00]$  radians and  $\psi_{i+1} \in [-104^\circ, -24^\circ] = [-1.82, -0.42]$  radians, respectively.

The data to be analysed in this section comprise the  $\phi$  and  $\psi$  triplets of 497 Gamma turns. Figures 4.4 and 4.5 display correlation plots of the data, with circular plots on the main diagonal, pairwise plots on the upper panels and circular correlation values on the lower panels. Circular correlations are calculated by replacing  $(x_i - \bar{x})$  and  $(y_i - \bar{y})$  in Pearson's product moment correlation for  $X$  and  $Y$  by  $\sin(x_i - \bar{x})$  and  $\sin(y_i - \bar{y})$ , where  $\bar{x}$  and  $\bar{y}$  in the latter two expressions are sample mean directions.  $p$ -values for testing the significance of the correlation coefficients are also given. For full details of this sample circular correlation coefficient, the reader is referred to Jammalamadaka and SenGupta (2001, chap. 8). Briefly, under the hypothesis that the true value of  $\rho_c$  is zero, the distribution of the estimate  $r_c$ , for large  $n$ , is such that

$$\sqrt{\left(n\hat{\lambda}_{20}\hat{\lambda}_{02}/\hat{\lambda}_{22}\right)} r_c \sim N(0, 1),$$

where

$$\hat{\lambda}_{ij} = \frac{1}{n} \sum_{k=1}^n \sin^i(\alpha_k - \bar{\alpha}) \sin^j(\beta_k - \bar{\beta})$$

for a random sample (of size  $n$ ) of circular variables  $(\alpha, \beta)$  with unspecified joint density. It is from this relation that the  $p$ -values are calculated.

As an exploratory analysis, a univariate von Mises distribution with mean direction  $\mu$  and concentration parameter  $\kappa$  is fitted separately to each of  $\phi_j$  and  $\psi_j$ ,  $j = 1, 2, 3$ . Maximum likelihood estimates of  $\mu$  and  $\kappa$  are displayed in Table 4.8.

We now use both the pseudolikelihood approach and the full likelihood approach to fit

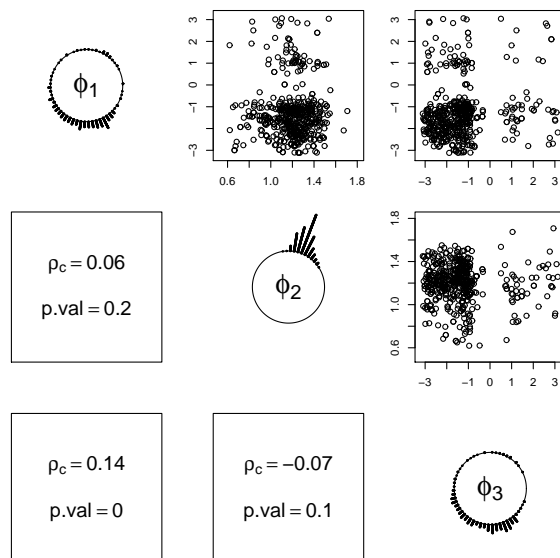


Figure 4.4: Matrix plot of  $\phi$  angles of gamma turn data, with circular plots on main diagonal, pairwise plots on upper panels and correlations on lower panels

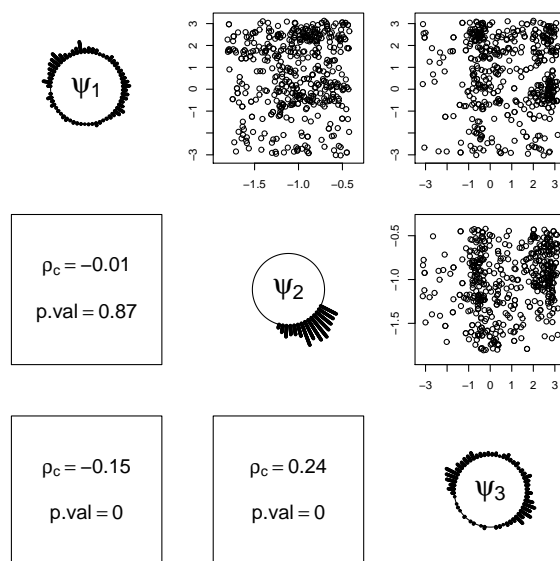


Figure 4.5: Matrix plot of  $\psi$  angles of gamma turn data, with circular plots on main diagonal, pairwise plots on upper panels and correlations on lower panels

a trivariate von Mises distribution to the  $\phi$  and  $\psi$  angles, separately, of the Gamma turn data, the results of which are shown in Table 4.9. Due to problems of convergence when the mean directions  $\mu_1$ ,  $\mu_2$  and  $\mu_3$  are included in the optimization process for the maximum likelihood approach applied to the  $\psi$  angles, we estimate these parameters using the marginal sample mean directions of the  $\psi$  data. The remaining parameters are estimated using the nlm approach (described in the previous section) applied to the mean (direction) corrected data. Standard errors of estimates are shown in brackets (except for the  $\psi$  mean directions using the MLE approach), and are calculated from the Hessian matrix obtained using the nlm estimation procedure, by taking the square roots of the diagonal elements of the inverse of the numerically calculated Hessian.

	$\phi_1$	$\phi_2$	$\phi_3$	$\psi_1$	$\psi_2$	$\psi_3$
$\hat{\mu}$	-1.64	1.20	-1.76	1.58	-1.02	1.03
$\hat{\kappa}$	1.58	31.46	1.67	0.44	8.19	0.31

Table 4.8: Marginal MLEs of  $\mu$  and  $\kappa$  for gamma turn data.

	<u>Gamma turn MLE</u>		<u>Gamma turn PLE</u>		<u>Simulated PLE</u>	
	$\phi$	$\psi$	$\phi$	$\psi$	$\phi$	$\psi$
$\kappa_1$	1.60 (0.09)	0.44 (0.07)	1.60 (0.09)	0.44 (0.07)	1.65 (0.10)	0.52 (0.07)
$\kappa_2$	31.72 (2.00)	8.87 (0.54)	31.73 (1.99)	8.87 (0.53)	33.25 (2.08)	7.75 (0.46)
$\kappa_3$	1.69 (0.10)	0.32 (0.67)	1.69 (0.10)	0.31 (0.07)	1.77 (0.10)	0.33 (0.07)
$\lambda_{12}$	0.55 (0.41)	0.23 (0.20)	0.65 (0.31)	0.15 (0.13)	1.04 (0.30)	0.06 (0.14)
$\lambda_{13}$	0.32 (0.12)	-0.43 (0.10)	0.39 (0.09)	-0.33 (0.06)	0.46 (0.08)	-0.48 (0.06)
$\lambda_{23}$	-0.71 (0.42)	1.35 (0.21)	-0.79 (0.31)	1.13 (0.13)	-0.66 (0.30)	0.94 (0.14)
$\mu_1$	-1.64 (0.05)	1.58 .	-1.63 (0.05)	1.46 (0.12)	-1.66 (0.05)	1.49 (0.09)
$\mu_2$	1.20 (0.01)	-1.02 .	1.20 (0.01)	-1.02 (0.02)	1.18 (0.01)	-1.02 (0.02)
$\mu_3$	-1.75 (0.04)	1.03 .	-1.73 (0.04)	1.27 (0.10)	-1.64 (0.04)	1.29 (0.09)

Table 4.9: Maximum likelihood estimates (MLEs) and pseudolikelihood estimates (PLEs) (and their standard errors) for  $\phi$  and  $\psi$  angles of Gamma turn data and for data simulated using these estimates.

Comparing Tables 4.8 and 4.9 we see that the univariate MLEs for  $\mu$  and  $\kappa$  are similar to



the estimates for  $\mu$  and  $\kappa$  in the trivariate case, with a possible tendency for the univariate  $\kappa$  estimates to be slightly smaller. Moreover, the  $\lambda$  estimates are generally quite close to zero, although most also have small standard errors relative to their magnitude.

As an assessment of the goodness-of-fit of the model, two sets of trivariate data are simulated using the MLEs and compared with the original  $\phi$  and  $\psi$  Gamma turn data. Figure 4.6 displays the data obtained.

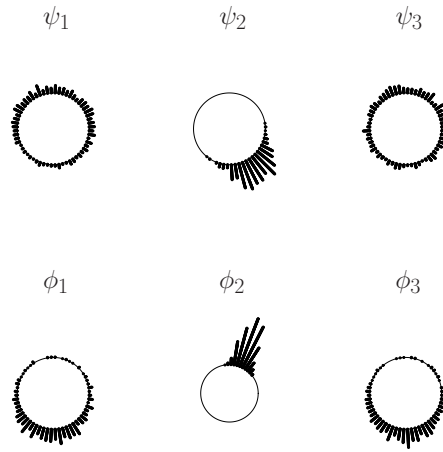


Figure 4.6: Circular plot of 497 simulated  $\phi$  and  $\psi$  triplets with true parameter values equal to the MLEs for the original gamma turn data.

It can be seen that the data simulated are reasonably similar to the original data. For the  $\psi$  angles there appears to be a slight bimodality in angles 1 and 3 of the original data that is not reproduced in the simulation. For both  $\phi$  and  $\psi$ , the plots for angles 1 and 3 of the simulated data appear slightly more gradual in terms of changes in density on the circle than the original data.

Using the pseudolikelihood to re-estimate the parameters based on the simulated data gives the results in the right hand part of Table 4.9, and we observe comparable results with those displayed in the left part of the table, indicating reasonable goodness-of-fit.

We extend the analysis of the Gamma turn data by developing likelihood ratio tests for the trivariate von Mises distribution.

### Hypothesis testing

An important part of fitting statistical models is the formulation and testing of hypotheses. In this section we compare likelihood ratio test statistics based on a von Mises maximum likelihood approach, a von Mises pseudolikelihood approach, and a normal approximation approach.

We will test two hypotheses regarding the mean directions of the angles composing Gamma turns. Firstly, the hypothesis  $\mu_1 = \mu_3$  for the  $\phi$  and  $\psi$  angles of such a turn will be tested. A further test on the mean directions of the  $\phi$  angles could be based on the hypothesis  $\mu_1 = \mu_3 = \mu_2 - \pi$ . Although these hypotheses may be criticised on the grounds that they have been generated by looking at the data, the focus of the present section is on formulating test procedures rather than the results of the tests per se. A test of independence of  $\phi_1$ ,  $\phi_2$  and  $\phi_3$  is equivalent to testing that all  $\lambda$  values are equal to zero. These three hypotheses will be tested in the three ways described above.

We outline the procedure for using a Normal approximation to the von Mises distribution in order to test the above hypotheses. Although  $\kappa$  values for angles 1 and 3 of both  $\phi$  and  $\psi$  are reasonably small, the trivariate Normal distribution

$$(\Theta_1, \Theta_2, \Theta_3) \sim N_3(\boldsymbol{\mu}_\Theta, \boldsymbol{\Sigma}_\Theta), \quad (4.14)$$

where  $(\boldsymbol{\Sigma}_\Theta^{-1})_{ii} = \kappa_i$ ,  $(\boldsymbol{\Sigma}_\Theta^{-1})_{ij} = -\lambda_{ij}$  and  $i \neq j$ , will be used to test the hypothesis  $\mu_1 = \mu_3$ . In (4.14),  $\Theta$  is to be replaced by  $\Phi$  or  $\Psi$ , depending on the variable of interest, whilst  $\boldsymbol{\mu}_\Theta$  is the vector comprising the mean directions of  $\Theta_1$ ,  $\Theta_2$  and  $\Theta_3$ . For  $\hat{\boldsymbol{\mu}}_\phi$  and  $\hat{\boldsymbol{\mu}}_\psi$  we take the maximum likelihood estimates. The covariance matrices obtained for  $\phi$  and

$\psi$ , based on the MLEs of the trivariate von Mises model, are respectively

$$\hat{\Sigma}_{\psi} = \begin{pmatrix} -6.28 & 1.43 & 12.06 \\ 1.43 & -0.11 & -1.96 \\ 12.06 & -1.96 & -16.98 \end{pmatrix} \quad \hat{\Sigma}_{\phi} = \begin{pmatrix} 0.67 & 0.01 & 0.15 \\ 0.01 & 0.03 & -0.01 \\ 0.15 & -0.01 & 0.63 \end{pmatrix}.$$

The matrix  $\hat{\Sigma}_{\psi}$  is not positive definite, a direct result of the values of the MLEs for the trivariate von Mises fit – since not all parameter values will lead to a positive definite normal approximation (in which case the normal approximation is inappropriate). Therefore the hypothesis described will be carried out for the  $\phi$  angles only. It is noted here that in Section 4.4.2 alternative, moment based estimators will be discussed, which would lead to a positive definite covariance matrix. Since these are based on a normal approximation, and the estimates of the  $\kappa$  parameters for the Gamma turn data are not particularly large (except for the middle value), discussion of these estimators is deferred until Section 4.4.2.

Test 1:  $\mu_1 = \mu_3$

The test statistic, which is approximately  $\chi_1^2$  distributed under the null hypothesis and for large  $\kappa$  values, is  $S = 2(l_{\text{full}} - l_{\text{red}})$ , where  $l_{\text{full}}$  is the loglikelihood of the distribution (4.14) with mean vector and covariance matrix  $\hat{\mu}_{\phi}$  and  $\hat{\Sigma}_{\phi}$  respectively.  $l_{\text{red}}$  is the corresponding value when the maximum likelihood estimates are calculated with  $\mu_1 = \mu_3$ . We get  $S = 2.91$ , with a  $p$ -value of 0.093. We therefore accept the null hypothesis that the means are equal.

Test 2:  $\mu_1 = \mu_3 = \mu_2 - \pi$

Using the same procedure as for test 1, we obtain a test statistic of 191.8 on two degrees of freedom. This is clearly a very large figure and we reject the null hypothesis. As seen in Table 4.9, the difference between the PLE of  $\mu_2$  and both  $\mu_1$  and  $\mu_3$  is slightly *less* than  $\pi$ , and  $\mu_2$  has a particularly small standard error.

Test 3:  $\lambda_{12} = \lambda_{13} = \lambda_{23} = 0$ 

Under this test of independence  $\hat{\Sigma}_{\phi}^{-1}$  for the reduced model is  $\text{diag}(\hat{\kappa})$ , where  $\hat{\kappa}$  is the MLE of  $(\kappa_1, \kappa_2, \kappa_3)$  when all  $\lambda$  values are equal to zero.  $\hat{\kappa}$  is calculated to be (1.59, 31.46, 1.67) (values that correspond almost exactly with those in Table 4.8), and the resulting test statistic is 119.8 on three degrees of freedom. Again this value is highly significant, although perhaps not surprisingly so, given the small magnitude of the standard errors of the  $\lambda$  estimates relative to the magnitude of the estimates themselves.

Of importance in fitting a normal distribution to directional data is the point at which the circle is cut. In the current situation we have done this at  $\pi$ , although this should perhaps be done separately for each variable in such a way that the concentration parameter is maximised. Of course, the more concentrated the data, the easier it is to select a point at which to cut the circle, and the more appropriate the use of the normal approximation. Since in the current setting  $\phi_1$  and  $\phi_3$  are quite dispersed, the normal approximation is perhaps not entirely suitable, and the problem is therefore not addressed in detail.

The left half of Table 4.10 gives loglikelihood values for the full and reduced models and for each of the three approaches described: full (joint) von Mises likelihood, von Mises pseudolikelihood and normal approximation. The right half of the table displays the likelihood ratio test statistics based on the values in the left part of the table. As can be seen from the table, the same conclusions are reached regarding the acceptance or rejection of the null hypothesis in each case, namely that the only null hypothesis not rejected is for test 1,  $\mu_1 = \mu_3$  (using a 5% significance level).

Of more interest, however, is comparison of loglikelihood values for the different methods, and the resulting differences in the test statistics. In particular, the log-pseudolikelihood appears to underestimate the loglikelihood slightly (in absolute value) in the first three rows of Table 4.10. In each case, however, the size of the underestimation

	Log-likelihood			Test statistic		
	VM joint	VM pseudo	Normal	VM joint	VM pseudo	Normal
Full	-1233.66	-1225.46	-1857.21			
Test 1	-1235.34	-1227.04	-1858.66	3.35	3.17	2.91
Test 2	-1259.57	-1250.30	-1953.10	51.81	49.70	191.79
Test 3	-1238.88	-1238.88	-1917.13	10.43	26.85	119.84

Table 4.10: Loglikelihood and log–pseudolikelihood values evaluated at MLEs and PLEs, and test statistics based on these values.

is similar, resulting in little difference in the test statistics for tests 1 and 2. In the case  $\lambda = 0$  (row 4 of the table), the full likelihood and pseudolikelihood values are the same. This is to be expected, since in this case each is the product of the same three independently distributed von Mises distributions. The effect is an inflated test statistic when the test is based on the pseudolikelihood.

#### 4.4.2 Serine and valine data

Of interest in the study of protein structures is the relationship between the backbone of a protein and its side–chain conformation (see, for example, Dunbrack and Cohen (1997)). The latter is measured by the angles  $\chi_j$ ,  $j = 1, \dots, 4$ , and each of the twenty commonly occurring amino acids has associated with it 0, 1, 2, 3 or all 4 of these  $\chi$  angles, as described in Section 1.2. In this section we use a trivariate von Mises distribution to model backbone and side–chain angles gathered from a database for two amino acids, serine and valine, that have a single side–chain angle  $\chi_1$ . The two amino acids are compared in terms of summary statistics and fitted model parameters, and for each amino acid the relationship between the  $\phi$ ,  $\psi$  and  $\chi_1$  angles is investigated using likelihood ratio tests.

### Summary statistics

The raw data comprise the  $\phi$ ,  $\psi$  and  $\chi_1$  (to be referred to as  $\chi$ ) angles for 10475 serine and 11617 valine amino acids, gathered from numerous proteins and obtained from the Backbone-Dependent Rotamer Library (<http://dunbrack.fccc.edu/bbdep/bbdepformat.php>).

Figures 4.7 and 4.8 show correlation plots, as described for the gamma turn data in the previous section, for the raw serine and valine data. For display purposes, the data are plotted in such a way as to be able to distinguish between clusters, so that axis endpoints are not necessarily  $\pi$  and  $-\pi$ . The figures show the data to be multimodal, whilst the  $p$ -values for the circular correlation coefficients show all pairwise correlations to be highly significant. The plots also reveal an apparent similarity between the valine and serine data. In this instance a mixture model appears appropriate, since it was conjectured in Section 2.3 that the marginal distributions associated with the trivariate von Mises distribution are either unimodal or bimodal.

As an exploratory analysis of the serine and valine data, we select regions of the plots in Figures 4.7 and 4.8 and propose a trivariate von Mises distribution for modelling the resulting data. Working in the interval  $[-\pi, \pi)$ , we take those data for which  $\phi \in (-\pi, 0)$ ,  $\psi \in \{(1.4, \pi] \cup (-\pi, -2\pi/3)\}$  and  $\chi \in (-1.75, 0)$ . This leaves data for 1171 serine and 1343 valine amino acids. Correlation plots of the resulting data are shown in Figures 4.9 and 4.10.

Based on the plots, the new data appear fairly unimodal. Again there are similarities between the valine and serine data, although the former appear to be slightly more concentrated. Based on the figures in the lower panels of the two plots, the only non-significant circular correlation for the serine data is that between  $\phi_s$  and  $\chi_s$ . For the valine data, that between  $\psi_v$  and  $\chi_v$  is the only non-significant correlation. Table 4.11 gives univariate maximum likelihood estimates for each angle of the reduced serine and valine

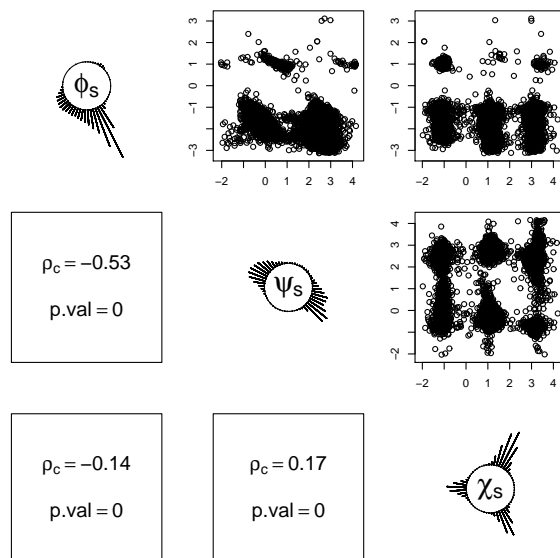


Figure 4.7: Correlation plot of raw serine data, with circular plots on main diagonal, pairwise plots on upper panels and correlations on lower panels.

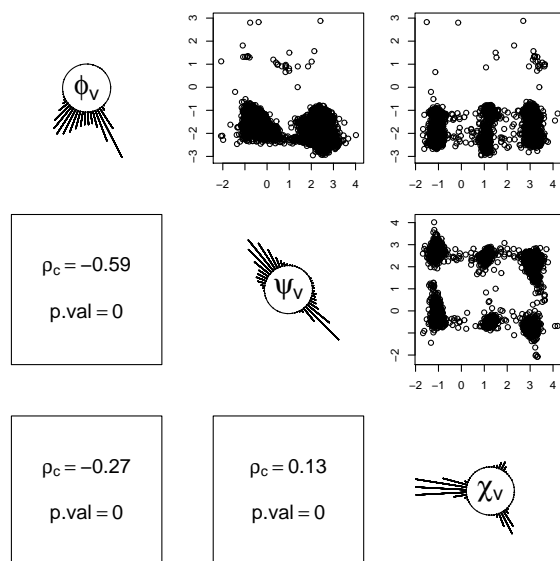


Figure 4.8: Correlation plot of raw valine data, with circular plots on main diagonal, pairwise plots on upper panels and correlations on lower panels.

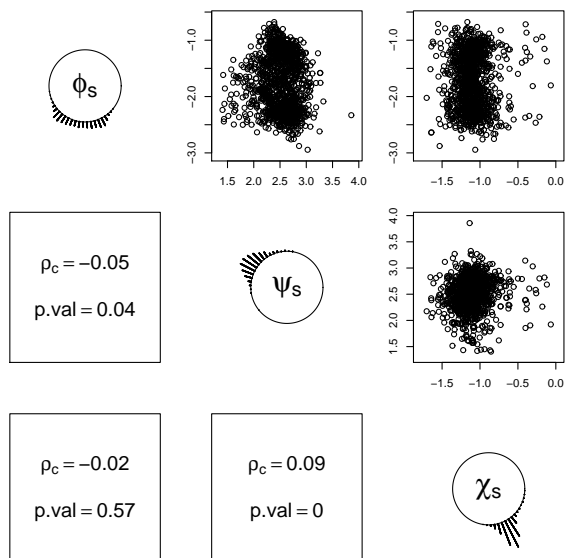


Figure 4.9: Correlation plot of selected serine data, with circular plots on main diagonal, pairwise plots on upper panels and correlations on lower panels.

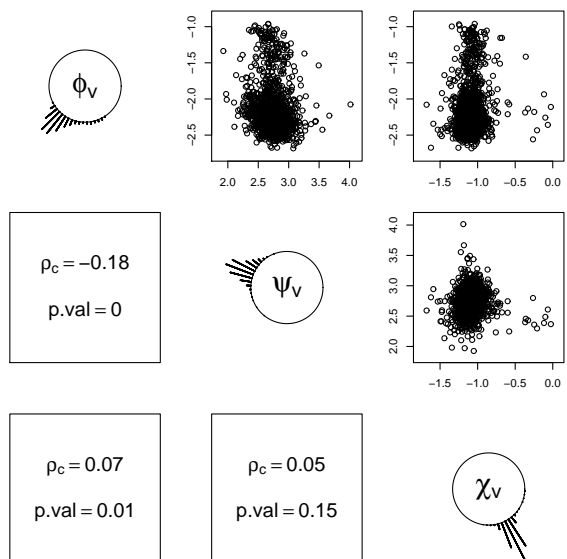


Figure 4.10: Correlation plot of selected valine data, with circular plots on main diagonal, pairwise plots on upper panels and correlations on lower panels.



data.

Immediately it can be seen from the MLEs, as it can from the plots, that the data are highly concentrated, particularly the valine data. The suitability of a trivariate normal distribution should therefore also be considered. The sample mean directions for the two data sets are reasonably similar, particularly for the  $\chi$  values.

	$\phi_s$	$\psi_s$	$\chi_s$	$\phi_v$	$\psi_v$	$\chi_v$
$\hat{\mu}$	-1.73	2.51	-1.10	-2.16	2.75	-1.08
$\hat{\kappa}$	4.56	11.55	28.29	9.32	25.46	52.83

Table 4.11: Univariate von Mises MLEs of  $\mu$  and  $\kappa$  for serine and valine data

### Fitting a trivariate von Mises distribution

We now fit the proposed trivariate von Mises distribution to the serine and valine data. The results are displayed in Table 4.12. The maximum likelihood estimates are obtained as described for the gamma turn data analysed in Section 4.4.1, and the pseudolikelihood estimates are used as the starting point for the algorithm. For the maximum likelihood approach, the same convergence issues arise as for the  $\psi$  angles of the gamma turn data. For this reason, we again estimate the mean directions in the MLE approach by taking the marginal sample mean directions. The mean (direction) corrected data are then used in estimation of the remaining parameters.

We observe from Table 4.12 that the MLEs and PLEs are very similar, particularly for the serine data. This should be expected, since the data are highly concentrated and we know that, as the concentration parameters tend to infinity, the trivariate von Mises distribution tends to a trivariate normal distribution, for which the pseudolikelihood estimators are fully efficient. There is some disagreement in the magnitude of the standard errors of the estimates based on the full likelihood and pseudolikelihood, and since these are only

	<u>MLE</u>				<u>PLE</u>				<u>PLE (simulated)</u>			
	Serine		Valine		Serine		Valine		Serine		Valine	
$\kappa_\phi$	4.59	(0.17)	9.61	(0.36)	4.58	(0.17)	9.63	(0.36)	4.80	(0.18)	9.84	(0.36)
$\kappa_\psi$	11.66	(0.47)	26.26	(1.00)	11.66	(0.47)	26.28	(0.99)	12.33	(0.50)	28.09	(1.06)
$\kappa_\chi$	28.49	(1.17)	53.25	(2.05)	28.50	(1.16)	53.28	(2.04)	28.52	(1.16)	54.55	(2.09)
$\lambda_{\phi\psi}$	-0.43	(0.23)	-2.85	(0.46)	-0.42	(0.16)	-2.99	(0.33)	-0.28	(0.17)	-2.98	(0.34)
$\lambda_{\phi\chi}$	-0.17	(0.36)	1.81	(0.64)	-0.16	(0.25)	1.94	(0.47)	-0.13	(0.26)	1.92	(0.46)
$\lambda_{\psi\chi}$	1.59	(0.55)	2.22	(1.04)	1.64	(0.40)	2.34	(0.75)	1.42	(0.40)	2.24	(0.76)
$\mu_\phi$	-1.73	.	-2.16	.	-1.73	(0.01)	-2.16	(0.01)	-1.73	(0.01)	-2.17	(0.01)
$\mu_\psi$	2.51	.	2.75	.	2.51	(0.01)	2.74	(0.01)	2.50	(0.01)	2.74	(0.01)
$\mu_\chi$	-1.10	.	-1.08	.	-1.10	(0.01)	-1.09	(0.00)	-1.09	(0.01)	-1.08	(0.00)

Table 4.12: Maximum likelihood estimates (MLEs) and pseudolikelihood estimates (PLEs) (and their standard errors) of parameters for serine and valine data, and for data simulated using these estimates.

approximate, the significance of parameters would be better explored with likelihood ratio tests. As with the gamma turn data, NA's in the table are the result of negative values on the diagonal of the inverse of the numerically obtained Hessian matrix.

Comparing the ability of the maximum likelihood and pseudolikelihood procedures to maximise the full likelihood function, the loglikelihood evaluated at the MLEs is  $-816.4236$  and evaluated at the PLEs is  $-816.4254$  for the serine data. The corresponding figures for the valine data are  $584.7315$  and  $584.6620$ . The maximum likelihood approach is therefore seen to be superior (as expected) in this respect, although the difference is practically zero. On the other hand, estimation of parameters takes over two hours for each data set based on the ML approach, whilst the PLEs are calculated in approximately ten seconds, giving an idea of the relative computational expense of the two approaches.

Comparing Tables 4.11 and 4.12, we see that the  $\kappa$  estimates based on independent univariate von Mises distributions are very similar to those based on the trivariate models, with those of the latter being very slightly larger in each case. We note that the same was true for the gamma turn data, and in both cases the estimated  $\lambda$  values are quite small.

The closeness of the agreement between the univariate and trivariate  $\kappa$  estimates could be an indication of the dependence of the variables in the trivariate model, with ‘closer’ agreement indicating less dependence, although this is not a very rigorous assessment. Based on the  $\lambda$  estimates and their standard errors, using either the MLEs or PLEs, the indication for the serine data is again that the only non-significant correlation is that between  $\phi_s$  and  $\chi_s$ . For the valine data, all three appear significant.

As an assessment of the goodness-of-fit of the model, we follow the same approach as for the gamma turn data and simulate data based on the pseudolikelihood estimates, then compare the resulting data with the original data in terms of plots and parameter estimates. The pseudolikelihood estimates of the simulated data are displayed in Table 4.12, whilst correlation plots of the simulated data are given in Figures 4.11 and 4.12.

We see from Table 4.12 that the PLEs for the simulated data are similar to the true parameters (the PLEs of the original data), but with a tendency for the  $\kappa$  parameters to be overestimated. Considering Figures 4.11 and 4.12, the circular plots on the diagonals appear similar for the original and simulated data, although slightly less so for  $\phi_s^{\text{sim}}$  than for the other variables. The plots on the upper panels appear slightly different for the original and simulated data, although this is partly due to the slightly different axis scales. Specifically, the simulated data are not subject to the range restrictions imposed on the original data. In terms of the circular correlation coefficients and their significance, the former are very similar for the original and simulated data, although the correlation between  $\phi_s^{\text{sim}}$  and  $\psi_s^{\text{sim}}$  is noticeably less significant than that between  $\phi_s$  and  $\psi_s$  based on the  $p$ -values.

In summary, the model appears to be a reasonable fit for both the serine and valine data based on the comparison between the original data and those simulated using PLEs as true parameter values. Indeed, some of the discrepancies observed may be due to the problem of simulating data without the restriction in range imposed on the original data.

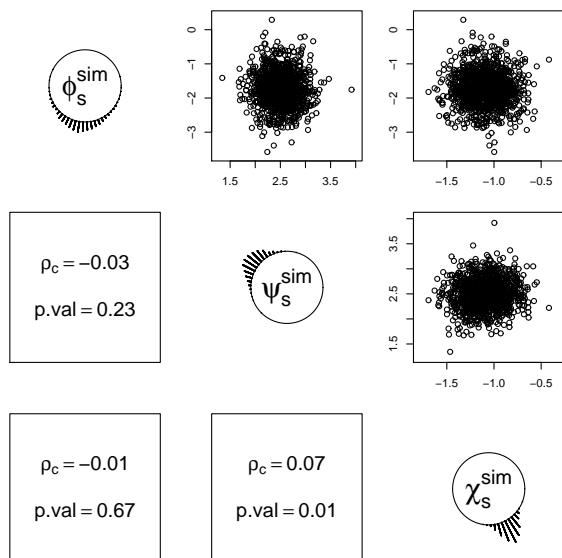


Figure 4.11: Correlation plot of simulated serine data, with circular plots on main diagonal, pairwise plots on upper panels and correlations on lower panels.

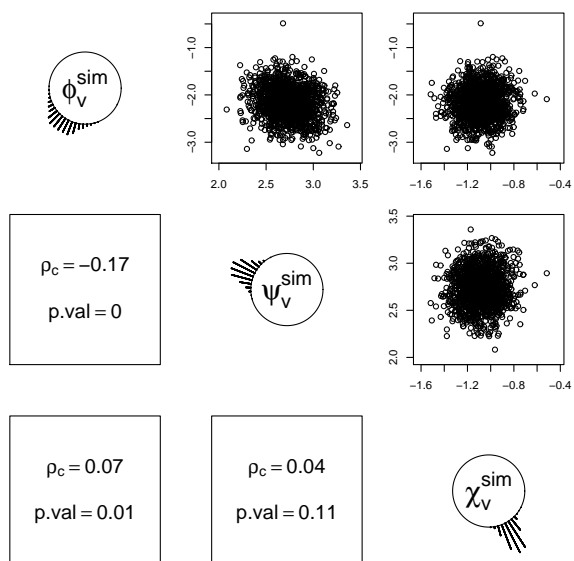


Figure 4.12: Correlation plot of simulated valine data, with circular plots on main diagonal, pairwise plots on upper panels and correlations on lower panels.

Since we have seen that the estimates of the  $\kappa$  parameters for both the serine and valine data are large, we here describe an alternative method of parameter estimation, based on moment estimators and an assumption of high concentration. In particular, we estimate the covariance matrix of the normal distribution under high concentration (independent of any particular circular model) using (Singh et al., 2007)

$$\hat{\Sigma} = \begin{pmatrix} 2(1 - \bar{R}_1) & \bar{S}_{12} & \bar{S}_{13} \\ & 2(1 - \bar{R}_2) & \bar{S}_{23} \\ & & 2(1 - \bar{R}_3) \end{pmatrix} \quad (4.15)$$

where

$$\bar{R}_j = \frac{\sum_{i=1}^n \cos(\theta_{ji} - \bar{\theta}_j)}{n}, \quad \bar{S}_{jl} = \frac{\sum_{i=1}^n \sin(\theta_{ji} - \bar{\theta}_j) \sin(\theta_{li} - \bar{\theta}_l)}{n},$$

$j \neq l$  and  $\bar{\theta}_j$  is the sample mean direction of the  $\theta_{ji}$  values,  $i = 1, \dots, n$ . We note that an alternative would be to use  $\bar{S}_{jl}$  for  $j = l$  (ie. the diagonal elements of the matrix). The inverse of this matrix can then be compared with the covariance matrix of the limiting normal distribution of the trivariate von Mises distribution, which was derived in Section 2.3 to be of the form  $(\Sigma^{-1})_{jj} = \kappa_j$ ,  $(\Sigma^{-1})_{jl} = -\lambda_{jl}$ ,  $j \neq l$ . Estimates of the  $\kappa$  and  $\lambda$  parameters based on the matrix (4.15) can then be obtained and compared with the maximum likelihood estimates and pseudolikelihood estimates from Table 4.12. Table 4.13 shows moment estimates of the  $\kappa$  and  $\lambda$  parameters for the serine and valine data. The estimates are obtained by relating the inverse of the matrix

	$\kappa_1$	$\kappa_2$	$\kappa_3$	$\lambda_{\phi\psi}$	$\lambda_{\phi\chi}$	$\lambda_{\psi\chi}$
Serine	4.27	11.39	28.22	-0.33	-0.14	1.44
Valine	9.33	25.94	52.96	-2.53	1.62	2.08

Table 4.13: Estimates of  $\kappa$  and  $\lambda$  parameters for serine and valine data based on moment estimator of covariance matrix of the normal distribution.

in (4.15) to the inverse covariance matrix under high concentration of the trivariate von

Mises distribution. Specifically, the diagonal elements of the former are the estimates of the  $\kappa$  parameters, whilst the remaining elements are negated in order to obtain the  $\lambda$  parameter estimates. Comparing Tables 4.12 and 4.13, we see that the moment estimates are very similar to the estimates obtained using a maximum likelihood approach and a pseudolikelihood approach. The absolute value of each parameter estimate is slightly smaller for the moment estimates. Since the moment estimators are constructed assuming high concentration and using a normal approximation approach, the similarity of the moment estimates and the likelihood estimates indicates that a normal distribution may be reasonable for modeling the serine and valine data.

In the next section we investigate hypothesis testing for the serine and valine amino acid data.

### Hypothesis testing

We now use those methods employed for the gamma turn data in order to test hypotheses on the  $\lambda$  values of the serine and valine data. Since it is of interest to investigate the interrelationships between the  $\phi$ ,  $\psi$  and  $\chi$  values, we first test the hypothesis that all  $\lambda$  values for a particular data set are equal to zero ( $\lambda = 0$ ). In this case the pseudolikelihood and full likelihood are identical, the joint density of  $\phi$ ,  $\psi$  and  $\chi$  being the product of three independent von Mises distributions. The MLEs and PLEs for this model are therefore simply the univariate maximum likelihood estimates based on each column of the data.

For serine, the subsequent test  $\lambda_{\phi_s \psi_s} = \lambda_{\phi_s \chi_s} = 0$  is also performed. In this case,  $\phi_s$  follows a univariate von Mises distribution whilst  $\psi_s$  and  $\chi_s$  have a joint bivariate von Mises distribution. Comparison of a full likelihood and a pseudolikelihood approach in this case therefore reduces essentially to a comparison of the likelihood and pseudolikelihood for a bivariate von Mises distribution.

The results of testing the hypotheses described are displayed in Table 4.14. The following

four methods are used to test the hypotheses:

1. Full and reduced models are fitted using the joint trivariate von Mises distribution and likelihood ratio tests are performed using values of the loglikelihood function at the MLEs.
2. The pseudolikelihood is used for fitting the full and reduced models, and likelihood ratio tests are performed using values of the log-pseudolikelihood at the PLEs.
3. A trivariate normal distribution is fitted to the data for the full and reduced models and likelihood ratio tests are performed.
4. A trivariate normal approximation based on the MLEs is used to obtain loglikelihood values based on a normal distribution and likelihood ratio tests are performed using these values.

In addition to the conclusions of the results of the tests, also of interest is comparison of the four approaches. They provide a means of comparing the pseudolikelihood and full likelihood for hypothesis testing, and also for assessing the appropriateness of a normal distribution by comparing the results of methods 3 and 4 above.

	Log-likelihood				Test statistic			
	VM joint	VM pseudo	Normal	Norm appr.	VM joint	VM pseudo	Normal	Norm appr.
<b>Serine</b>								
Full	-816.42	-810.25	-813.02	-817.94				
$\lambda = 0$	-822.67	-822.67	-818.35	-823.20	12.49	24.83	10.66	10.53
$\lambda_{\phi\psi} = \lambda_{\phi\chi} = 0$	-818.40	-813.86	-814.66	-819.60	3.96	7.21	3.29	3.33
<b>Valine</b>								
Full	584.73	613.01	551.23	547.73				
$\lambda = 0$	559.87	559.87	526.77	523.24	49.72	106.27	48.90	48.98

Table 4.14: Loglikelihood values and test statistics based on four different approaches.

A number of conclusions and observations can be made based on the results of the tests. For both serine and valine, the hypothesis that all  $\lambda$  values are equal to zero is

strongly rejected by all the tests, based on a chi-squared distribution with 3 degrees of freedom, the 2% critical value of which is 9.84. Again for both data sets, the maximised von Mises log-pseudolikelihood is greater than the maximised von Mises log-likelihood when based on the full model. When all  $\lambda$  parameters are equal to zero, however, the pseudolikelihood and likelihood are, as mentioned above, identical. We therefore obtain a greatly inflated test statistic when this test is performed using the pseudolikelihood. Comparison of both loglikelihood and test statistic values for the von Mises and normal distribution approaches reveals close similarities between the two, particularly for serine. The closeness of the agreement could be an indication of the appropriateness of a normal distribution for modelling the data. When the normal distribution model is fitted and evaluated it is important to cut the circle in such a way as to obtain a unimodal distribution. In the current context, this is achieved by adding  $2\pi$  to all negative  $\psi$  values, as for the plots in Figures 4.9 and 4.10.

For the test  $\lambda_{\phi\psi} = \lambda_{\phi\chi} = 0$  for serine we obtain non-significant test statistics for all tests except that based on the pseudolikelihood. The test statistic 7.21 for the pseudolikelihood approach has a  $p$ -value of 0.027 based on a chi-squared distribution with two degrees of freedom, whilst that based on the likelihood approach, 3.96, has a  $p$ -value of 0.13. As discussed above, in fitting this reduced model we are able to compare the full likelihood and pseudolikelihood for the *bivariate* von Mises distribution. As for the trivariate case, the latter, at least for these data, overestimates the former. Since the overestimation is not as great (in absolute terms) as for the trivariate model, we again obtain an inflated test statistic relative to the other approaches. We conclude that there is not significant evidence to reject the hypothesis that both  $\lambda_{\phi\psi}$  and  $\lambda_{\phi\chi}$  are equal to zero.

We observe from Table 4.14 that the test statistics based on the von Mises pseudolikelihood approach are approximately twice those based on the von Mises full likelihood approach. A similar observation can be made for the hypothesis that all  $\lambda$  parameters are equal to zero for the Gamma turn data (Table 4.10). If we consider normal



variates  $x$  and  $y$  in the context of a regression problem, then the hypothesis  $\rho_{xy} = 0$  is equivalent to each of the hypotheses  $\beta_x = 0$  in the regression of  $y$  on  $x$  and  $\beta_y = 0$  in the regression of  $x$  on  $y$ . In employing the full likelihood approach we are effectively testing either  $\beta_x = 0$  or  $\beta_y = 0$ , whilst a pseudolikelihood approach uses both. The likelihood ratio test statistic based on the latter should therefore be twice that based on the former. We conjecture that a similar statement is true in higher dimensions for both the partial pseudolikelihood and the full pseudolikelihood, although this is not discussed further here.

We conclude the section by noting that it may also be instructive to compare data sets in terms of their parameter estimates by using likelihood ratio tests. For example, we could test the hypothesis that the value of each parameter for serine is equal to the corresponding parameter value for valine. This would be done by fitting a trivariate von Mises model to the pooled data, and comparing the loglikelihood with the sum of the loglikelihoods of the two separate models. In the present situation the parameter estimates are very different (especially for the  $\kappa$  parameters) and so this hypothesis will not be explored further.

## 4.5 Conclusions

In this chapter we have derived the efficiency of the pseudolikelihood for the bivariate von Mises distribution. This efficiency has been shown to depend on the values of parameters, and is reasonably high so long as  $\lambda$  is not very large. We have performed a thorough analysis of properties of pseudolikelihood estimates for the trivariate von Mises distribution, implementing a Gibbs sampling approach to data simulation and comparing pseudolikelihood estimates with maximum likelihood estimates. For all parameter combinations considered, the two are shown to have similar properties in terms of accuracy and precision of estimates, whilst the former are calculated at a fraction of the computational cost. An analysis of protein fold data shows the trivariate model to

be a reasonable fit for those data studied, and likelihood ratio tests have been developed for the trivariate distribution. A further analysis has investigated the relationship between the backbone and side-chain conformation of two amino acids, serine and valine. Based on the region of data analysed, it was concluded that the  $\phi$  angles for the serine data are independent of both the  $\psi$  and  $\chi$  values. The results concerning efficiency for the bivariate case and regarding efficiency and model fitting for the trivariate case are summarised in Mardia et al. (2007a) and Singh et al. (2007) respectively.

Using an extension of the bivariate Cosine model, discussed in detail by Mardia et al. (2007b) and defined in Equation (1.13), an alternative multivariate von Mises model, the multivariate Cosine model, can be defined for  $\Theta = (\Theta_1, \Theta_2, \dots, \Theta_p)$  as follows:

$$f(\Theta) = C_p^{-1}(\kappa, \Delta) \exp \left\{ \kappa^T c(\theta, \mu) - s(\theta, \mu)^T \Delta s(\theta, \mu) - c(\theta, \mu)^T \Delta c(\theta, \mu) \right\},$$

where  $-\pi < \theta_j \leq \pi$ ,  $-\pi < \mu_j \leq \pi$ ,  $\kappa_j \geq 0$ ,  $\delta_{jl} \geq 0$ . The vectors  $c(\theta, \mu)$ ,  $s(\theta, \mu)$ ,  $\mu$  and  $\kappa$  are defined as for the multivariate Sine model (Equation (2.2)), while  $[\Delta]_{jl} = \delta_{jl} = \delta_{lj}$  and  $\delta_{jj} = 0$ . The normalizing constant is  $C_p^{-1}(\kappa, \Delta)$ . This model can be investigated in ways similar to the analysis of Chapters 2 to 4 for the multivariate Sine model. Comparisons analogous to those made by Mardia et al. (2007b) for the bivariate Sine and Cosine models can then be made for the multivariate models. The two can also be compared, for example, in terms of the efficiency of the pseudolikelihood for parameter estimation.

## **Part II**

### **Time series models with application to protein conformational angles**



## Chapter 5

# Directional time series models

### 5.1 Introduction

Directional time series occur in a number of meteorological, oceanographic, geological and biological contexts. The first of these has received perhaps the most attention in the literature, Breckling (1989) providing the most comprehensive account with his analysis of wind speeds and directions, whilst Fisher and Lee (1994) and Coles (1998) also use wind directions in their examples. On the other hand, biological applications, for example, are seldom found, and with the emergence of bioinformatics the potential for exploiting existing models and developing new ones provides the opportunity for advancement and increased usage of the subject of directional time series.

In contrast to the wealth of tools at the disposal of those wishing to analyse linear time series, various challenges associated with the analysis of directional data have hindered the development of analogous techniques for time series comprising directional measurements. In this section we survey the existing models and highlight some of the problems associated with the further development of such methods.

Attention focuses on exploratory analysis (in Section 5.2) and on four directional time

series models: the von Mises autoregressive (AR) process (Section 5.3); processes based on link functions (Section 5.4), the wrapped AR process (Section 5.5) and projected processes (Section 5.6). Criteria for choosing between the models are discussed in Section 5.7, whilst Section 5.8 summarises the chapter.

## 5.2 Directional time series and exploratory analysis

In many statistical analyses, the first step is to use exploratory or descriptive techniques to investigate features of the data being studied. For linear time series, examples include the autocorrelation coefficient and a simple time plot to identify potential trend, non-stationarity, seasonal effects, etc (see, for example, Chatfield (1975) for comprehensive details).

Fisher (1993) describes some analogous methods for the exploratory analysis of directional time series. For highlighting trends in a directional time series  $\{\theta_i\}$ ,  $i = 1, \dots, n$ , he describes kernel estimates  $\hat{x}(t)$  and  $\hat{y}(t)$  based on  $x_i = \cos \theta_i$  and  $y_i = \sin \theta_i$ ,  $i = 1, \dots, n$ . In practice, the  $x_i$  and  $y_i$  values are smoothed individually, each being a weighted average of values measured at times  $t_i$  near  $t$ . In particular, the following expression gives  $\hat{x}(t)$  whilst an analogous expression can be used to obtain  $\hat{y}(t)$ :

$$\hat{x}(t) = \frac{\sum_{i=1}^n K_h(t - t_i) x_i}{\sum_{i=1}^n K_h(t - t_i)}$$

where  $K_h(t) = h^{-1}K(t/h)$  and  $K$  is the desired kernel function.  $\hat{x}(t)$  and  $\hat{y}(t)$  are used to obtain a smoothed estimate  $\hat{\theta}(t)$  of the form

$$\hat{\theta}_t = \tan^{-1} \{ \hat{y}(t) / \hat{x}(t) \}.$$

Analogous to the autocorrelation coefficient for linear variables, he describes a  $k$ -lag circular autocorrelation coefficient of  $\{\theta_i\}$  based on the  $(n - k)$  data pairs

$$(\theta_1, \theta_{k+1}), (\theta_2, \theta_{k+2}), \dots, (\theta_{n-k}, \theta_n).$$

His circular autocorrelation coefficient is based on an estimator of the circular correlation

$$\rho_T = \frac{E [\sin(\Theta_1 - \Theta_2) \sin(\Phi_1 - \Phi_2)]}{\{E [\sin^2(\Theta_1 - \Theta_2)] E [\sin^2(\Phi_1 - \Phi_2)]\}^{\frac{1}{2}}} \quad (5.1)$$

between two circular variables  $\Theta$  and  $\Phi$  introduced by Fisher and Lee (1983). Here,  $(\Theta_1, \Phi_1)$  and  $(\Theta_2, \Phi_2)$  are distributed independently as  $(\Theta, \Phi)$ . The estimator of  $\rho_T$  used is

$$\hat{\rho}_T = \frac{\sum_{1 \leq i < j \leq n} \sin(\theta_i - \theta_j) \sin(\phi_i - \phi_j)}{\left[ \sum_{1 \leq i < j \leq n} \sin^2(\theta_i - \theta_j) \sum_{1 \leq i < j \leq n} \sin^2(\phi_i - \phi_j) \right]^{\frac{1}{2}}}, \quad (5.2)$$

and forms the basis for calculation of the circular autocorrelation coefficient. For convenience, if we write  $\phi_i = \theta_{i+k}$ ,  $i = 1, \dots, n - k$ , then we have the  $(n - k)$  pairs

$$(\theta_1, \phi_1), (\theta_2, \phi_2), \dots, (\theta_{n-k}, \phi_{n-k}),$$

based on which we can calculate the  $k$ -lag circular autocorrelation  $\hat{\rho}_T(k)$  as

$$\hat{\rho}_T(k) = \frac{\sum_{1 \leq i < j \leq n-k} \sin(\theta_i - \theta_j) \sin(\phi_i - \phi_j)}{\left[ \sum_{1 \leq i < j \leq n-k} \sin^2(\theta_i - \theta_j) \sum_{1 \leq i < j \leq n-k} \sin^2(\phi_i - \phi_j) \right]^{\frac{1}{2}}}. \quad (5.3)$$

Plotting  $\hat{\rho}_T(k)$  versus  $k$  then gives a circular correlogram. In order to assess the significance of these coefficients, permutation tests can be carried out, in which  $k$ -lag circular autocorrelation coefficients are calculated for a large number (or preferably all) of the  $n! - 1$  possible orderings of the data other than that observed, and compared with that value calculated based on the actual data.

Fisher (1993) also gives a detailed account of rank cumulative sum (CUSUM) methods due to Lombard (1988), based on which one can investigate the possibility that a directional time series comprises blocks of data differing in their population mean directions.

### 5.3 The von Mises autoregressive process

Breckling (1989) develops a directional process, the von Mises autoregressive (AR)

process, by direct analogy with the linear AR process. Specifically, for a directional process  $\{\theta_t\}$ , if the conditional distribution of  $\theta_t$  given  $(\theta_{t-1}, \dots, \theta_{t-p})$  is von Mises with concentration vector (see Definition 1.3.2)

$$\mathbf{v}_t = \sum_{j=1}^p \kappa_j (\cos \theta_{t-j}, \sin \theta_{t-j})^T + (\kappa_0, 0)^T$$

then  $\{\theta_t\}$  is a von Mises process and

$$f(\theta_t | \theta_{t-1}, \dots, \theta_{t-p}) = [2\pi I_0(v_t)]^{-1} \exp \left\{ \sum_{j=1}^p \kappa_j \cos(\theta_t - \theta_{t-j}) + \kappa_0 \cos \theta_t \right\}$$

where  $v_t = \|\mathbf{v}_t\|$  is the length of  $\mathbf{v}_t$ .

An important distinction between this directional AR process and its linear counterpart is that the former does *not* have constant concentration whilst the latter does have constant variance. For the special case  $p = 1$  and  $\kappa_0 = 0$ ,  $\{\theta_t\}$  becomes a process with constant concentration and  $\theta_t$  has a uniform marginal distribution with density  $(2\pi)^{-1}$ .

Maximum likelihood estimators and least squares estimators for  $\boldsymbol{\kappa} = (\kappa_0, \dots, \kappa_p)^T$  are shown to agree, the estimating equation for  $\boldsymbol{\kappa}$  in each case being

$$\frac{d \log L_c}{d\boldsymbol{\kappa}} = \sum_{t=p}^n \left\{ -\frac{A_1(v_t)}{v_t} \Gamma_t \Gamma_t^T \boldsymbol{\kappa} + \gamma_t \right\} = \mathbf{0},$$

where  $L_c$  is the likelihood function conditional on  $\theta_0$ ,  $A_\nu(\cdot) = I_\nu/I_0(\cdot)$ ,

$$\Gamma_t = \begin{pmatrix} 1 & 0 \\ \cos \theta_{t-1} & \sin \theta_{t-1} \\ \vdots & \vdots \\ \cos \theta_{t-p} & \sin \theta_{t-p} \end{pmatrix} \quad \text{and} \quad \gamma_t = \begin{pmatrix} \cos \theta_t \\ \cos(\theta_t - \theta_{t-1}) \\ \vdots \\ \cos(\theta_t - \theta_{t-p}) \end{pmatrix}.$$

In general, an iterative technique is required to obtain  $\hat{\boldsymbol{\kappa}}$ , the exception being for  $p = 1$  and  $\kappa_0 = 0$ . Breckling (1989) applies the model with  $p = 1$  to a series of wind directions.



## 5.4 Linked processes

Fisher and Lee (1994) use the idea of link functions to adapt linear time series models to a directional context using two approaches.

Firstly, they define a circular stationary process  $\{\theta_t\}$  with mean direction  $\mu$  to be a linked autoregressive moving average (LARMA) process if and only if the process  $\{x_t\} = \{g^{-1}(\theta_t - \mu)\}$  is an autoregressive moving average (ARMA) process, where  $g(x)$  is an odd, monotone function mapping the real line onto  $(-\pi, \pi]$  and  $g(0) = 0$ . The  $k$ -lag circular autocorrelation function of  $\{\theta_t\}$  is given as

$$\rho_T(k) = \rho_T \{g(X_t), g(X_{t+k})\},$$

where  $\rho_T$  is as in Equation (5.1). They use this to plot acf's for LAR(1) models based on the probit link, where the directional series  $\{\theta_t\}$  is obtained from the simulated linear series  $\{x_t\}$ . Given a series  $\{\theta_t\}$  it is suggested that standard methods for linear series be used to model the series  $x_t = g^{-1}(\theta_t - \hat{\mu})$ , where  $\hat{\mu}$  is the sample mean direction of the series. The suggested approach is therefore that  $g(\cdot)$  is chosen so as to make the associated linear process approximately normal.

An alternative specification of a circular autoregressive (CAR) process using a link function is also described by Fisher and Lee (1994), in which  $\Theta_t$  given  $\theta_{t-1}, \dots, \theta_{t-p}$  follows a von Mises distribution with concentration parameter  $\kappa$  and mean direction

$$\mu_t = \mu + g \{ \omega_1 g^{-1}(\theta_{t-1} - \mu) + \dots + \omega_p g^{-1}(\theta_{t-p} - \mu) \}. \quad (5.4)$$

In contrast to Breckling's use of the conditional likelihood for parameter estimation, Fisher and Lee (1994) assume a univariate von Mises distribution  $M(\mu, \kappa)$  for each  $\theta_t$  (but observe that these are not the true marginals). It is claimed estimation is not adversely effected by making this assumption for series that are not too short. The likelihood function is then

$$\prod_{t=p+1}^n f(\theta_t - \mu - g[\omega_1 g^{-1}\{(\theta_{t-1} - \mu)/2\} \cdots + \omega_p g^{-1}\{(\theta_{t-p} - \mu)/2\}]) \prod_{t=1}^p f(\theta_t - \mu)$$

where  $f$  is the density of the von Mises distribution with zero mean and concentration parameter  $\kappa$ . Again an iterative procedure is needed for parameter estimation. The two linked processes are used to fit directional AR(1) models to a time series of wind directions and  $g(\cdot)$  is taken to be the probit link.

## 5.5 Wrapped processes

The directional time series model that has received the most attention in the literature is the wrapped autoregressive (WAR) process, developed by Breckling (1989). There are various reasons for this. It provides a natural circular analogue to linear processes and interpretation is aided by the relationship with the normal distribution on the line. The covariance matrix, for example, maintains many of its properties under the wrapping procedure (Coles, 1998). On the other hand, inference for the model is complicated, and various methods have therefore been proposed with the aim of overcoming this difficulty in order that the WAR process can realise its potential.

The wrapped AR process is most simply viewed as the result of wrapping a linear AR process  $\{X_t\}$  around the circle. Given the linear AR( $p$ ) process

$$X_t = \sum_{j=1}^p \alpha_j X_{t-j} + \epsilon_t; \quad \epsilon_t \sim N(0, \sigma^2) \text{ i.i.d.},$$

the WAR( $p$ ) process  $\{\eta_t\}$  is defined by

$$\eta_t = X_t \pmod{2\pi}, \quad t = 1, \dots, n.$$

Thus, if we observe the series  $\{\eta_t\}$ , the linear process is given by  $X_t = \eta_t + 2\pi k_t$ , where the  $k_t$  are unobserved integers, called wrapping coefficients. The inferential complexity arises from the fact that that we need to estimate the parameters  $\alpha_1, \dots, \alpha_p, \sigma^2$  of the linear process, given only the wrapped process.

Breckling (1989) developed inferential methods based on the correlogram of wrapped processes. More recently, treating the problem as one of missing data, Fisher and Lee (1994) proposed use of the expectation maximisation (EM) algorithm. The calculations involved in the E-step are ‘prohibitive’, and an approximation is used, although for higher order models the resulting equations are even more complicated. They assert, therefore, that the wrapped normal approach is only feasible for AR models of low order.

Coles (1998) also notes the inferential difficulties associated with the model, and develops a Bayesian approach using Markov chain monte carlo (MCMC) methods. Some problems arise in the case of large variance, for which stronger priors are needed to overcome the problem of wrapping coefficients and large variance components confounding. He simulates a linear AR(1) process, which is then wrapped around the circle, and uses the MCMC procedure in order to estimate parameters. Model selection is explored briefly by showing the AR(2) parameter to be redundant. Estimates are shown to agree closely with the original parameter values  $\sigma = 1$  and  $\alpha_1 = 0.8$ . The approach is also used to estimate the parameters of a WAR(2) process applied to a series of wind data.

Fisher and Lee (1994) give the circular correlation function  $\rho_T$  of  $\Theta_1 = Y_1(\text{mod}2\pi)$  and  $\Theta_2 = Y_2(\text{mod}2\pi)$  for the bivariate normally distributed  $(Y_1, Y_2)$  and derive from it the circular autocorrelation function of the circular WAR( $p$ ) process. The former is given by

$$\rho_T = \frac{\sinh(2\rho\sigma_1\sigma_2)}{\{\sinh(2\sigma_1^2)\sinh(2\sigma_2^2)\}^{\frac{1}{2}}},$$

where  $\text{var}(X_i) = \sigma_i^2$ ,  $i = 1, 2$  and  $\text{corr}(X_1, X_2) = \rho$ . The latter, with  $\{\eta_t\}$  viewed as the result of wrapping the AR( $p$ ) process  $X_t$  around the circle, is given by

$$\rho_T(k) = \frac{\sinh\{2\rho(k)\sigma_2/(1 - \alpha_1\rho(1) - \dots - \alpha_p\rho(p))\}}{\sinh\{2\sigma_2/(1 - \alpha_1\rho(1) - \dots - \alpha_p\rho(p))\}},$$

where  $\rho(k)$  is the  $k$ -lag (linear) autocorrelation function of  $\{X_t\}$ , and  $\alpha_1, \dots, \alpha_p, \sigma^2$  are its AR( $p$ ) parameters. They suggest solving

$$\hat{\rho}_T(k) = \sinh(2c_0\hat{\rho}(k))/\sinh(2c_0)$$

in order to estimate  $\rho(k)$ , where  $\hat{\rho}_T(k)$  is as in (5.3),  $c_0$  is obtained from the equation  $\overline{R} = \exp(-c_0/2)$  and  $\overline{R}$  is the mean resultant length of the elements of the series  $\{\eta_t\}$ .

## 5.6 Projected processes

Another situation in which the problem of missing data arises is in that of the projected Gaussian process, due to Fisher and Lee (1994). Suppose  $\{X_t\}$  and  $\{Y_t\}$  are two independent, stationary Gaussian time series, each with zero mean. The point  $(X_t, Y_t)$  can be represented in polar coordinates as  $(R_t, \Theta_t)$ , say. Then  $\{\Theta_t\}$  is a stationary circular time series with uniform marginal distributions (Fisher and Lee, 1994).

Noting that we observe  $\Theta_t$  and not  $R_t$ , the EM algorithm is again proposed for parameter estimation. As for the wrapped model, computational complexity limits potential application to low order AR processes.

Fisher and Lee (1994) show that, if  $(X_1, Y_1)$  and  $(X_2, Y_2)$  are independent random vectors from a bivariate normal distribution with variances both equal to  $\sigma_2$  and correlation  $\rho$ , and if  $\Theta_1$  and  $\Theta_2$  are defined by  $(X_i, Y_i) = R_i(\cos \Theta_i, \sin \Theta_i)$ ,  $i = 1, 2$ , then the circular correlation  $\rho_T$  is given by

$$\rho_T = \frac{\pi^2}{16} \rho^2 (1 - \rho^2)^2 \left\{ {}_2F_1 \left( \frac{3}{2}, \frac{3}{2}; 2; \rho^2 \right) \right\}, \quad (5.5)$$

where  ${}_2F_1$  is the hypergeometric function. If the linear processes  $\{X_t\}$  and  $\{Y_t\}$  have the common  $k$ -lag autocorrelation function  $\rho(k)$ , then the  $k$ -lag circular autocorrelation function  $\rho_T(k)$  of  $\{\Theta_t\}$  is obtained by replacing  $\rho^2$  by  $\rho^2(k)$  in (5.5). An estimate of  $\rho(k)$  is given by replacing  $\rho_T(k)$  in the resulting expression by  $\hat{\rho}_T(k)$  from (5.3), and then solving for  $\rho(k)$ .

## 5.7 Choosing between the models

Fisher and Lee (1994) discuss the types of data that the models they describe (linked, wrapped and projected) can accommodate. They give the general advice that the first of these should be preferred for data that are reasonably highly concentrated, whilst the latter two better accommodate dispersed data. This is qualified by the observation that each is fairly flexible and able to accommodate different types of data.

Breckling (1989) gives a comparison of von Mises and wrapped normal variables preceding his time series analysis, noting that the two approximate each other reasonably well, especially for very large or very small  $\kappa$ . He also notes the complications associated with inference based on the wrapped normal model, and suggests therefore that the von Mises process may be preferred when inference is the aim. Breckling's von Mises process is characterised by changing concentration over time, indicating use of this model for non-stationary series.

## 5.8 Conclusions

We have surveyed and summarised the main contributions to the literature of directional time series models with discussion of four approaches – the von Mises AR process, linked, wrapped and projected AR models. It is evident that the wealth of methods and techniques available for the analysis of linear time series is not easily extendable to a directional setting. In a comment on work by Erwin et al. (2002), Kent and Mardia (2002) comment that intractability of the equilibrium distribution or the conditional distribution (or both) of directional models “appears to be universal law”. More direct methods involving link functions or wrapping lead to problems of missing data. Jammalamadaka and SenGupta (2001) also succinctly summarise the work done on directional time series, including work from PhD theses.



# Chapter 6

## Sine and Cosine time series models

### 6.1 Introduction

In this chapter we present two first order directional time series models based on the conditional distributions of the bivariate Sine and Cosine models (first introduced in Chapter 1 and further studied in Part I of the thesis in the case of the former). For the latter, we present the model in vectorial form, a slightly generalised version of the von Mises AR(1) model of Breckling (1989) described in Section 5.3. Section 6.2 focuses on the Sine model, whilst Section 6.3 discusses the Cosine model.

The Sine model (first described in Section 1.4) is adapted to a time series setting in Section 6.2.1, which section also considers the model under high concentration. The behaviour of the deterministic component of the model is explored in detail in Section 6.2.2, and this behaviour related to the values of model parameters. Data simulation is considered in Section 6.2.3, and the behaviour of the simulated data related both to the deterministic and random components of the model. The issue of bimodality is also discussed in this section. In Section 6.2.4 we derive the likelihood function for the Sine

time series model and investigate parameter estimation using this likelihood function, based on simulated data sets and various parameter configurations. The Sine time series model is applied to the  $\phi$  and  $\psi$  angles of a particular protein in Section 6.2.5. An analysis of the goodness-of-fit of the model is explored in some detail in this section, including the use of smoothing and kernel density estimation to compare features of the raw data with features of the fitted model. A conclusion of the observations made throughout Section 6.2 is given in Section 6.2.6.

The analysis of the Cosine time series model proceeds in a very similar fashion to that of the Sine time series model. The model is outlined in Section 6.3.1, and related to the von Mises AR process of Breckling (1989) defined in Section 5.3. The model under high concentration is also considered in this section. The relationship with the Cosine model (defined in Section 1.4) is highlighted in Section 6.3.2. Model properties are discussed in Section 6.3.3, with the main focus being on an analysis of the behaviour of the deterministic component of the model. As for the Sine time series model, this is followed by a discussion of data simulation in Section 6.3.4, in which section simulated data are related, in terms of their behaviour, to the values of model parameters. The likelihood function and parameter estimation are discussed in Section 6.3.5, whilst application of the model to the same data as considered for the Sine time series model is outlined in Section 6.3.6. This section also compares the Sine and Cosine time series models in terms of their goodness-of-fit to the protein data. Finally, Section 6.3.7 summarises the observations made on the Cosine model.

The possible extension of the Sine and Cosine time series models to  $AR(p)$  models is discussed in Section 6.4. Also outlined in this section is another potential extension of the Cosine model to a Kalman filter setting, as proposed by Kent and Mardia (2002) in the discussion of a paper by Erwin et al. (2002).



## 6.2 The Sine time series model

### 6.2.1 The model

In this section we revisit the Sine model and adapt it to a circular time series setting. The joint density of the two circular random variables  $\theta_1$  and  $\theta_2$  for the Sine model is given by

$$f(\theta_1, \theta_2) = C \exp\{\kappa_1 \cos(\theta_1 - \mu_1) + \kappa_2 \cos(\theta_2 - \mu_2) + \lambda \sin(\theta_1 - \mu_1) \sin(\theta_2 - \mu_2)\}, \quad (6.1)$$

where  $-\pi \leq \theta_1, \theta_2 < \pi$ ,  $\kappa_1, \kappa_2 \geq 0$ ,  $-\infty < \lambda < \infty$  and  $-\pi \leq \mu_1, \mu_2 < \pi$ .  $C$  is the appropriate normalization constant, given by

$$C^{-1} = 4\pi^2 \sum_{m=0}^{\infty} \binom{2m}{m} \left( \frac{\lambda^2}{4\kappa_1\kappa_2} \right)^m I_m(\kappa_1) I_m(\kappa_2).$$

For large  $\kappa_1, \kappa_2$  the distribution (6.1) reduces to a bivariate normal distribution with covariance parameters

$$\sigma_1^2 = \frac{\kappa_2}{\kappa_1\kappa_2 - \lambda^2}, \quad \sigma_2^2 = \frac{\kappa_1}{\kappa_1\kappa_2 - \lambda^2}, \quad \rho = \frac{\lambda}{\sqrt{\kappa_1\kappa_2}}, \quad (6.2)$$

imposing the restriction  $\lambda^2 < \kappa_1\kappa_2$ .

In order to formulate a time series model we replace  $\theta_2$  with  $\theta_t$ , and  $\theta_1$  with  $\theta_{t-1}$  in (6.1), and use the conditional density of  $\theta_t$  given  $\theta_{t-1}$ . As parameters of the equilibrium distribution of the process, we also make the replacements  $\kappa_1 = \kappa_2 = \kappa$ , say, and  $\mu_1 = \mu_2 = \mu$ , say. Equation (1.15) gave the conditional distribution of  $\theta_2$  given  $\theta_1$  as von Mises with density

$$\begin{aligned} f(\theta_2|\theta_1) &= [2\pi I_0(a_1)]^{-1} e^{a_1 \cos(\theta_2 - \mu_2 - b_1)} \\ &= [2\pi I_0(a_1)]^{-1} e^{\kappa_2 \cos(\theta_2 - \mu_2) + \lambda \sin(\theta_1 - \mu_1) \sin(\theta_2 - \mu_2)} \end{aligned} \quad (6.3)$$

where  $a_1 = \{\kappa_2^2 + \lambda^2 \sin^2(\theta_1 - \mu_1)\}^{1/2}$ . Adapting this density as described above to a time series setting, we obtain the distribution of  $\theta_t$  given  $\theta_{t-1}$  as von Mises with density

$$\begin{aligned} f(\theta_t|\theta_{t-1}) &= [2\pi I_0(\kappa_t)]^{-1} e^{\kappa_t \cos(\theta_t - \mu_t)} \\ &= [2\pi I_0(\kappa_t)]^{-1} e^{\kappa \cos(\theta_t - \mu) + \lambda \sin(\theta_t - \mu) \sin(\theta_{t-1} - \mu)} \end{aligned} \quad (6.4)$$

where

$$\kappa_t = \{\kappa^2 + \lambda^2 \sin^2(\theta_{t-1} - \mu)\}^{1/2} \quad \text{and} \quad \mu_t = \mu + \tan^{-1} \left\{ \frac{\lambda}{\kappa} \sin(\theta_{t-1} - \mu) \right\},$$

for  $t = 2, \dots, n$ , say. As a circular AR(1) process, the model can be equivalently stated as

$$\theta_t = \mu + \tan^{-1} \left\{ \frac{\lambda}{\kappa} \sin(\theta_{t-1} - \mu) \right\} + \epsilon_t, \quad (6.5)$$

where  $\epsilon_t \sim M(0, \kappa_t)$ ,  $t = 2, \dots, n$ . Immediately it can be seen that the model has the feature of changing concentration over time.

Since the joint density of  $\theta_t$  and  $\theta_{t-1}$  (Equation (6.1) with the appropriate substitutions) is symmetric in  $\theta_t$  and  $\theta_{t-1}$  it follows that the model is time-reversible and the equilibrium distribution  $\Pi(\theta)$  of the process is given by the univariate marginal distribution of either  $\theta_1$  or  $\theta_2$ , the former given in Equation (1.14). That is,

$$\Pi(\theta) = 2\pi C I_0 \left\{ [\kappa^2 + \lambda^2 \sin^2 \theta]^{1/2} \right\} e^{\kappa \cos \theta}. \quad (6.6)$$

If the  $\theta$  values are highly concentrated around  $\mu$ , then the differences  $\theta_t^* = \theta_t - \mu$  will be small  $\forall t = 1, \dots, n$ . In this case, from (6.4) and using a second order Taylor series approximation, we have

$$\begin{aligned} f(\theta_t^* | \theta_{t-1}^*) &\propto \exp \left\{ \kappa \cos \theta_t^* + \lambda \sin \theta_t^* \sin \theta_{t-1}^* \right\} \\ &\approx \exp \left\{ \kappa \left[ 1 - (\theta_t^*)^2/2 \right] + \lambda \theta_t^* \theta_{t-1}^* \right\} \\ &\propto \exp \left\{ -\frac{\kappa}{2} (\theta_t^*)^2 + \lambda \theta_t^* \theta_{t-1}^* \right\}. \end{aligned} \quad (6.7)$$

Rewriting the quadratic  $c_1(\theta_t^*)^2 + c_2\theta_t^*$  in (6.7) as  $c_1(\theta_t^* + \frac{c_2}{2c_1})^2 - \frac{c_2^2}{4c_1}$ , where  $c_1$  and  $c_2$  are constants (with respect to  $\theta_t^*$ ), (6.7) becomes

$$f(\theta_t^* | \theta_{t-1}^*) \propto \exp \left\{ -\frac{\kappa}{2} \left( \theta_t^* - \frac{\lambda \theta_{t-1}^*}{\kappa} \right)^2 \right\},$$

which is proportional to the density of a normal distribution with mean  $\lambda \theta_{t-1}^* / \kappa$  and variance  $1/\kappa$ . We therefore see that, for  $\theta$  values highly concentrated around  $\mu$ , the process

$\{\theta_t\}$ ,  $t = 1, \dots, n$  can be approximated by the linear AR(1) process

$$\theta_t^* = \rho\theta_{t-1}^* + \epsilon_t, \quad \epsilon_t \sim N(0, 1/\kappa),$$

where  $\rho = \lambda\theta_{t-1}^*/\kappa$ . Conditions under which the  $\theta$  values can be highly concentrated around  $\mu$  are outlined in Section 6.2.3.

We begin by analysing the deterministic component of the model, namely the mean direction  $\mu_t = \mu + \tan^{-1} \left\{ \frac{\lambda}{\kappa} \sin(\theta_{t-1} - \mu) \right\}$ .

## 6.2.2 The deterministic component

The deterministic component of the model is given by the equation

$$\mu_t = \mu + \tan^{-1} \left\{ \frac{\lambda}{\kappa} \sin(\theta_{t-1} - \mu) \right\}, \quad t = 2, \dots, n. \quad (6.8)$$

We investigate the behaviour of this mean direction for different values of  $\lambda$  and  $\kappa$ , restricting attention, without loss of generality, to  $\mu = 0$ . We illustrate the problem by plotting  $\mu_t$  for four different parameter combinations and a range of values for  $\theta_{t-1}$ . Figure 6.1 shows  $\mu_t$  plotted against  $\theta_{t-1}$ , the latter over the interval  $[-\pi, \pi)$ . Figure 6.1 (a) shows the mean direction for the parameter combinations  $(\lambda, \kappa) = (4, 1), (7, 8), (4, 8)$  and  $(1, 8)$ . Plot (b) of Figure 6.1 shows the effect of negating  $\lambda$ , namely a reflection in the line  $\mu = 0$  on the  $y$ -axis. For reference, the lines  $\mu_t = \theta_{t-1}$  on plot (a) and  $\mu_t = -\theta_{t-1}$  on plot (b) are also displayed.

A number of features of the deterministic component of the model can be observed from both Equation (6.8) and Figure 6.1. We will focus first on the case  $\lambda > 0$ . Adhering for the time being to the restriction  $|\lambda| < \kappa$  that ensures the bivariate density (6.1) is asymptotically normal, we see from Equation (6.8) that  $|(\lambda/\kappa) \sin \theta_{t-1}| < 1$ , and hence that, since  $\mu = 0$ , we have  $|\mu_t| < \pi/4$ . This means that, regardless of the value of  $\theta_{t-1}$ ,  $\mu_t$  is no further than  $\pi/4$  from the overall mean direction.

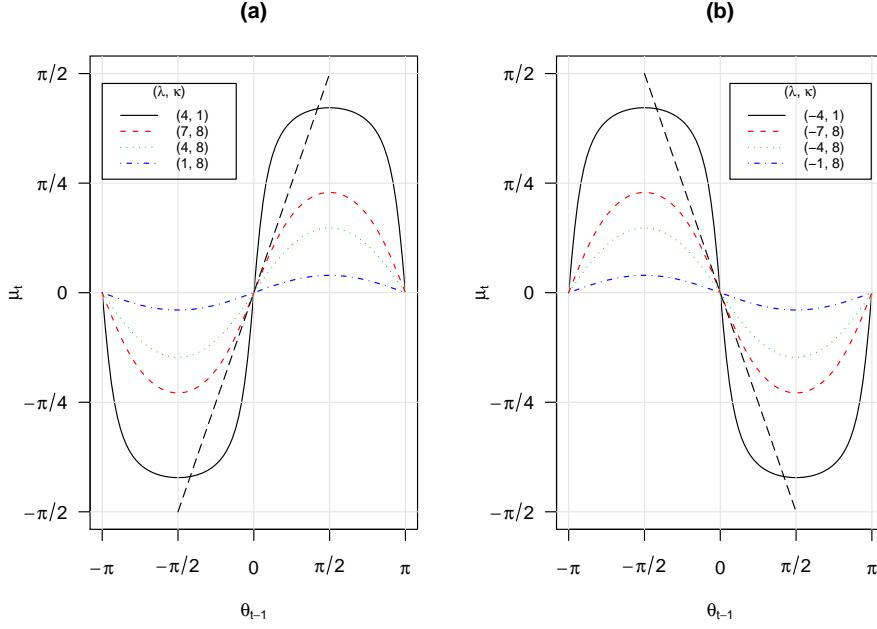


Figure 6.1: Plots of  $\mu_t$  versus  $\theta_{t-1}$  for parameter configurations  $(\lambda, \kappa) = (\pm 4, 1)$ ,  $(\pm 7, 8)$ ,  $(\pm 4, 8)$  and  $(\pm 1, 8)$  for the Sine time series model.

From Figure 6.1 we see that, for  $|\theta_{t-1}| > \pi/2$ , the further  $\theta_{t-1}$  is from the overall mean (zero), the closer  $\mu_t$  is to this value, a feature that would perhaps not be expected of an observed data set. Another unusual feature of the model is that the value of  $\mu_t$  given  $\theta_{t-1}$  is the same as the value of  $\mu_t$  given  $\pi - \theta_{t-1}$ . Since, based on the deterministic component of the model, we observe unusual behaviour for values of  $\theta_{t-1}$  greater than  $\pi/2$  in absolute value, and, as mentioned above, there can be at most one value of  $\theta$  outside the range  $(-\pi/4, \pi/4)$ , namely  $\theta_1$ , it is instructive to consider the probability of this value being outside the interval  $(-\pi/2, \pi/2)$ . In doing this we are no longer limiting attention to the deterministic component of the model. The required probability can be calculated using the equilibrium distribution of the process (Equation (6.6)). The probability  $p$  that the value  $\theta_1$  falls outside the interval  $(-\pi/2, \pi/2)$  is given by the expression

$$p = 2 \int_{\pi/2}^{\pi} \Pi(\theta) d\theta.$$

For the parameter configurations  $(|\lambda|, \kappa) = (7, 8)$ ,  $(4, 8)$  and  $(1, 8)$  we obtain (using numerical integration)  $p = 0.0007$ ,  $p = 0.0002$  and  $p = 0.00001$  respectively. This feature of the model is therefore unlikely to cause a problem if it is considered for fitting a highly concentrated data set.

Also observable in Figure 6.1 is the effect of changing the value of the ratio  $\lambda/\kappa$ . In particular, as this ratio approaches  $\pm 1$ , so the maximum at  $\theta_{t-1} = \pm\pi/2$  approaches  $\pi/4$  and the minimum at  $\theta_{t-1} = \mp\pi/2$  approaches  $-\pi/4$ . If  $\kappa$  is large relative to  $|\lambda|$ , then the ratio  $|\lambda|/\kappa$  is small, so that  $\mu_t$  is constrained to a small arc around zero, which is intuitively appropriate. If on the other hand  $\lambda$  is large relative to  $\kappa$ , then  $\mu_t$  is less constrained and the series less concentrated.

For  $\lambda < 0$ , we observe from the right hand plot of Figure 6.1 that if  $\theta_{t-1} > 0$  then  $\mu_t < 0$ , and conversely a negative value of  $\theta_{t-1}$  leads to a positive value of  $\mu_t$ . More specifically,

$$\mu_t|\theta_{t-1}; (\lambda, \kappa) = \mu_t|-\theta_{t-1}; (-\lambda, \kappa) = -\mu_t|\theta_{t-1}; (-\lambda, \kappa).$$

We now consider the effect of allowing  $|\lambda| > \kappa$ . We observe immediately that the value of  $\mu_t$  is no longer constrained to the interval  $(-\pi/4, \pi/4)$ , but instead to the interval  $(-\pi/2, \pi/2)$ . From Figure 6.1 (a), comparing the curve for  $(\lambda, \kappa) = (4, 1)$  with the line  $\mu_t = \theta_{t-1}$ , we see that for values of  $\theta_{t-1}$  between 0 and  $\pi/2$ , we have  $\mu_t > \theta_{t-1}$ . In other words, values of  $\theta_{t-1}$  between zero and  $\pi/2$  are rotated positively toward  $\tan^{-1} \left\{ \frac{\lambda}{\kappa} \sin \frac{\pi}{2} \right\} = \tan^{-1} \left\{ \frac{\lambda}{\kappa} \right\}$ . The latter is an increasing function of  $\lambda/\kappa$ , and tends to  $\pi/2$  as the ratio tends to infinity. In other words, the larger the ratio  $\lambda/\kappa$ , the closer to  $\pi/2$  the deterministic model rotates  $\theta$  values. As can be seen from Figure 6.1 (a), there is a value  $\theta_{t-1} \in (0, \pi/2]$  for which  $\mu_t = \theta_{t-1}$  when  $\lambda > \kappa$ . This point is given by the solution to the equation  $\tan \theta = \frac{\lambda}{\kappa} \sin \theta$ , namely  $\theta = \cos^{-1} \frac{\kappa}{\lambda}$ . Again, this tends to  $\pi/2$  as  $\lambda/\kappa \rightarrow \infty$ . If  $\theta_{t-1} = \cos^{-1} \frac{\kappa}{\lambda}$ , then all future  $\theta$  values are also equal to this value, based on the deterministic component of the model. Similar observations can be made for  $\theta_{t-1} \in (-\pi/2, 0)$ . The unusual behaviour of the deterministic model for  $|\theta_{t-1}| > \pi/2$

is again apparent when  $\lambda > \kappa$ . Similar observations can be made in the present case regarding the effect of negative values of  $\lambda$  as above for the case  $\lambda < \kappa$ .

In the next section we combine the deterministic and random components of the model in order to simulate data, and investigate properties of the resulting data in view of both the observations made in the present section and the effect of including the random component.

### 6.2.3 Data simulation

In order to simulate data from the Sine time series model (6.5) we will take  $\mu$  to be zero. The value of  $\theta_1$  is simulated using steps 1–3 of the algorithm described in Section 2.4. The values of  $\theta_2, \dots, \theta_n$  are then simulated according to (6.5). Figure 6.2 displays six data sets each based on different parameter combinations  $(\lambda, \kappa)$ . From plots (a), (b) and (c) of Figure 6.2 we can see the effect of the values of  $\lambda$  and  $\kappa$  with  $\lambda > \kappa$ . Plots (d) and (e) can be compared to see the effects of the parameters when  $\lambda < \kappa$ , whilst Figure 6.2 (f) gives an idea of the kind of data obtained for negative  $\lambda$  with  $|\lambda| > \kappa$ .

A summary of the effects of parameter values is given below. For clarity, we consider  $\lambda > 0$  and  $\lambda < 0$  separately. For each of these, we also consider separately the cases  $|\lambda| > \kappa$  and  $|\lambda| \leq \kappa$ .

1.  $\lambda > 0$ :

- (a)  $\lambda > \kappa$

In order to understand the behaviour observed in plots (a)–(c) of Figure 6.2 we need to consider the equilibrium distribution of the process, given in Equation (6.6). As discussed in Section 1.4, this distribution is symmetric around  $\mu$  and either unimodal with mode at  $\mu$  or bimodal with modes at  $\mu \pm \theta^*$ , where  $\theta^*$  satisfies Equation (1.18). In the current context, the distribution is unimodal

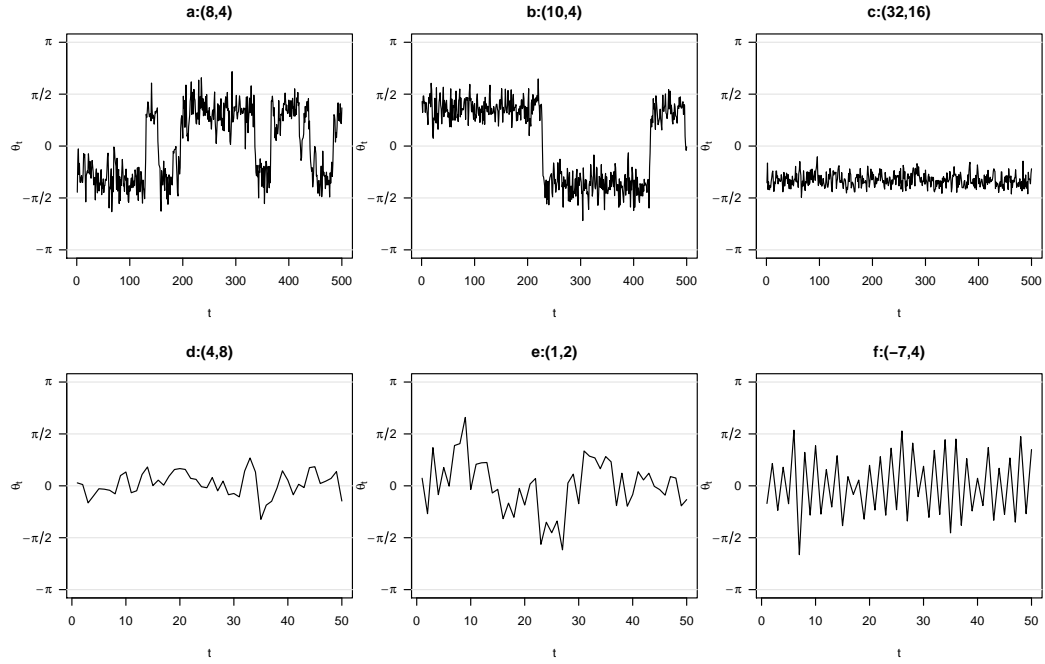


Figure 6.2: Simulated data for the Sine time series model. Values of  $(\lambda, \kappa)$  are given in brackets above each plot.

if and only if  $A_1(\kappa) \leq \kappa^2/\lambda^2$ . Since  $A_1(\kappa) \leq 1$ , a sufficient condition for unimodality is given by  $|\lambda| \leq \kappa$ . Although  $|\lambda| > \kappa$  is not a sufficient condition for bimodality, if  $\lambda$  and  $\kappa$  are not very close in absolute value and  $\kappa$  is not very small,  $|\lambda| > \kappa$  tends to give a bimodal equilibrium distribution.

Moving to data simulation, starting with  $\kappa = \lambda$  and increasing  $\lambda$  whilst keeping  $\kappa$  fixed creates clusters of  $\theta$  values close to  $\pi/2$  and  $-\pi/2$ . This is due to the deterministic component of the model inflating  $\theta$  values that are small in absolute value towards either  $\pi/2$  or  $-\pi/2$  (see Section 6.2.2). If the ratio  $\lambda/\kappa$  and/or the values of  $\kappa$  and  $\lambda$  themselves are small enough, then the  $\theta$  values can switch between clustering close to  $\pi/2$  and clustering close to  $-\pi/2$ . For the former, the smaller the ratio  $\lambda/\kappa$ , the further from  $\pi/2$  the  $\theta$  values cluster and the more likely a  $\theta$  value passes zero. For the latter, smaller

$\kappa$  and  $\lambda$  values lead to smaller values of  $\kappa_t = \sqrt{\{\kappa^2 + \lambda^2 \sin^2 \theta_{t-1}\}}$  and more dispersed values of the random component  $\epsilon_t$ , thus increasing the probability of  $\theta$  values passing through either zero or  $2\pi$ . Using a similar argument, if the ratio  $\lambda/\kappa$  and/or the values of  $\kappa$  and  $\lambda$  themselves are sufficiently large, then  $\theta$  values cluster either around  $\pi/2$  or around  $-\pi/2$ , but do not switch between the two. All of these features are observable in Figure 6.2 (a)–(c).

(b)  $\lambda \leq \kappa$

In this case  $\theta$  values are rotated towards the overall mean (zero) before a random term is added. Figure 6.2 (d) and (e) show the effect of keeping the ratio  $\lambda/\kappa$  fixed while decreasing (or increasing) the values of  $\kappa$  and  $\lambda$  themselves, namely that the lower values give rise to a more dispersed series. Fixing the value of  $\lambda$  and increasing  $\kappa$  leads to both a smaller ratio  $\lambda/\kappa$  and larger values of  $\kappa_t$ , giving a series more concentrated around zero. Fixing  $\kappa$  and increasing  $\lambda$  gives a larger ratio  $\lambda/\kappa$ , meaning  $\theta$  values are not shrunk towards zero as much, and random terms are more concentrated around zero.

2.  $\lambda < 0$ :

Figure 6.2 (f) shows a simulated data set with a negative value of  $\lambda$ . As can be seen, the data set comprises a series of  $\theta$  values that tend to alternate between positive and negative values. Based on the model (6.5) we see that, with  $\mu = 0$ , neither the sign of  $\theta_{t-1}$  nor the sign of  $\lambda$  affects the value of  $\kappa_t$ . As discussed in Section 6.2.2, the deterministic component of the model satisfies the equation  $\mu_t|\theta_{t-1}; (\lambda, \kappa) = -\mu_t|\theta_{t-1}; (-\lambda, \kappa)$ . Thus, negating every second  $\theta$  value of a series for which  $(\lambda, \kappa) = (-l, k)$  gives a series for which  $(\lambda, \kappa) = (l, k)$ . If  $\mu \neq 0$  then the above holds for the mean corrected data  $\theta_t - \mu$ .

Having successfully simulated data sets, attention turns to the likelihood function and parameter estimation based on such a data set.



### 6.2.4 Likelihood function and parameter estimation

Given a series  $\{\theta_t\}$  based on Equation (6.4), the likelihood function is given by

$$\begin{aligned} L &= \Pi(\theta_1) \prod_{t=2}^n f(\theta_t | \theta_{t-1}) \\ &= C I_0 \{[\kappa^2 + \lambda^2 \sin^2 \theta_1]^{1/2}\} e^{\kappa \cos \theta_1} \prod_{t=2}^n \frac{\exp \{\kappa \cos \theta_t + \lambda \sin \theta_t \sin \theta_{t-1}\}}{I_0 \{[\kappa^2 + \lambda^2 \sin^2 \theta_{t-1}]^{1/2}\}} \\ &= \frac{C \exp \{\kappa \sum_{t=1}^n \cos \theta_t\} \exp \{\lambda \sum_{t=2}^n \sin \theta_t \sin \theta_{t-1}\}}{\prod_{t=3}^n I_0 \{[\kappa^2 + \lambda^2 \sin^2 \theta_{t-1}]^{1/2}\}}. \end{aligned}$$

The log likelihood  $l = \log L$  is therefore

$$l = \log C - \sum_{t=3}^n \log I_0 \{[\kappa^2 + \lambda^2 \sin^2 \theta_{t-1}]^{1/2}\} + \kappa \sum_{t=1}^n \cos \theta_t + \lambda \sum_{t=2}^n \sin \theta_t \sin \theta_{t-1}. \quad (6.9)$$

In order to obtain parameter estimates for a given data set, the negative of the loglikelihood (6.9) is minimised using the `nlm` function in R. Table 6.1 shows the results of fitting the Sine time series model to simulated data sets for the parameter configurations  $(\lambda, \kappa) = (4, 8), (0.5, 1), (8, 4)$  and  $(-3, 7)$ . For each configuration, 100 series of length  $n = 200$  are simulated, and the table gives the mean and standard deviation of estimates for  $\kappa$  and  $\lambda$ . For  $\mu$ , the elements in the table are the mean direction of estimates and the circular variance  $1 - \overline{R}$  of the estimates, where  $\overline{R}$  is the mean resultant length of the estimates.

$(\lambda, \kappa)$	$\hat{\mu}$	$\hat{\lambda}$	$\hat{\kappa}$
(4, 8)	-0.004 (0.001)	4.036 (0.726)	8.018 (0.804)
(0.5, 1)	0.011 (0.008)	0.505 (0.163)	1.005 (0.118)
(8, 4)	0.010 (0.031)	7.735 (1.718)	4.351 (1.008)
(-3, 7)	-0.006 (0.000)	-2.976 (0.558)	7.053 (0.677)

Table 6.1: Mean values (and standard deviations) of estimates for simulated Sine time series data.

It can be seen from the table that, based on the mean values, the true parameter values are recovered reasonably successfully, although  $\kappa$  and  $\lambda$  estimates tend to be slightly overestimated (the exception being  $\hat{\lambda}$  when  $(\lambda, \kappa) = (8, 4)$ ). We further investigate the behaviour of the loglikelihood function in the vicinity of the true parameter values by plotting the loglikelihood function in various ways for a given data set. Figure 6.3 shows profile loglikelihood plots for the two parameter configurations  $(\mu_0, \lambda_0, \kappa_0) = (0, 0.5, 1)$  and  $(0, -3, 7)$ . Plots (a) and (d) show the loglikelihood for a range of  $\lambda$  values in the neighbourhood of the true value  $\lambda_0$  with  $\mu = \mu_0$  and  $\kappa = \kappa_0$ , whilst plots (b) and (e) show the loglikelihood for a range of  $\kappa$  values in the neighbourhood of the true value  $\kappa_0$  with  $\mu = \mu_0$  and  $\lambda = \lambda_0$ . Finally, plots (c) and (f) show the likelihood for a range of  $\kappa$  and  $\lambda$  values with  $\mu$  fixed at  $\mu_0$ . On plots (c) and (f), horizontal and vertical lines show the true values of  $\lambda$  and  $\kappa$ , and are seen to be slightly away from where the loglikelihood is maximised for the given data sets. As seen from plots (a), (b), (d) and (e), the gradient of the loglikelihood in the vicinity of the maxima is such that the maxima are easily discernible.

As we appear to have a method of estimation that is reasonably successful at recovering true parameter values, we next apply the sine time series model to a real data set.

### 6.2.5 Application to protein data

We apply the Sine time series model to the 249  $\phi$  and  $\psi$  angles of the protein triosephosphate isomerase. Table 6.2 shows the maximum likelihood estimates of  $\mu$ ,  $\lambda$  and  $\kappa$  for each series  $\{\phi_t\}$  and  $\{\psi_t\}$ . Approximate standard errors are also given, and are obtained from the Hessian matrix calculated in the nlm estimation procedure.

Using these estimates we simulate two series  $\{\phi_t^{\text{sim}}\}$  and  $\{\psi_t^{\text{sim}}\}$ . Maximum likelihood estimates for these data sets are also given in Table 6.2, and are seen to be similar to the estimates for the real data set. As can also be seen from the table, we have  $\hat{\lambda} > \hat{\kappa}$  for

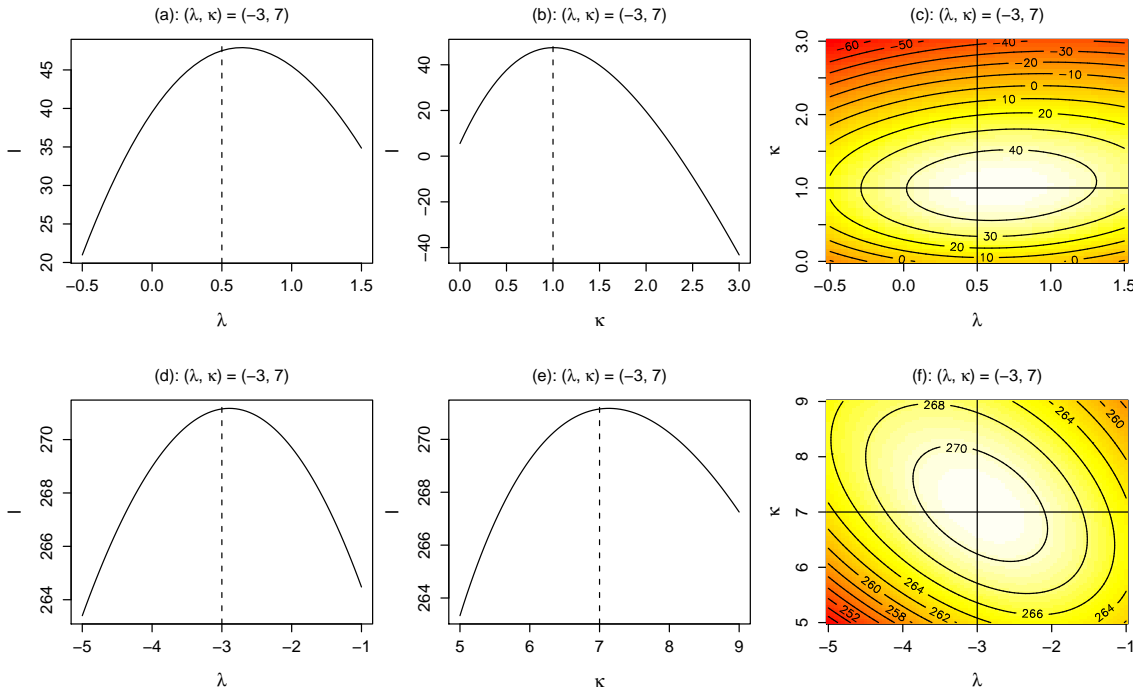


Figure 6.3: Profiles of loglikelihood function for simulated data sets in the vicinity of the true parameter values. True values of  $(\lambda, \kappa)$  are given in brackets above each plot.

the  $\psi$  series. Figure 6.4 displays circular and time series plots for the protein data and the simulated data.

From the plots for the  $\psi$  series of the protein data, the data appear to be bimodal, which would suggest  $\lambda > \kappa$ , as is the case for their estimates. Observable more from the circular plot of the  $\psi$  series is that the data are more highly concentrated around one of the modes than the other. This is not a feature that was observed in any of the data sets simulated in Section 6.2.3. The plots for  $\psi^{\text{sim}}$  give a representation of what data might be expected to look like if the true parameter values are the maximum likelihood estimates of the  $\psi$  protein data. Since the features of the plots for  $\psi$  and  $\psi^{\text{sim}}$  are quite different, this would indicate a poorly fitting model. One reason for this may be precisely that the model is

	$\phi$	$\phi^{\text{sim}}$	$\psi$	$\psi^{\text{sim}}$
$\hat{\mu}$	-1.564 (0.075)	-1.630 (0.067)	0.697 (0.069)	0.676 (0.054)
$\hat{\lambda}$	2.174 (0.364)	2.134 (0.299)	1.689 (0.149)	1.999 (0.180)
$\hat{\kappa}$	3.154 (0.253)	3.184 (0.256)	0.314 (0.107)	0.369 (0.103)

Table 6.2: Estimates (and standard errors) of Sine time series model parameters for protein and simulated data.

unable to capture the bimodality with differing concentration at each mode. As can be seen from the circular plots for  $\psi$  and  $\psi^{\text{sim}}$ , the overall distribution of values around the circle is quite different for the two. On the other hand, the general tendency to switch between two modes is evident from the time series plots for both  $\psi$  and  $\psi^{\text{sim}}$ .

Moving to the  $\phi$  series, the circular plots of  $\phi$  and  $\phi^{\text{sim}}$  show the distribution of the latter to be more symmetric. In the time series plots, there is a repeating feature observable for  $\phi$  in which short series of values are highly concentrated close to  $-\pi/2$ , which is not the case for the  $\phi^{\text{sim}}$  values. These differences again bring into question the goodness-of-fit of the model.

Certainly based on the time series plots for both  $\phi$  and  $\psi$  it does appear that a model with changing concentration over time is appropriate. Based on comparison of the true data and the simulated data, however, it also appears that the Sine time series model is not particularly appropriate for modelling these angles.

With the aims of further developing methods of measuring the goodness-of-fit of directional time series models and investigating the goodness-of-fit of the Sine model to the protein data being studied, we now consider two more plots that could potentially diagnose a poorly fitting model.

The left hand side of Figure 6.5 shows plots of  $\phi_t$  versus  $\phi_{t-1}$  and  $\psi_t$  versus  $\psi_{t-1}$  for the protein data being studied. The right hand side shows corresponding plots for  $\phi$  and  $\psi$

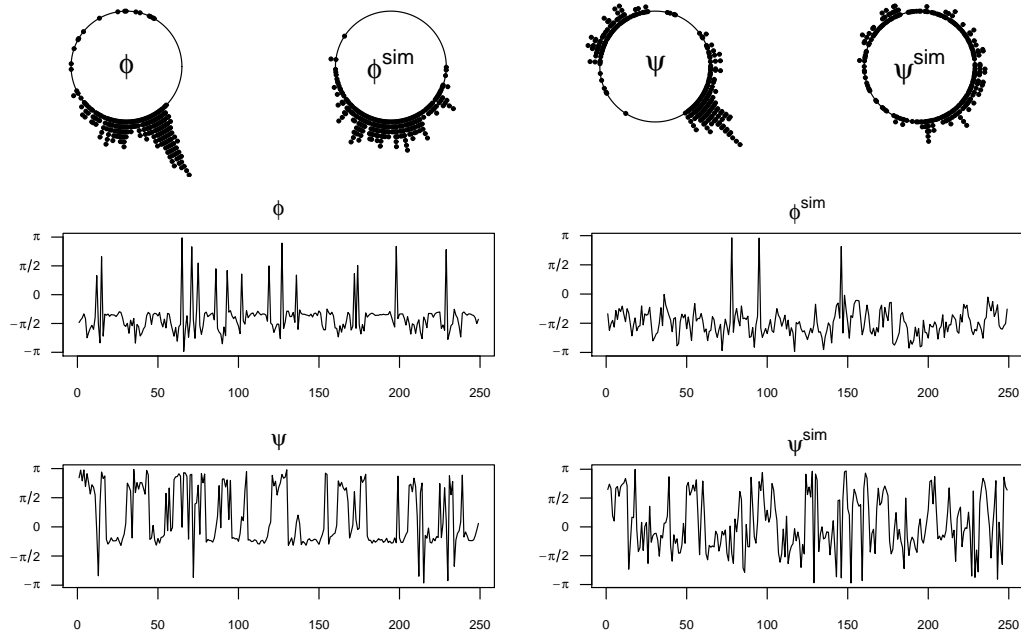


Figure 6.4:  $\phi$  and  $\psi$  conformational angles for a protein data set, and simulated  $\phi$  and  $\psi$  values based on maximum likelihood estimates for the true data.

values simulated using the maximum likelihood estimates for the protein data. In each of the four plots, the mle  $\hat{\mu}$  of  $\mu$  for the relevant data set has been subtracted from the original values prior to plotting. The solid black lines represent  $\hat{\mu}_t = \tan^{-1}\{\hat{\lambda}/\hat{\kappa} \sin \theta_{t-1}\}$ , where  $\theta$  is to be replaced by the relevant variable and ranges from  $-\pi$  to  $\pi$ . The dashed lines represent  $\hat{\mu}_t \pm 1/\sqrt{\hat{\kappa}_t}$ , where  $\hat{\kappa}_t = \sqrt{\hat{\kappa}^2 + \hat{\lambda}^2 \sin^2 \theta_{t-1}}$  and  $\theta$  is again to be replaced by the relevant variable. (The use of  $\hat{\mu}_t \pm 1/\sqrt{\hat{\kappa}_t}$  is motivated by the normal approximation to the von Mises distribution, for which, if  $\kappa$  is large, we have  $\sigma \approx 1/\sqrt{\kappa}$ ). The plots of  $\hat{\mu}_t \pm 1/\sqrt{\hat{\kappa}_t}$  in Figure 6.5 are representative of those obtained more generally, and highlight the tendency of non-stationarity being more prevalent when  $\lambda > \kappa$ . Finally, the red line represents a smoothed version of  $\theta_t$  versus  $\theta_{t-1}$ , obtained using the R function

```
smth.circ=function(x,y,grid=seq(-pi,pi,length=101),h){
  w=dvm(outer(x,grid,"-"),0,1/h)
```

```
w.mean=atan2(apply(w*sin(y),2,sum),apply(w*cos(y),2,sum))
w.mean}
```

in which  $x$  and  $y$  are circular variables. This function essentially calculates a weighted sample mean direction of  $y$  values at each point on the “grid” of values between  $-\pi$  and  $\pi$ . The weight  $w_i$ , say, applied to  $y_i$  at grid point  $j$ , is the von Mises density centred at  $x_i$ , with concentration  $1/h$  and evaluated at grid point  $j$ . Mathematically, the estimate  $\hat{m}_h(x)$ , at point  $x$  of the “grid”, dependent on bandwidth  $h$ , is

$$\hat{m}_h(x) = \tan^{-1} \left\{ \sum_{i=1}^n w_i(x) \sin y_i, \sum_{i=1}^n w_i(x) \cos y_i \right\}$$

where  $w_i \sim M(x_i, 1/h)$ .

The value of the smoothing parameter  $h$  is chosen by cross-validation, by minimising the sum of squared errors  $\sum_{i=1}^n (\hat{y}_i - y_i)^2$ , where  $\hat{y}_i$  is a weighted sample mean direction of  $y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n$ , given  $x_1, \dots, x_n$ . The weight  $v_j$ , say, applied to  $y_j$  in the calculation of  $\hat{y}_i$ , ( $i \neq j$ ), is the von Mises density, centred at  $x_i$ , with concentration  $1/h$  and evaluated at  $x_j$ . The following R function is used to this end:

```
smth.CV=function(x,y,h.start){
  smth.cv=function(h){
    n=length(y); y.hat=y; s=10^10; if (h>0){
      v=dvm(outer(x,x,"-"),0,1/h)
      v=v-diag(v)*diag(n)
      y.hat=atan2(apply(v*sin(y),2,sum),apply(v*cos(y),2,sum))
      s=(pmin(abs(y.hat-y),abs(y.hat+2*pi-y),abs(y.hat-2*pi-y)))
      s=sum(s^2)}
    s}
  nlm(smth.cv,h.start)}
```

If the model is a good fit to the data, it should be expected that the smoothed values approximately follow the line for  $\hat{\mu}_t$ . As can be seen from the plots, the approximation

does appear slightly superior for the simulated data than the protein data. The highly concentrated cluster of values at approximately  $(\psi_t, \psi_{t-1}) = (-\pi/2, -\pi/2)$  of the protein data appears to be keeping the smoothed line down whilst the line for  $\hat{\mu}_t$  increases. The asymmetric nature of the smoothed line in this case, in contrast to the symmetry of the smoothed line for the simulated data, perhaps indicated that a mixture of models would be more appropriate for these data. Based on these plots, therefore, we have some evidence to suggest that the Sine time series model is not such a good fit to the protein data being studied, particularly the  $\psi$  series.

Figure 6.6 assesses the goodness-of-fit based on the equilibrium distribution of the process (Equation (6.6)). The plots for the protein data are again on the left hand side of the figure, whilst those for the simulated data are on the right hand side. The dashed line in each case is the equilibrium distribution evaluated for  $(\mu, \lambda, \kappa) = (0, \hat{\lambda}, \hat{\kappa})$  and plotted between  $-\pi$  and  $\pi$ . The solid line in each case is a kernel density estimate, obtained using the R function `bkde`, based on the raw data minus  $\hat{\mu}$  for each variable.

It is important to note that for directional data a slight amendment of the kernel density estimation method is required. Since the data are only in the interval  $[-\pi, \pi)$  but in reality are cyclical with period  $2\pi$ , an adjustment is required to obtain the kde. In particular, the data  $Z$  are augmented with  $Z - 2\pi$  and  $Z + 2\pi$ , the kernel density estimated in the interval  $[-3\pi, 3\pi]$  and the final estimate taken as three times the middle third (interval  $[-\pi, \pi)$ ) of the overall estimate.

The bandwidth of the kernel is chosen visually to best approximate the plot of the equilibrium distribution. It may be argued that, for a data set for which the Sine time series model is appropriate, the raw data may not constitute a sample from the equilibrium distribution. In the present case, basing the kernel density estimate on values sampled four apart makes no discernible difference to the kernel density estimate of either the protein data or the simulated data. Another issue, the effects of which are perhaps observable in the plots for  $\psi$  and  $\psi^{\text{sim}}$  in Figure 6.6, is that if  $\lambda \gg \kappa$ , then even though the equilibrium

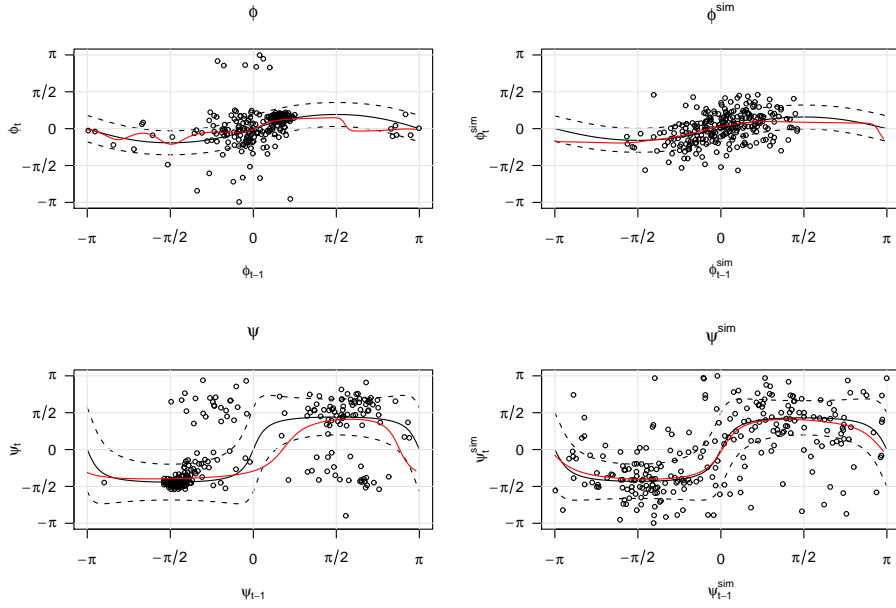


Figure 6.5: Plots of  $\theta_t$  versus  $\theta_{t-1}$ , a smoothed version (red line) and  $\hat{\mu}_t \pm 1/\sqrt{\hat{\kappa}_t}$ , where  $\theta$  is one of  $\phi, \phi^{\text{sim}}, \psi$  or  $\psi^{\text{sim}}$ .

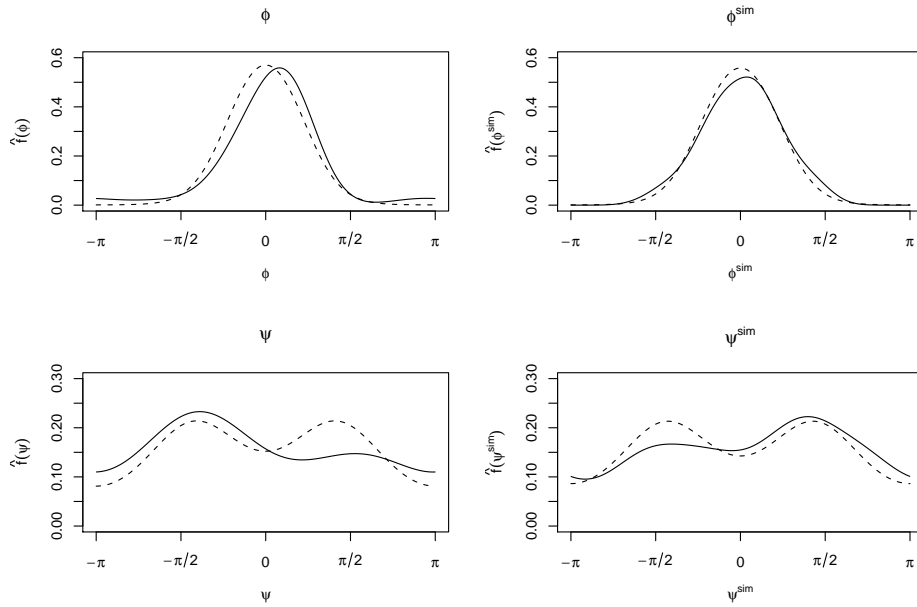


Figure 6.6: Marginal kde (solid) and pdf (dashed) based on data and MLEs respectively.



distribution is bimodal, a data set simulated using the Sine time series model may be unimodal (see Figure 6.2 (c)). Even if the simulated data are bimodal, they may not be symmetric about  $\mu$  (contrast the first 300 values of the data in Figure 6.2 (b) with the first 400: the latter are approximately symmetric about zero but the former are not). Therefore if the equilibrium distribution based on the fitted model is bimodal, then the goodness-of-fit plots of the type displayed in Figure 6.6 may not be appropriate. For those MLEs giving rise to a unimodal equilibrium distribution, however, it may be reasonable to expect a close agreement between the kernel density estimate and the fitted equilibrium distribution plot. In the present case for the protein and simulated  $\phi$  values, a poorly fitting model is not diagnosed by either plot.

### 6.2.6 Conclusions

In this section we have adapted the bivariate circular density studied by Singh et al. (2002) to a time series context using the associated conditional distribution. A number of features of the model have been investigated. In particular, we have investigated the behaviour of the deterministic model for various parameter configurations and related the observed behaviour to simulated data sets. A method of parameter estimation has been established, and the likelihood function of the model investigated. Fitting the model to a real protein data set has highlighted aspects of a poorly fitting model. The appropriateness of a model with changing concentration over time, however, is indicated by time series plots of the protein data. A number of goodness-of-fit measures have been developed which could be used more generally in the fitting of circular time series model. In particular, methods have been devised to compare “observed” conditional and equilibrium distributions with their “fitted” counterparts. The former are those based on smoothing and kernel density estimation of the raw data, the latter are those obtained from fitting a particular model, in this case the Sine time series model.

## 6.3 The Cosine time series model

In this section we present the system equation component of a filtering approach suggested by Kent and Mardia (2002) in a discussion of a paper by Erwin et al. (2002). This part of the model is also the first order AR von Mises process introduced by Breckling (1989) (see Section 5.3) without the restriction of a zero mean for the equilibrium distribution. It turns out that the conditional distribution on which the model is based is the conditional distribution associated with the Cosine model, which was introduced in Section 1.4.

### 6.3.1 The model

The system equation of the filter is determined by modelling the unit vector  $\mathbf{x}_t = (\cos \theta_t, \sin \theta_t)^T$  conditionally on  $\mathbf{x}_{t-1}$  as von Mises (see Definition 1.3.2)

$$\mathbf{x}_t | \mathbf{x}_{t-1} \sim VM(a\mathbf{x}_{t-1} + b\mathbf{e}), \quad (6.10)$$

where  $a$  and  $b > 0$  are scalars and  $\mathbf{e} = (\cos \mu, \sin \mu)^T$  is a unit vector. The constraint  $b > 0$  is imposed for identifiability, since in practice  $\mathbf{e}$  could be any unit vector. The equilibrium distribution will be seen to be symmetric about  $\mu$ . The pdf of  $\mathbf{x}_t | \mathbf{x}_{t-1}$ , from Equation (1.9), is

$$f(\mathbf{x}_t | \mathbf{x}_{t-1}) = [2\pi I_0(\| a\mathbf{x}_{t-1} + b\mathbf{e} \|)]^{-1} e^{\mathbf{x}_t^T (a\mathbf{x}_{t-1} + b\mathbf{e})}.$$

The model can be specified equivalently, and in the notation used for the Sine time series model, as  $\theta_t \sim M(\mu_t, \kappa_t)$ , where  $\mu_t = \tan^{-1}(a \sin \theta_{t-1} + b \sin \mu, a \cos \theta_{t-1} + b \cos \mu)$ ,  $\kappa_t = \sqrt{a^2 + b^2 + 2ab \cos(\theta_{t-1} - \mu)}$  and  $\tan^{-1}(q, p) \in [-\pi, \pi]$  is the angle between the positive  $x$ -axis and the vector  $(p, q)$ . With this specification we have

$$f(\theta_t | \theta_{t-1}) = \{2\pi I_0[\sqrt{(a^2 + b^2 + 2ab \cos(\theta_{t-1} - \mu))}]\}^{-1} \times \exp\{a \cos(\theta_t - \theta_{t-1}) + b \cos(\theta_t - \mu)\}. \quad (6.11)$$

If the  $\theta$  values are highly concentrated around  $\mu$  then small values of  $\theta_t - \mu = \theta_t^*$ , say, result. Using the approximation  $\cos \psi \approx 1 - \psi^2/2$  for small  $\psi$ , the density of  $\theta_t^*$  given  $\theta_{t-1}^*$  in this case is given by

$$\begin{aligned} f(\theta_t^* | \theta_{t-1}^*) &\propto \exp\{a \cos(\theta_t^* - \theta_{t-1}^*) + b \cos \theta_t^*\} \\ &\approx \exp\{a[1 - \frac{1}{2}(\theta_t^* - \theta_{t-1}^*)^2] + b[1 - (\theta_t^*)^2/2]\}. \end{aligned} \quad (6.12)$$

Rewriting the quadratic in  $\theta_t^*$  in the exponential of (6.12) as

$$c_1(\theta_t^*)^2 + c_2\theta_t^* + c_3 = c_1(\theta_t^* + c_2/2c_1)^2 + c_3 - c_2^2/4c_1,$$

(6.12) is proportional (in terms of  $\theta_t^*$ ) to

$$\exp\{-\frac{1}{2}(a+b)(\theta_t^* - \frac{a\theta_{t-1}^*}{a+b})^2\}. \quad (6.13)$$

Equation (6.13) is proportional to the density of the distribution  $\theta_t^* | \theta_{t-1}^* \sim N(a\theta_{t-1}^*/[a+b], 1/[a+b])$ , giving the linear AR(1) model

$$\theta_t^* = \rho\theta_{t-1}^* + \epsilon_t, \quad \epsilon_t \sim N(0, 1/[a+b]) \quad (6.14)$$

where  $\rho = a/(a+b)$ . The conditions under which the  $\theta$  values are highly concentrated about  $\mu$  are discussed in the next section.

### 6.3.2 Relationship with the Cosine model

In Section 1.4 we outlined the two bivariate von Mises distributions referred to as the Sine model and the Cosine model. In order to highlight the relationship between the latter and the model currently being considered, we review some of the details here. Rivest (1987) considered the set of bivariate models

$$\begin{aligned} f_1(\theta, \phi) &= C \exp\{\kappa_1 \cos(\theta - \mu) + \kappa_2 \cos(\phi - \nu) + \\ &\quad + \alpha \cos(\theta - \mu) \cos(\phi - \nu) + \beta \sin(\theta - \mu) \sin(\phi - \nu)\}, \end{aligned} \quad (6.15)$$

itself a submodel of a class considered by Mardia (1975). With  $\alpha = \beta = a$ ,  $\kappa_1 = \kappa_2 = b$ ,  $\mu = \nu = 0$ ,  $\theta = \theta_t$  and  $\phi = \theta_{t-1}$ , (6.15) reduces to

$$f_1(\theta_t, \theta_{t-1}) = C \exp \{b(\cos \theta_t + \cos \theta_{t-1}) + a \cos(\theta_t - \theta_{t-1})\}, \quad (6.16)$$

which is the bivariate density of the Cosine model given originally in Equation (1.13). The conditional distribution of  $\theta_t$  given  $\theta_{t-1}$  in this case is given by Equation (1.16). That is,

$$f(\theta_t | \theta_{t-1}) = \{2\pi I_0[\sqrt{(a^2 + b^2 + 2ab \cos(\theta_{t-1} - \mu))}]^{-1} \exp\{a \cos(\theta_t - \theta_{t-1}) + b \cos(\theta_t - \mu)\},$$

which is precisely the conditional distribution of the time series model being studied. In other words, the joint density of  $(\theta_t, \theta_{t-1})$  corresponding to the conditional density of  $\theta_t$  given  $\theta_{t-1}$  (Equation (6.11)) is nothing but the bivariate Cosine density.

As for the Sine time series model, since the bivariate Cosine density is symmetric in  $\theta_t$  and  $\theta_{t-1}$  it follows that the model is time-reversible. The equilibrium distribution  $\Pi(\theta)$  of the process is given by the univariate marginal distribution of either  $\theta_1$  or  $\theta_2$ . Specifically, the probability density function of the equilibrium distribution is

$$\Pi(\theta) = 2\pi C I_0[\sqrt{(a^2 + b^2 + 2ab \cos(\theta - \mu))}] e^{b \cos(\theta - \mu)}. \quad (6.17)$$

The normalizing constant  $C$  in the current setting is given by

$$C^{-1} = 4\pi^2 \{I_0(a)[I_0(b)]^2 + 2 \sum_{p=1}^{\infty} I_p(a)[I_p(b)]^2\}. \quad (6.18)$$

In terms of the random vector  $\mathbf{x}$ , the density of the equilibrium distribution is

$$\Pi(\mathbf{x}) = 2\pi C I_0(\|a\mathbf{x} + b\mathbf{e}\|) e^{b\mathbf{e}^T \mathbf{x}}. \quad (6.19)$$

The distribution (6.17) is not von Mises, but is seen to be approximately von Mises for small values of  $a$ . The distribution is also symmetric about  $\mu$ . It turns out (Mardia et al.,

2007b) that the equilibrium distribution (Equation (6.17)) is unimodal with mode at  $\mu$  for all  $a > 0$ , and for  $a < 0$  is unimodal if and only if

$$A_1(|a + b|) + \frac{|a + b|}{a} \leq 0, \quad (6.20)$$

where  $A_1(\cdot) = I_1(\cdot)/I_0(\cdot)$ . If (6.20) does not hold then  $\Pi(\theta)$  is bimodal with modes at  $\mu \pm \theta^*$ , where  $\theta^*$  is given by Equation (1.19). Since Equation (6.20) is a necessary condition for unimodality with mode at  $\mu$ , it also gives a necessary condition for  $\theta$  values to be highly concentrated around  $\mu$ , and therefore a necessary condition for the normality of the process under high concentration discussed in Section 6.3.1. Figure 6.7 shows an example of a unimodal and a bimodal equilibrium distribution for  $a < 0$ . In the unimodal case  $a = -2$  and  $b = 3$ , whilst the bimodal plot corresponds to when  $a = -4$  and  $b = 3$ .

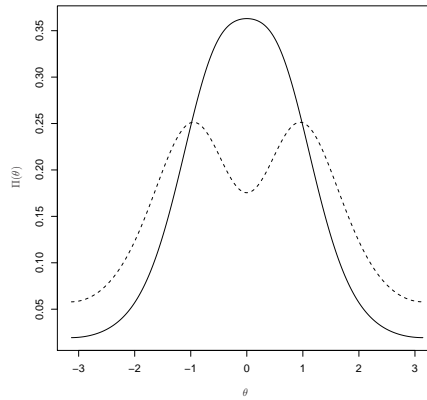


Figure 6.7: Plots of  $\Pi(\theta)$  versus  $\theta$  for  $a = -2$ ,  $b = 3$  (unimodal—solid line) and  $a = -4$ ,  $b = 3$  (bimodal—dashed line)

We next conduct an investigation into the behaviour of the model by analysing its deterministic and random components for various values of  $a$  and  $b$ .

### 6.3.3 Model properties

From the expression for  $\mu_t$ , one aspect of the deterministic component  $\mu_t$  of the model that is immediately apparent is that  $\mu_t|(a, b) = \mu_t|(ca, cb)$  for any constant  $c$ . From the expression for  $\kappa_t$ , on the other hand, we see that  $\kappa_t|(ca, cb) = c\kappa_t|(a, b)$ .

A further investigation into the properties of the model can be made by plotting the conditional mean  $\mu_t = \arg E(e^{i\theta_t}|\theta_{t-1})$ , for  $\theta_{t-1}$  ranging from  $-\pi$  to  $\pi$ . The top left plot of Figure 6.8 shows a plot of  $\mu_t$  (solid line) for  $\theta_{t-1}$  ranging from  $-\pi$  to  $\pi$  and when  $a = b = 2$  and  $\mu = 0$ . The dashed lines above and below the solid line represent  $\mu_t \pm 1/\sqrt{\kappa_t}$ . As departures from the shape of this plot, the cases  $|a| > b$  and  $|a| < b$  are also considered.

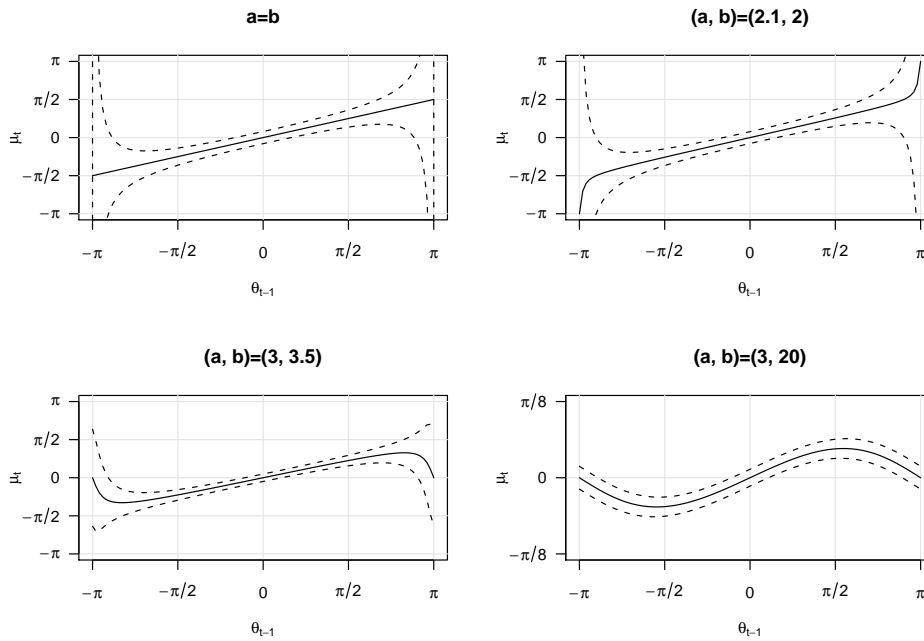


Figure 6.8: Plots of  $\mu_t(\pm 1/\sqrt{\kappa_t})$  versus  $\theta_{t-1}$  for (a)  $a = b$ , (b)  $a = 2.1$ ,  $b = 2$ , (c)  $a = 3$ ,  $b = 3.5$  and (d)  $a = 3$ ,  $b = 20$

The particular case  $a = b$  requires individual consideration. In this case, from (6.11), we

have

$$\begin{aligned} f(\theta_t|\theta_{t-1}) &\propto \exp\{a[\cos(\theta_t - \theta_{t-1}) + \cos \theta_t]\} \\ &= \exp\{2a \cos(\theta_{t-1}/2) \cos(\theta_t - \theta_{t-1}/2)\}, \end{aligned}$$

a von Mises distribution with concentration (at time  $t$ )  $\kappa_t = 2a \cos(\theta_{t-1}/2)$  and conditional mean  $\mu_t = \theta_{t-1}/2$ . The expression for  $\mu_t$  explains why, in Figure 6.8 (a), the domain of  $\mu_t$  when  $a = b$  and  $\theta_{t-1} \in (-\pi, \pi)$  is  $-\pi/2$  to  $\pi/2$ . Considering the expression for  $\kappa_t$ , if  $\theta_{t-1} = \pi (= -\pi)$ , then  $\kappa_t = 0$  so that, given  $\theta_{t-1} = \pi$ , the distribution of  $\theta_t$  is uniform on the circle. This explains the two dashed lines from  $-\pi$  to  $\pi$  on Figure 6.8 (a), each representing the possible values  $\mu_t$  can take for those particular values of  $\theta_{t-1}$ .

Starting at  $|a| = b$  and increasing  $|a|$ , the conditional mean plot displayed in the top right of Figure 6.8 is similar to that for  $a = b$ , without the vertical dashed lines and with the line extending from the existing solid line up to  $\pi$  and down to  $-\pi$  with slight curvature introduced. Further increasing  $|a|$  pushes the turning points towards zero in both directions on the plot, so that eventually, as  $|a|$  becomes very large, the conditional mean plot approaches that of the line  $\mu_t = \theta_{t-1}$ . Increasing the absolute value of  $a$  also brings in the dashed lines representing  $\mu_t \pm 1/\sqrt{\kappa_t}$  so that they turn in the same direction as  $\mu_t$ .

Consider now starting at  $|a| = b$  and increasing the value of  $b$ , as in the bottom left plot of Figure 6.8, where  $b = 3.5$ , and in (d), where  $b = 20$ . It can be seen that the solid line representing  $\mu_t$  is no longer monotonic, and that as  $b$  increases, the periodicity in the curvature gets closer to  $2\pi$ , whilst the actual conditional mean is constrained to an ever smaller interval. For  $b > |a|$ , similar behaviour to that of the Sine model is observed for the Cosine model and is apparent in the bottom two plots of Figure 6.8. Firstly, except for the turning points, there are two possible values of  $\theta_{t-1}$  that lead to any single value of  $\mu_t$ . Secondly, moving away from the overall mean direction (zero), in either direction on the  $\theta_{t-1}$  axis but past the turning point, the further  $\theta_{t-1}$  is from the overall mean, the closer

$\mu_t$  is to this value. These are features that perhaps one might not expect to observe in an actual data set. In order to further investigate these features, an expression is obtained for the probability that any particular  $\theta$  value occurs further away from  $\mu = 0$  than the turning points. Since

$$\tan \mu_t = \frac{a \sin \theta_{t-1}}{a \cos \theta_{t-1} + b}$$

we have

$$\frac{d\mu_t}{d\theta_{t-1}} \sec^2 \mu_t = \frac{a \cos \theta_{t-1} (a \cos \theta_{t-1} + b) + a^2 \sin^2 \theta_{t-1}}{(a \cos \theta_{t-1} + b)^2}.$$

It then follows that

$$\frac{d\mu_t}{d\theta_{t-1}} = \frac{a(a + b \cos \theta_{t-1})}{a^2 + b^2 + 2ab \cos \theta_{t-1}}. \quad (6.21)$$

The turning points are then found as the solutions to  $a + b \cos \theta_{t-1} = 0$ , ie.  $\theta_{t-1} = \cos^{-1}(-a/b)$  and  $\theta_{t-1} = -\cos^{-1}(-a/b)$ . For such turning points to exist, therefore, we require  $a < b$ , as suggested by the plots of Figure 6.8. Moreover, substituting this solution back into the denominator of (6.21) gives  $(b - a)(b + a)$ , which is equal to zero if  $a = b$ , again highlighting  $a = b$  as an irregular case.

We can now obtain the probability, given  $a$  and  $b$ , that  $\theta_{t-1}$  falls outside the turning points, as

$$p = 2 \int_{\cos(-a/b)}^{\pi} \Pi(\theta) d\theta$$

where  $\Pi(\theta)$  is given by Equation (6.17). Table 6.3 displays these probabilities for various values of  $a$  and  $b$ . As can be seen from the table,  $a$  and  $b$  have to be quite small before the probability exceeds 0.05. Since the concentration parameter associated with  $\theta_t$  is  $\kappa_t = \sqrt{[a^2 + b^2 + 2ab \cos(\theta_{t-1} - \mu)]}$ , the maximum  $\kappa_t$  can attain is  $(a + b)$ . From the evidence in Table 6.3 therefore, the concentration parameter has to be quite small for the probability that any particular  $\theta$  value falls beyond the turning points. It could therefore be argued that the feature observed in Figure 6.8 (c) and 6.8 (d) should not have too much of an impact on the model unless  $a$  and  $b$  are very small.

We now consider the effect of switching the sign of  $a$  on the plots of Figure 6.8. With



$a$	3	1.2	0.5	0.4	0.4	0.3
$b$	3.5	1.6	1.5	1.2	0.6	0.5
$p$	0.00	0.01	0.05	0.09	0.13	0.17

Table 6.3: Probabilities, for various  $a$  and  $b$  values, that any particular  $\theta$  value will fall in the tail beyond the turning points.

$\mu = 0$  we have  $\mu_t = \tan^{-1}(a \sin \theta_{t-1}, a \cos \theta_{t-1} + b)$ . If  $a$  is switched to  $-a$  then the conditional mean becomes  $\mu_t^*$ , say, where

$$\begin{aligned}\mu_t^* &= \tan^{-1}(-a \sin \theta_{t-1}, -a \cos \theta_{t-1} + b) \\ &= \tan^{-1}(a \sin(\theta_{t-1} - \pi), a \cos(\theta_{t-1} - \pi) + b).\end{aligned}$$

That is, the effect of switching the sign of  $a$  is to shift the plots of Figure 6.8 along the  $\theta_{t-1}$  axis by  $\pi$ .

Having considered the deterministic component of the model, we next turn our attention to data simulation, by combining the deterministic and random components of the model.

### 6.3.4 Data simulation

As for the Sine model, we will consider simulation of data with  $\mu = 0$ . Again similarly to the Sine model, we simulate an initial value  $\theta_1$  using steps 1–3 of the algorithm described in Section 2.4 adapted to the Cosine model and as described by Mardia et al. (2007b). The conditional distribution (6.11) is then used to simulate the values  $\theta_2, \dots, \theta_n$  sequentially.

Figure 6.9 shows three data sets, each of length 100, for the parameter combinations  $(a, b) = (3, 3.5)$ ,  $(2.1, 2)$  and  $(10, 2)$ . Since  $a > 0$  for all of these parameter combinations, the equilibrium distribution for each is unimodal and symmetric about  $\mu = 0$ . For the parameter values  $(a, b) = (3, 3.5)$  and  $(a, b) = (2.1, 2)$ , we plotted in Figure 6.8  $\mu_t \pm$

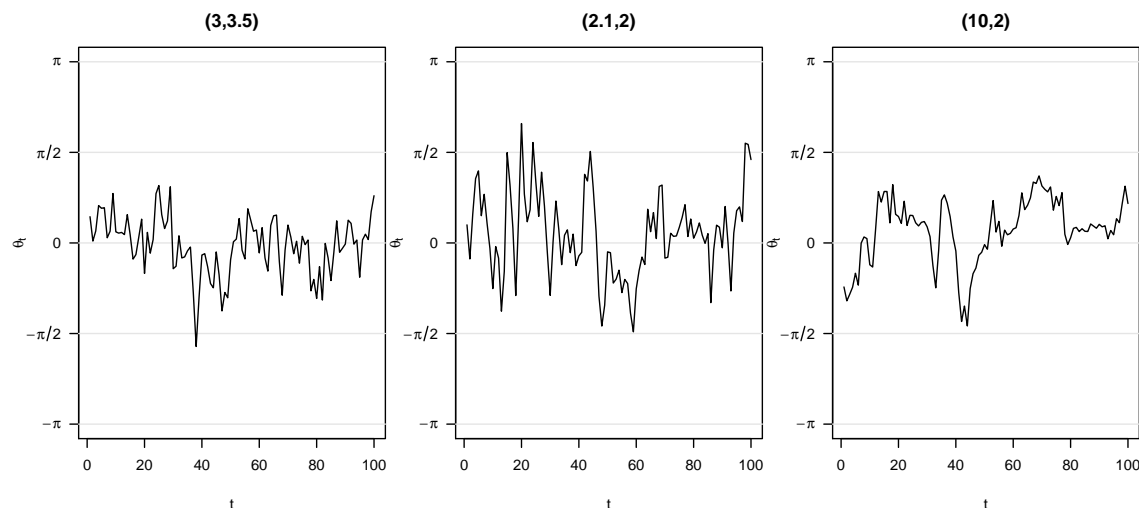


Figure 6.9: Simulated unimodal Cosine time series data for the values of  $(a, b)$  given above each plot.

$1/\sqrt{\kappa_t}$ , and the simulated data are best compared with these plots in mind. For the plots in Figure 6.8, the lines for  $\mu_t \pm 1/\sqrt{\kappa_t}$  are slightly further from the line  $\mu_t$  for  $(a, b) = (2.1, 2)$  than for  $(a, b) = (3, 3.5)$ . This is reflected in the simulated data, the plots of which show the data based on the former parameter combination to be slightly more dispersed overall. As predicted in Section 6.3.3, the fact that  $\mu_t$  is not a monotonic function of  $\theta_{t-1}$  does not affect the simulated data for the data set for which  $b > a$ , since the range of values is not wide enough for this to be so. Looking again at the plots in Figure 6.8, the gradient of the slope of the line  $\mu_t$  for  $(a, b) = (2.1, 2)$  is slightly greater than that for  $(a, b) = (3, 3.5)$  (for the approximately straight middle line segment). This being the case, the value of  $\mu_t$  is rotated less in the direction of  $\mu = 0$  so that  $\mu_t$  is closer to  $\theta_{t-1}$ . The effect is difficult to discern for the simulated data in this case since there is not a great difference in the slopes for  $(a, b) = (2.1, 2)$  and  $(a, b) = (3, 3.5)$ . It is perhaps more readily observable in the third plot in Figure 6.9 in which  $(a, b) = (10, 2)$ . As discussed in Section 6.3.3, if we increase the value of  $a$  whilst keeping  $b$  fixed then the conditional mean plot approaches that of the line  $\mu_t = \theta_{t-1}$ . Moreover, increasing the value of  $a$  increases  $\kappa_t$  and thus leads to

a more highly concentrated series. The combined effects on the deterministic and random component of the model are observable in the third plot of Figure 6.9, which exhibits slower changes in the values of  $\theta_t$  and periods of higher concentration than those apparent in the other two plots of Figure 6.9.

Next we consider two bimodal data sets and consider more generally the interpretation and effects of parameters in cases when the equilibrium distribution of the process is bimodal. As described in Section 6.3.2, this occurs if and only if  $a < 0$  and

$$A_1(|a + b|) + \frac{|a + b|}{a} > 0. \quad (6.22)$$

Unfortunately this is a more complicated condition than for the Sine model, for which it was observed that if  $\lambda$  is not approximately equal to  $\kappa$  and  $\kappa$  is not very small,  $|\lambda| > \kappa$  tends to lead to a bimodal equilibrium distribution. No analogous statement can be made about the Cosine time series model. As an example,  $(a, b) = (-3, 1)$  gives a bimodal equilibrium distribution, but then (6.22) implies that  $(a, b) = (-3, 5)$  also gives a bimodal equilibrium distribution. Based on (6.22) a necessary condition for bimodality is given by  $b \leq 2|a|$ .

Figure 6.10 shows plots based on the two parameter configurations  $(a, b) = (-3, 1)$  and  $(a, b) = (-16, 10)$ . The left hand plots show time series plots for simulated data sets of length 50. The middle plots show dashed lines connecting  $\theta_t$  for  $t = 1, 3, 5, \dots, 49$  and solid lines connecting  $\theta_t$  for even  $t$ . The right hand plots show  $\mu_t$  and  $\mu_t \pm 1/\sqrt{\kappa_t}$ . For display purposes, the  $y$ -axis for the right hand plots range from zero to  $2\pi$  in order that the plot of  $\mu_t$  is continuous. These right hand plots give an insight into the behaviour observed for the simulated data. In particular, the plots of  $\mu_t$  show that the effect of the deterministic component of the model is to rotate  $\theta_{t-1} - \pi$  negatively towards zero if  $\theta_{t-1} - \pi \in (0, \pi)$  and positively towards zero if  $\theta_{t-1} - \pi \in (-\pi, 0)$ . As can be seen from the closeness of the lines  $\mu_t \pm 1/\sqrt{\kappa_t}$  to  $\mu_t$  when  $(a, b) = (-16, 10)$ , the random components  $\epsilon_t$  of the model should be small. Indeed, the range of  $\kappa_t$  given  $\theta_{t-1}$  for the data set shown is

$(6.97, 24.75)$ . These effects are visible in the middle plot for  $(a, b) = (-16, 10)$ , in which it can be seen that the line for  $\theta_t$  for even values of  $t$  approximates  $\theta_t - \pi$  for odd  $t$ .

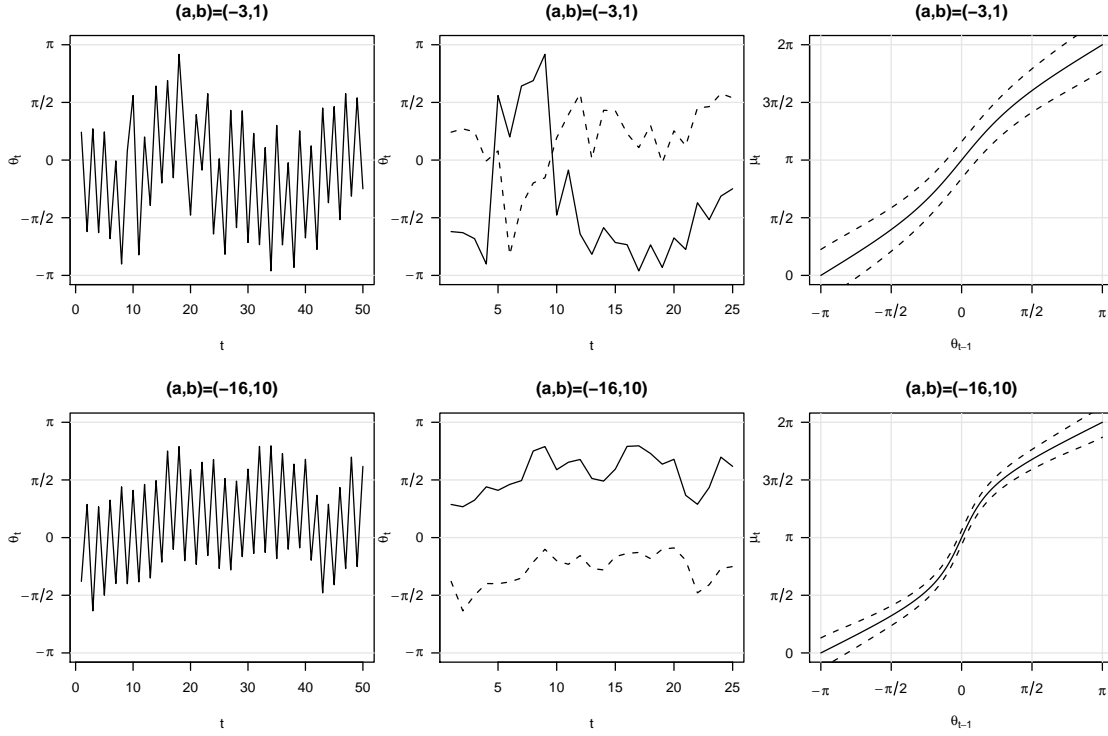


Figure 6.10: Simulated bimodal Cosine time series data. Left: Time series plots. Middle:  $\theta_t$  for  $t$  even (solid) and  $t$  odd (dashed). Right:  $\mu_t \pm 1/\sqrt{\kappa_t}$  for the given  $(a, b)$  values

By comparing the behaviour of the data for  $(a, b) = (-3, 1)$  with the behaviour described for  $(a, b) = (-16, 10)$ , we can assess the effect of changing  $a$  and  $b$ . Again we see that  $\theta_t$  for even  $t$  is approximately equal to  $\theta_t - \pi$  for odd  $t$  (this is more clearly visible if we subtract  $2\pi$  from the values of  $\theta_t$  for even  $t \in [4, 10]$ ). The approximation, however, is not as close as when  $(a, b) = (-16, 10)$ , due to the fact that the  $\kappa_t$  values are smaller, as indicated by the lines  $\mu_t \pm 1/\sqrt{\kappa_t}$  on the top right plot. For the data set for which  $(a, b) = (-3, 1)$ , the range of observed values of  $\kappa_t$  given  $\theta_{t-1}$  was  $(2.00, 3.99)$ . Due to the ratio  $|a|/b$  being larger when  $(a, b) = (-3, 1)$  than when  $(a, b) = (-16, 10)$ ,  $\mu_t$  is closer to  $\theta_{t-1} - \pi$  for this data set. The smaller values of  $|a|$  and  $b$  however lead to more

dispersion contributed from the random component of the model.

Having successfully simulated data and interpreted parameter values for the Cosine time series model, we next investigate the likelihood function and parameter estimation for the model.

### 6.3.5 Likelihood function and parameter estimation

The likelihood function for the model can be obtained from the equation

$$L = L(\theta_1, \dots, \theta_n; a, b, \mu) = \Pi(\theta_1) \prod_{t=2}^n f(\theta_t | \theta_{t-1}). \quad (6.23)$$

Using equations (6.11) and (6.17) we have

$$\begin{aligned} L &= C I_0[\sqrt{(a^2 + b^2 + 2ab \cos(\theta_1 - \mu))}] e^{b \cos(\theta_1 - \mu)} \times \\ &\quad \prod_{t=2}^n \frac{e^{a \cos(\theta_t - \theta_{t-1}) + b \cos(\theta_t - \mu)}}{I_0[\sqrt{(a^2 + b^2 + 2ab \cos(\theta_{t-1} - \mu))}]} \\ &= C \prod_{t=1}^n e^{b \cos(\theta_t - \mu)} \prod_{t=2}^n e^{a \cos(\theta_t - \theta_{t-1})} / \prod_{t=3}^n I_0[\sqrt{(a^2 + b^2 + 2ab \cos(\theta_{t-1} - \mu))}] \\ &= \frac{C \exp\{b \sum_{t=1}^n \cos(\theta_t - \mu) + a \sum_{t=2}^n \cos(\theta_t - \theta_{t-1})\}}{\prod_{t=3}^n I_0[\sqrt{(a^2 + b^2 + 2ab \cos(\theta_{t-1} - \mu))}]} \end{aligned} \quad (6.24)$$

The log likelihood  $l = \log L$  is therefore given by

$$\begin{aligned} l &= \log C + b \sum_{t=1}^n \cos(\theta_t - \mu) + a \sum_{t=2}^n \cos(\theta_t - \theta_{t-1}) \\ &\quad - \sum_{t=3}^n \log I_0[\sqrt{(a^2 + b^2 + 2ab \cos(\theta_{t-1} - \mu))}]. \end{aligned} \quad (6.25)$$

We next investigate the behaviour of this loglikelihood function and discuss parameter estimation based on it.

In order to investigate the estimation of parameters based on different values of  $a$  and  $b$  ( $\mu$  will be set to zero in each case),  $m = 100$  data sets, each of size  $n = 200$ , for each

possible  $(a_0, b_0)$  pairing from  $a_0 \in \{-3, -1, 1, 3, 5\}$  and  $b_0 \in \{1, 2\}$ , are simulated. The `nlm` function in R is then used to minimise the function  $-l$  in each case, with  $a = 0, b = 0$  and  $\mu = \bar{\theta}$  used as initial values for the estimation process, where  $\bar{\theta}$  is the directional mean of the angles in the data set.

Table 6.4 displays the mean values (and standard deviations) of the maximum likelihood estimates  $\hat{a}$ ,  $\hat{b}$  and  $\hat{\mu}$  of  $a, b$  and  $\mu$  respectively. Each entry in the table is based on  $m = 100$  data sets. For  $\mu$  the elements in the table are, as usual, the mean direction of estimates and the circular variance  $1 - \bar{R}$  of the estimates, where  $\bar{R}$  is the mean resultant length of the estimates. As can be seen, the standard deviation of the estimates is quite small, and the parameter estimates are generally very close to the true parameter values. There does, however, appear to be a tendency for  $a$  and  $b$  to be slightly overestimated in absolute value.

$\hat{a} \searrow$		$a_0$									
		-3		-1		1		3		5	
$b_0$	1	-3.073	(0.285)	-1.013	(0.142)	1.001	(0.141)	2.983	(0.315)	5.009	(0.547)
	2	-3.089	(0.292)	-1.045	(0.177)	0.945	(0.211)	3.050	(0.461)	5.014	(0.614)
$\hat{b} \searrow$											
		-3		-1		1		3		5	
$b_0$	1	1.034	(0.183)	1.027	(0.164)	1.007	(0.163)	1.058	(0.244)	1.199	(0.355)
	2	2.081	(0.251)	2.037	(0.215)	2.083	(0.279)	2.148	(0.335)	2.138	(0.422)
$\hat{\mu} \searrow$											
		-3		-1		1		3		5	
$b_0$	1	0.019	(0.015)	-0.020	(0.006)	0.010	(0.007)	-0.010	(0.015)	0.009	(0.018)
	2	-0.011	(0.003)	-0.008	(0.001)	-0.002	(0.002)	-0.007	(0.003)	-0.008	(0.004)

Table 6.4: Mean values (and standard deviations) of estimates of  $a, b$  and  $\mu$  for 10 combinations of  $(a_0, b_0)$  and with  $\mu_0 = 0$ , each based on 100 simulated data sets of 200 observations. Estimates obtained using the `nlm` function in R.

In order to investigate the behaviour of the loglikelihood function in the vicinity of the true parameter values, data are simulated and various plots produced. Figure 6.11 shows profile loglikelihood plots based on two data sets, each of size  $n = 200$ , in which  $(a_0, b_0) = (2, 1)$  for data set one and  $(a_0, b_0) = (-3, 1)$  for data set two. For both data sets,

$\mu_0 = 0$ . Figure 6.11(A) and (D) show plots of  $l(\boldsymbol{\theta}; a, b_0, \mu_0)$  versus  $a$  for  $a \in [a_0 - 2, a_0 + 2]$  for data sets one and two respectively, where  $\boldsymbol{\theta}$  is the simulated data set for the particular values of  $a_0$  and  $b_0$ . (C) and (E) show plots of  $l(\boldsymbol{\theta}; a_0, b, \mu_0)$  versus  $b$  for  $b$  between 0 and 3. Finally, (C) and (F) show contoured image plots of  $l(\boldsymbol{\theta}; a, b, \mu_0)$  for  $a \in [a_0 - 2, a_0 + 2]$  and  $b \in [0, 3]$ .

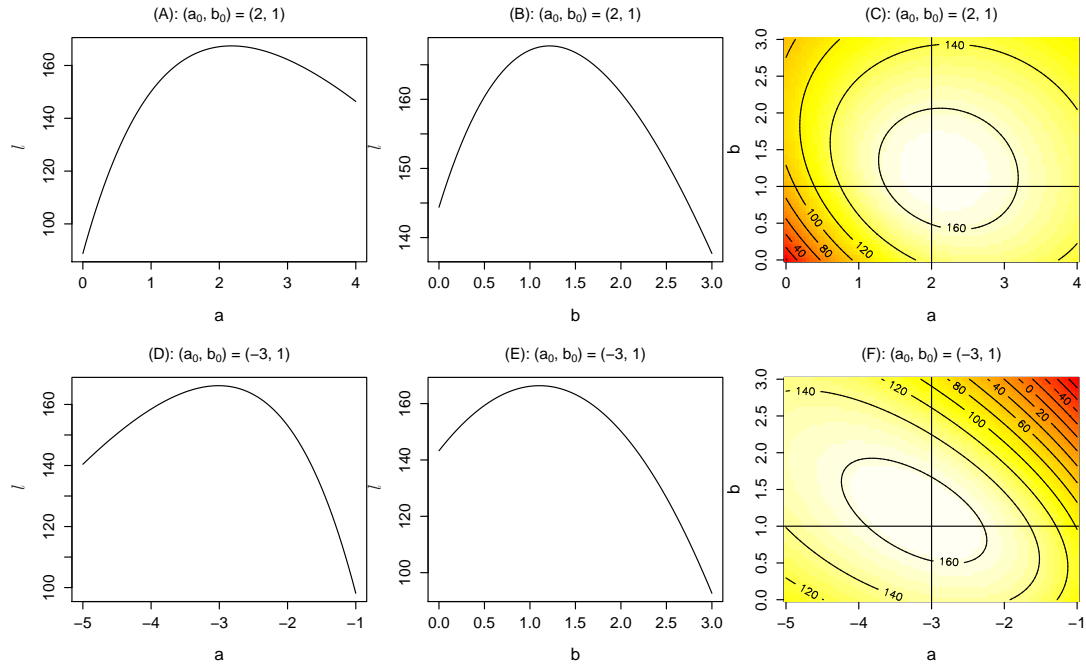


Figure 6.11: Profile loglikelihood plots. (A) and (D):  $l(\boldsymbol{\theta}; a, b_0, \mu_0)$  for  $a \in [a_0 - 2, a_0 + 2]$ ; (B) and (E):  $l(\boldsymbol{\theta}; a_0, b, \mu_0)$  for  $b \in [0, 3]$ ; (C) and (F)  $l(\boldsymbol{\theta}; a, b, \mu_0)$  for  $a \in [a_0 - 2, a_0 + 2]$  and  $b \in [0, 3]$ .

It can be seen from Figure 6.11 that the profile of the log likelihood function, taken with respect to either  $a$  or  $b$ , is well behaved at the true value of the other two parameters, inasmuch as a unique maximum is clearly visible. Moreover, with  $\mu$  fixed at  $\mu_0$ , the contoured image plot clearly highlights an area in which the maximum occurs, which area is around the true values of  $a$  and  $b$ . Worthy of note is that, based on the plots of Figure 6.11, the maximum of  $l$  appears to be at values of  $a$  and  $b$  that are slightly greater

in absolute value than the true  $a_0$  and  $b_0$ , a feature reminiscent of the maximum likelihood estimates given in Table 6.4.

### 6.3.6 Application to protein data

Having tested an estimation method for the model, it is now fitted to the 249  $\phi$  and  $\psi$  angles of the same protein, triosephosphate isomerase, that was used for the Sine model. MLEs of model parameters (and approximate standard errors obtained via `nlm`) are calculated and displayed in Table 6.5. As a measure of goodness-of-fit, data are simulated based on the MLEs obtained and the resulting data compared with the original. MLEs for the simulated data are also given in Table 6.5 and are seen to be reasonably similar to the MLEs for the original data.

	$\phi$	$\phi^{\text{sim}}$	$\psi$	$\psi^{\text{sim}}$
$\hat{\mu}$	-1.425 (0.056)	-1.400 (0.057)	-0.357 (0.304)	0.079 (0.275)
$\hat{a}$	1.113 (0.218)	1.215 (0.225)	1.345 (0.123)	1.336 (0.125)
$\hat{b}$	2.226 (0.239)	2.230 (0.255)	0.338 (0.117)	0.392 (0.113)

Table 6.5: Estimates (and approximate standard errors) of Cosine time series parameters for protein and simulated data.

Comparison of the simulated data set with the original protein data can be made from Figure 6.12, which shows circular and time series plots for the protein data and simulated data of both  $\phi$  and  $\psi$  angles. Similar comparisons between the protein data and the simulated data can be made from Figure 6.12 as were made from Figure 6.4 in Section 6.2.5 for the Sine model. As for that model, the considerable discrepancies between the protein data and the simulated data would indicate that the model is a poor fit.

In order to further investigate the goodness-of-fit of the Cosine model to the protein data, plots analogous to those of Figures 6.5 and 6.6 for the Sine model are produced for the



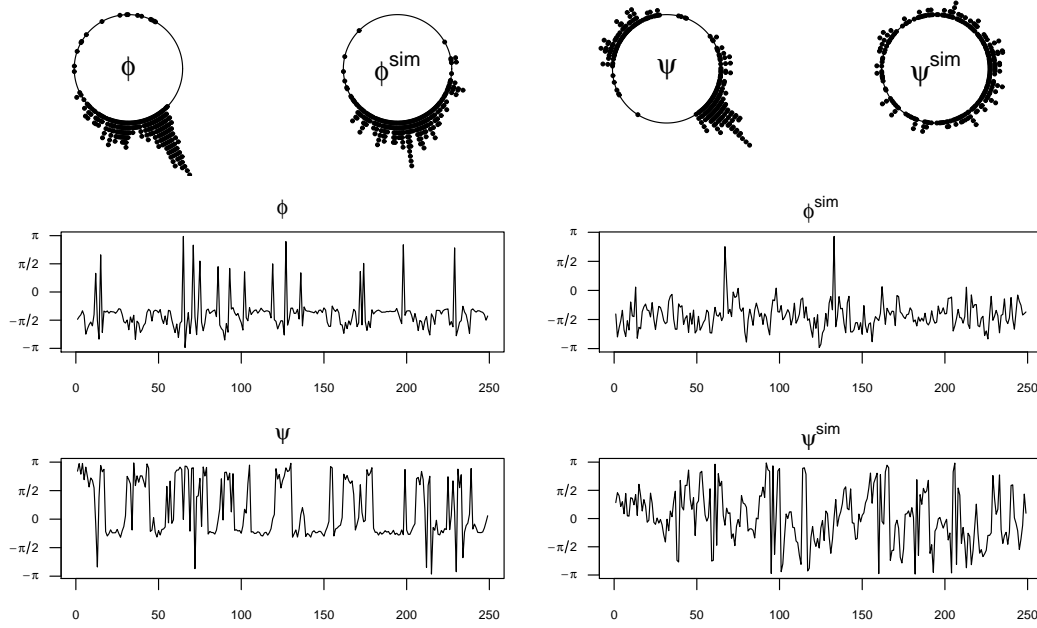


Figure 6.12:  $\phi$  and  $\psi$  conformational angles for a protein data set, and simulated  $\phi$  and  $\psi$  values based on maximum likelihood estimates for the true data using the Cosine time series model.

Cosine model and are displayed in Figures 6.13 and 6.14 respectively. A description of the methods used to obtain these plots and what they represent was given in Section 6.2.5.

Based on Figure 6.13 we see that again there is some evidence to suggest that the model is not such a good fit to the  $\phi$  values of the protein data, although the approximation of the smoothed line to the line for  $\mu_t$  is not too bad. For the simulated  $\phi$  series, the potential adverse effect of an outlier or influential observation can be seen. The solid red line represents the smoothed values of all the data, whilst the dashed red line represents the smoothed values omitting the data value in the top left of the plot. Even for the latter, similar behaviour on the right hand side of the plot is observed as for the original protein data, namely that the smoothed line dips away from the line  $\mu_t$ . Bearing this in mind,

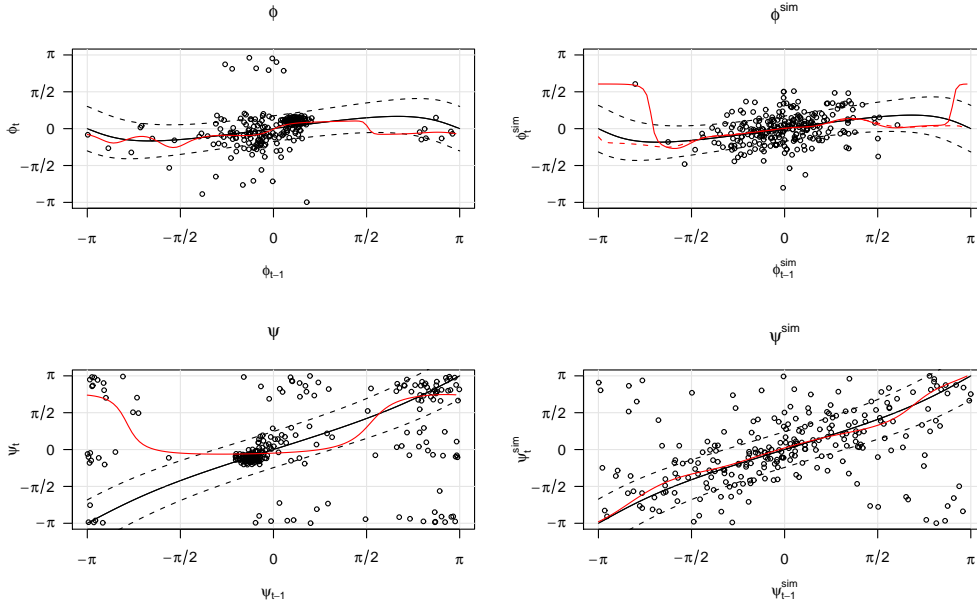


Figure 6.13: Plots of  $\theta_t$  versus  $\theta_{t-1}$ , a smoothed version (red line) and  $\hat{\mu}_t \pm 1/\sqrt{\hat{\kappa}_t}$ , where  $\theta$  is one of  $\phi, \phi^{\text{sim}}, \psi$  or  $\psi^{\text{sim}}$ .

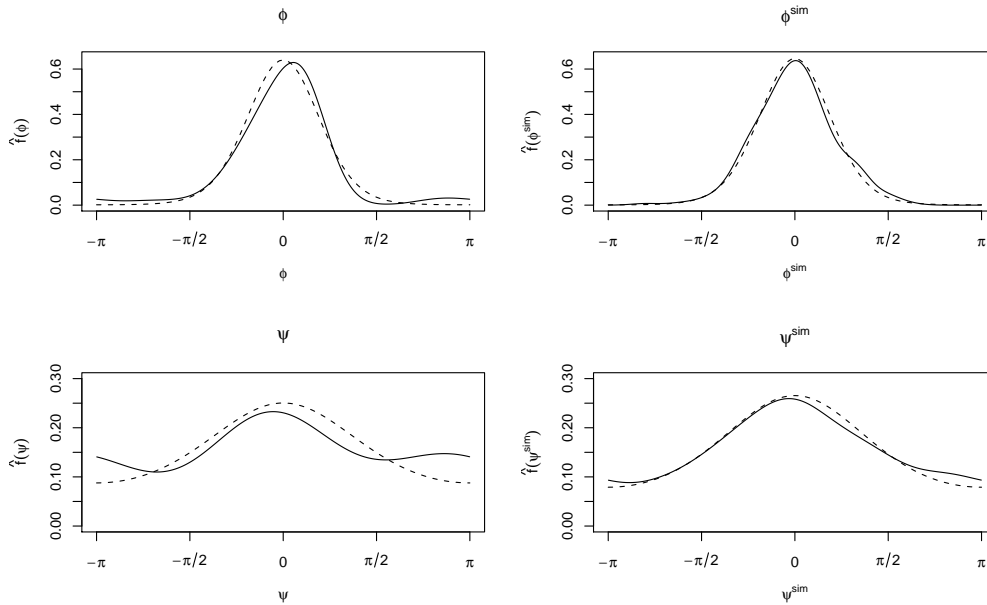


Figure 6.14: Marginal kde (solid) and pdf (dashed) based on data and MLEs respectively.

there is little evidence from these plots to suggest that the model is a poor fit to the  $\phi$  values of the protein data.

There is clearly very strong evidence, on the other hand, that the model is a poor fit to the  $\psi$  protein data. In particular, the maximum likelihood estimates for these data show that a unimodal model has been fitted ( $\hat{a} > 0$ ) to a data set that is clearly bimodal, and the smoothed line is nothing like the line for  $\mu_t$ .

By comparing the plots of Figure 6.13 with those of Figure 6.5 we can directly compare the goodness-of-fit of the Cosine model with that of the Sine model. In particular, since the plots of  $\theta_t$  versus  $\theta_{t-1}$  for the protein data are the same save that a different MLE of  $\mu$  has been subtracted in each case, the smoothed red lines should have the same features for a given data set between models (by construction the choice of the smoothing parameter  $h$  is invariant under rotation of the data). Comparison of Figures 6.5 and 6.13 shows this indeed to be the case. Comparison of Figures 6.5 and 6.13 shows the Sine model to be a far superior fit to the  $\psi$  values of the protein data, whilst the plots for the  $\phi$  values not only show the goodness-of-fit of the two models to be comparable for these data, but also the models themselves to be similar, based on the lines for  $\mu_t$  and  $\mu_t \pm 1/\sqrt{\kappa_t}$ .

Moving on to Figure 6.14, which compares kernel density estimates (solid lines) with the density of the equilibrium distribution (dashed lines) based on the parameter estimates of the fitted models, we clearly have further evidence to suggest that the Cosine time series model is a poor fit to the  $\psi$  protein data, since the kernel density estimate is not unimodal. None of the remaining plots give strong evidence to suggest a poorly fitting model, although the kernel density estimate for the  $\phi$  protein data does appear slightly skewed.

### 6.3.7 Conclusions

In this section we have investigated a circular time series model that can be seen as the adaptation of the bivariate Cosine density to a time series context through use of the associated conditional distribution. The behaviour of the deterministic and random components of the model have been investigated and interpreted in terms of simulated data. The likelihood function of the model has been investigated and the accuracy of parameter estimation using the R function `nlm` has been studied. The latter has revealed a tendency to slightly overestimate the absolute values of the parameters  $a$  and  $b$ . Further investigation into the behaviour of the loglikelihood function in the vicinity of the MLEs has shown this to perhaps be a feature of the likelihood function itself rather than of the estimation procedure.

Fitting the model to the  $\phi$  and  $\psi$  values of a protein data set has indicated a very poor fit for the  $\psi$  angles, and an indication that the model may not be a good fit for the  $\phi$  series either. Comparison of the goodness-of-fit plots established and discussed in Section 6.2.5 for the Sine model with the resulting plots for the fitted Cosine model indicate the former to be a far superior fit for the  $\psi$  values of the protein data, whilst comparison of the plots for the  $\phi$  series indicates both that the goodness-of-fit and the Sine and Cosine models themselves are similar for the fitted parameter values.

## 6.4 Further work

Sections 6.2.6 and 6.3.7 summarised the conclusions drawn from analysis of the Sine and Cosine models, and a comparison of the two in terms of goodness-of-fit to a specific data set was made in Section 6.3.6. In this section we discuss possible extensions of the two models and potential further work.

An obvious way in which both the Sine and Cosine time series models could be extended

is with the inclusion of higher order terms. Starting with the former, the simplest way to extend the Sine time series to an AR( $p$ ) model is by defining  $\theta_t$  given  $(\theta_{t-1}, \dots, \theta_{t-p})$  as  $\theta_t | (\theta_{t-1}, \dots, \theta_{t-p}) \sim M(\mu_t, \kappa_t)$ , where

$$\mu_t = \tan^{-1} \left\{ \sum_{j=1}^p \lambda_j \sin(\theta_{t-j} - \mu), \kappa \right\}, \quad (6.26)$$

$$\kappa_t = \left\{ \kappa^2 + \left[ \sum_{j=1}^p \lambda_j \sin(\theta_{t-j} - \mu) \right]^2 \right\}^{1/2}$$

and, as usual,  $\tan^{-1}(q, p)$  is the angle between the positive  $x$ -axis and the vector  $(p, q)$ . From the expression for  $\mu_t$  in Equation (6.26) we see that, as for the AR(1) model,  $\mu_t$  is constrained to lie in the interval  $[\pi/2, -\pi/2]$ .

For the Cosine AR( $p$ ) model, the conditional density specifying the model can be given either in terms of the vectorial form (Equation (1.9)) of the von Mises distribution or using the angular form (Equation (1.8)). The former is given by  $\mathbf{x}_t | (\mathbf{x}_{t-1}, \dots, \mathbf{x}_{t-p}) \sim VM(\boldsymbol{\alpha})$ , where

$$\boldsymbol{\alpha} = \sum_{j=1}^p a_j \mathbf{x}_{t-j}^T + b \mathbf{e}^T$$

and  $\mathbf{x}_{t-j} = (\cos \theta_{t-j}, \sin \theta_{t-j})^T$  and  $\mathbf{e} = (\cos \mu, \sin \mu)^T$  are unit vectors. The alternative specification is  $\theta_t | (\theta_{t-1}, \dots, \theta_{t-p}) \sim M(\tan^{-1}(s, c), \sqrt{(s^2 + c^2)})$ , where

$$s = \sum_{j=1}^p a_j \sin \theta_{t-j} + b \sin \mu \quad \text{and} \quad c = \sum_{j=1}^p a_j \cos \theta_{t-j} + b \cos \mu.$$

Another way in which the Cosine model could be extended is as suggested by Kent and Mardia (2002, p. 288). They propose treating the  $\{\mathbf{x}_t\}$  process as an unobserved signal and combining it with an observation process

$$\mathbf{y}_t | \mathbf{x}_t \sim VM(\kappa \mathbf{x}_t), \quad \kappa > 0.$$

A Kalman filter approach is then suggested as a way of recovering the signal from the observations.



# Chapter 7

## A Möbius time series model

### 7.1 Introduction

In this chapter a circular regression model due to Downs and Mardia (2002) is adapted to a time series context. The distribution of  $\theta_t$  given  $\theta_{t-1}$  is modeled using a particular von Mises distribution. In Section 7.2 we outline the regression model studied by Downs and Mardia (2002) and adapt it to a time series setting. The likelihood function (conditional on the first observation) is derived in Section 7.3 and potential methods of parameter estimation are introduced. For certain values of one of the parameters in the model ( $\omega$ ) the time series model takes a special form. These special cases are considered in Section 7.4. The deterministic component of the model and its behaviour in terms of the values of parameters is discussed in Section 7.5, whilst Section 7.6 focuses on simulating data from the model. The properties of the simulated data are related to the observations made in Section 7.5. Time-reversibility is an important part of time series modeling, and we investigate whether the Möbius model has this property in Section 7.7. It will be seen in Section 7.8 that the equilibrium distribution of the process is difficult to obtain analytically. We therefore use a numerical approach in this section in order to obtain the

equilibrium distribution. The results are compared with the von Mises distribution and with a sample from the equilibrium distribution. An analytic approach is used to derive the equilibrium distribution under high concentration. Estimation of the parameters of the model is investigated in Section 7.9, with the focus on two approaches: a grid search method and use of the R function `nlm`. In Section 7.10 we apply the model to the  $\phi$  and  $\psi$  angles of a particular protein and investigate the goodness-of-fit of the model. Finally the conclusions are summarised in Section 7.11, in which section we compare the three time series models that have been studied, namely the Sine model, the Cosine model and the Möbius model. Possible extensions of the work are also outlined in Section 7.11.

## 7.2 The model

The deterministic component of the regression model studied by Downs and Mardia (2002) links the dependent angular variable  $v$  to the independent angular variable  $u$  via

$$\tan \frac{1}{2}(v - \beta) = \omega \tan \frac{1}{2}(u - \alpha), \quad (7.1)$$

where  $\omega \in [-1, 1]$  is a slope parameter and  $-\pi \leq \alpha, \beta < \pi$  are angular location parameters. Equation (7.1) gives

$$v = \beta + 2 \tan^{-1} \{ \omega \tan \frac{1}{2}(u - \alpha) \}. \quad (7.2)$$

Since the application of (7.2) will be to time series data, we replace the dependent angle  $v$  with  $\theta_t$  and  $u$  with  $\theta_{t-1}$ ,  $t = 2, \dots, n$ . These substitutions also suggest the use of a single location parameter,  $\alpha = \beta$  say, giving

$$\tan \frac{1}{2}(\theta_t - \alpha) = \omega \tan \frac{1}{2}(\theta_{t-1} - \alpha) \quad (7.3)$$

and

$$\theta_t = \alpha + 2 \tan^{-1} \{ \omega \tan \frac{1}{2}(\theta_{t-1} - \alpha) \} \quad (7.4)$$



as analogues of (7.1) and (7.2).

For the time series model (7.4) in Section 7.2 it is assumed that  $\theta_t|\theta_{t-1}$  has a Von Mises distributon such that

$$\theta_t|\theta_{t-1} \sim M\left(\alpha + 2 \tan^{-1}\left\{\omega \tan \frac{1}{2}(\theta_{t-1} - \alpha)\right\}, \kappa\right). \quad (7.5)$$

In this case the time series model becomes

$$\theta_t = \alpha + 2 \tan^{-1}\left\{\omega \tan \frac{1}{2}(\theta_{t-1} - \alpha)\right\} + \epsilon_t \quad (7.6)$$

where  $\epsilon_t \sim M(0, \kappa)$ . We will refer to the mean direction of the conditional distribution of  $\theta_t$  given  $\theta_{t-1}$  as  $\mu_t$ . That is

$$\mu_t = \alpha + 2 \tan^{-1}\left\{\omega \tan \frac{1}{2}(\theta_{t-1} - \alpha)\right\}. \quad (7.7)$$

In what follows, all angles and their sums and differences are taken to be in the half-open interval  $[-\pi, \pi)$ , otherwise numerical errors will result (Downs and Mardia, 2002).

The model (7.6) is seen to be of the form of the linked autoregressive process discussed by Fisher and Lee (1994), for which the mean direction of the conditional distribution of  $\theta_t$  given  $(\theta_{t-1}, \dots, \theta_{t-p})$  has the form of Equation (5.4). Specifically, with  $p = 1$ ,  $g(\cdot) = 2 \tan^{-1}(\cdot)$ ,  $\omega_1 = \omega$  and  $\mu = \alpha$ , Equation (5.4) gives the mean direction of the distribution in (7.5).

The value of  $\omega$  is restricted to the interval  $[-1, 1]$  so that  $\alpha$  is uniquely identifiable. To see this, consider subtracting  $\pi$  from the value of  $\alpha$  in Equation (7.3). Then

$$\tan \frac{1}{2}(\theta_t - \alpha + \pi) = \omega \tan \frac{1}{2}(\theta_{t-1} - \alpha + \pi) \quad (7.8)$$

Now since  $\tan \frac{1}{2}(\phi + \pi) = -\cot \frac{1}{2}\phi$ , (7.8) is equivalent to

$$-\cot \frac{1}{2}(\theta_t - \alpha) = -\omega \cot \frac{1}{2}(\theta_{t-1} - \alpha)$$

which gives

$$\tan \frac{1}{2}(\theta_t - \alpha) = \frac{1}{\omega} \tan \frac{1}{2}(\theta_{t-1} - \alpha). \quad (7.9)$$

The equivalence of (7.8) and (7.9) shows that, if  $\omega$  were not restricted to the interval  $[-1, 1]$ , and if  $(\hat{\alpha}, \hat{\omega})$  is a solution to (7.3), then  $(\hat{\alpha} - \pi, \frac{1}{\hat{\omega}})$  is an equivalent solution.

### 7.3 Likelihood function

Since

$$\theta_t | \theta_{t-1} \sim M\left(\alpha + 2 \tan^{-1}\left\{\omega \tan \frac{1}{2}(\theta_{t-1} - \alpha)\right\}, \kappa\right)$$

we have

$$f(\theta_t | \theta_{t-1}) = \{2\pi I_0(\kappa)\}^{-1} \exp\{\kappa \cos(\theta_t - \mu_t)\} \quad (7.10)$$

where  $\mu_t$  is given by Equation (7.7)

Now the conditional pdf of  $\theta_2, \dots, \theta_n$  given  $\theta_1$  is

$$\begin{aligned} f(\theta_2, \dots, \theta_n | \theta_1) &= f(\theta_2 | \theta_1) f(\theta_3, \dots, \theta_n | \theta_1, \theta_2) \\ &= f(\theta_2 | \theta_1) f(\theta_3 | \theta_1, \theta_2) f(\theta_4, \dots, \theta_n | \theta_1, \theta_2, \theta_3) \\ &= \dots \\ &= f(\theta_2 | \theta_1) f(\theta_3 | \theta_1, \theta_2) f(\theta_4 | \theta_1, \theta_2, \theta_3) \dots f(\theta_n | \theta_1, \dots, \theta_{n-1}). \end{aligned}$$

But from Equation (7.6) the value  $\theta_t$  depends only on the previous value  $\theta_{t-1}$ . Therefore

$$f(\theta_t | \theta_1, \dots, \theta_{t-1}) = f(\theta_t | \theta_{t-1}),$$

$\forall t = 2, \dots, n$ . The conditional likelihood is therefore

$$L_C(\alpha, \omega, \kappa) = \{2\pi I_0(\kappa)\}^{-(n-1)} \exp\left\{\kappa \sum_{t=2}^n \cos\left[\theta_t - \alpha - 2 \tan^{-1}\left\{\omega \tan \frac{1}{2}(\theta_{t-1} - \alpha)\right\}\right]\right\}$$

giving the loglikelihood

$$l_C(\alpha, \omega, \kappa) = \text{const.} - (n-1) \log I_0(\kappa) + \kappa \sum_{t=2}^n \cos[\theta_t - \alpha - 2 \tan^{-1}\{\omega \tan \frac{1}{2}(\theta_{t-1} - \alpha)\}].$$

This is maximised with respect to the unknown parameters  $\alpha$  and  $\omega$  by maximising

$$l(\alpha, \omega) = \sum_{t=2}^n \cos[\theta_t - \alpha - 2 \tan^{-1}\{\omega \tan \frac{1}{2}(\theta_{t-1} - \alpha)\}]. \quad (7.11)$$

The maximisation of (7.11) will be investigated, in Section 7.9, in the following ways:

1. The function in (7.11) will be evaluated for various fixed values of  $\alpha$  and  $\omega$ , given a set of data and the parameter values that generated those data.
2. The `nlm` function in R will be used in an attempt to minimise the function  $-l(\alpha, \omega)$  for numerous data sets.

Once  $l(\alpha, \omega)$  has been maximised with respect to  $\alpha$  and  $\omega$ , a profile likelihood approach may be used to obtain a maximum likelihood estimate of  $\kappa$ , by maximising

$$l_C(\hat{\alpha}, \hat{\omega}, \kappa) = \text{const.} - (n-1) \log I_0(\kappa) + \kappa l_{\hat{\alpha}, \hat{\omega}} \quad (7.12)$$

with respect to  $\kappa$ . Differentiating (7.12) with respect to  $\kappa$  and noting that  $d(I_0(\kappa))/d\kappa = I_1(\kappa)$ , the modified Bessel function of the first kind and order one, gives

$$\frac{\partial}{\partial \kappa} [l_C(\hat{\alpha}, \hat{\omega}, \kappa)] = -(n-1) \frac{I_1(\kappa)}{I_0(\kappa)} + l_{\hat{\alpha}, \hat{\omega}},$$

so that  $\hat{\kappa}$  is the solution to

$$\frac{I_1(\hat{\kappa})}{I_0(\hat{\kappa})} = \frac{l_{\hat{\alpha}, \hat{\omega}}}{n-1}.$$

## 7.4 Special values of $\omega$

Before the above estimation methods are investigated, special values of  $\omega$  for which the time series model takes a certain form are considered. If  $\omega = 1$ , then (7.6) becomes

$$\theta_t = \theta_{t-1} + \epsilon_t \quad (7.13)$$

which is a circular random walk. If  $\omega = 0$ , then (7.6) reduces to

$$\theta_t = \alpha + \epsilon_t \quad (7.14)$$

in which case the  $\theta_t$ 's are i.i.d von Mises distributed random circular variables with mean direction  $\alpha$  and concentration parameter  $\kappa$ , and the maximum likelihood estimate of  $\alpha$  is the mean direction of the  $\theta_t$  values,  $t = 2, \dots, n$ ;

$$\hat{\alpha} = \begin{cases} \tan^{-1}(S_{(1)}/C_{(1)}) & \text{if } C_{(1)} > 0 \\ \tan^{-1}(S_{(1)}/C_{(1)}) + \pi \text{sgn}(S_{(1)}) & \text{if } C_{(1)} < 0 \end{cases} \quad (7.15)$$

where  $S_{(1)} = \sum_{t=2}^n \sin \theta_t$  and  $C_{(1)} = \sum_{t=2}^n \cos \theta_t$ . The addition of  $\pi \text{sgn}(S_{(1)})$  (where  $\text{sgn}(\cdot)$  is the sign function) in the case  $C_{(1)} < 0$  ensures that the mle of  $\alpha$  is in the interval  $[-\pi, \pi)$ .

If  $\omega = -1$ , then (7.6) becomes

$$\theta_t = 2\alpha - \theta_{t-1} + \epsilon_t \quad (7.16)$$

and (7.11) becomes

$$\sum_{t=2}^n \cos(\theta_t + \theta_{t-1} - 2\alpha). \quad (7.17)$$

Differentiating (7.17) with respect to  $\alpha$  gives

$$\begin{aligned} \frac{\partial}{\partial \alpha} [l(\alpha, \omega = -1)] &\propto \sum_{t=2}^n \sin(\theta_t + \theta_{t-1} - 2\alpha) \\ &= \sum_{t=2}^n [\sin(\theta_t + \theta_{t-1}) \cos(2\alpha) - \sin(2\alpha) \cos(\theta_t + \theta_{t-1})]. \end{aligned}$$

Setting this equal to zero gives that

$$\cos(2\hat{\alpha}) \sum_{t=2}^n \sin(\theta_t + \theta_{t-1}) = \sin(2\hat{\alpha}) \sum_{t=2}^n \cos(\theta_t + \theta_{t-1})$$

in which case the mle of  $\alpha$  is

$$\hat{\alpha} = \begin{cases} \frac{1}{2} \tan^{-1} \frac{S_{\theta_t, \theta_{t-1}}}{C_{\theta_t, \theta_{t-1}}} & \text{if } C_{\theta_t, \theta_{t-1}} > 0 \\ \frac{1}{2} \left[ \tan^{-1} \frac{S_{\theta_t, \theta_{t-1}}}{C_{\theta_t, \theta_{t-1}}} + \pi \text{sign}(S_{\theta_t, \theta_{t-1}}) \right] & \text{if } S_{\theta_t, \theta_{t-1}} < 0 \end{cases} \quad (7.18)$$

where  $S_{\theta_t, \theta_{t-1}} = \sum_{t=2}^n \sin(\theta_t + \theta_{t-1})$  and  $C_{\theta_t, \theta_{t-1}} = \sum_{t=2}^n \cos(\theta_t + \theta_{t-1})$ . (7.18) is a mean axis (which identifies an angle  $\beta$  with either  $\beta + \pi$  or  $\beta - \pi$ , whichever is in the interval  $[-\pi, \pi)$ ).

## 7.5 The deterministic component of the model

An analysis of the behaviour of the deterministic component of the model (Equation (7.7)) is, as seen for the Cosine and Sine time series models, useful as a precursory investigation into the model properties. To this end, plots of  $\mu_t$  versus  $\theta_{t-1}$  are produced, for various values of  $\alpha$  and  $\omega$ . Figure 7.1 shows two such plots, in which  $\alpha = 0$  in both cases whilst  $\omega = 0.2$  for the left hand plot and  $\omega = 0.6$  for the right hand plot. Changing the value of  $\alpha$  results in a translation of the plot without altering its shape. In particular, increasing  $\alpha$  by an amount  $a$  translates the curves in Figure 7.1 positively along both  $\theta_{t-1}$  and  $\mu_t$  axes by  $a$ . Switching the sign of  $\omega$  results in a reflection in the line  $\mu_t = \alpha$ . A feature of the deterministic component of the model observable from Figure 7.1, and true more generally is that, for any  $\alpha$  and  $\omega$ , if  $\theta_{t-1} = \alpha - \pi$ ,  $\alpha$  or  $\alpha + \pi$ , then  $\mu_t = \theta_{t-1}$ . Figure 7.1 shows the effect of changing the absolute value of  $\omega$ , in the given plots from 0.2 to 0.6. The value of  $\alpha$  is fixed at 0. For reference, the line  $\mu_t = \theta_{t-1}$  is superimposed. As can be seen, if  $\alpha - \pi < \theta_{t-1} < \alpha$ , then the model rotates  $\theta_{t-1}$  positively toward  $\alpha$ . If  $\alpha < \theta_{t-1} < \alpha + \pi$  then  $\theta_{t-1}$  is rotated negatively toward  $\alpha$ . The magnitude of the rotation in each case is greater the smaller the absolute value of  $\omega$ , as shown by the increase in the curvature as  $\omega$  decreases towards zero.

For negative  $\omega$  and  $\alpha = 0$ , similar considerations show that the model rotates  $-\theta_{t-1}$  towards  $\alpha$  in whichever direction requires the least rotation to reach  $\alpha$ , the magnitude of the rotation decreasing (for fixed  $\theta_{t-1}$ ) as  $\omega \rightarrow -1$ . A similar statement can be made for non-zero  $\alpha$  by defining  $\alpha$  as the zero direction.

Also apparent from the plots is that the deterministic component of the model is both

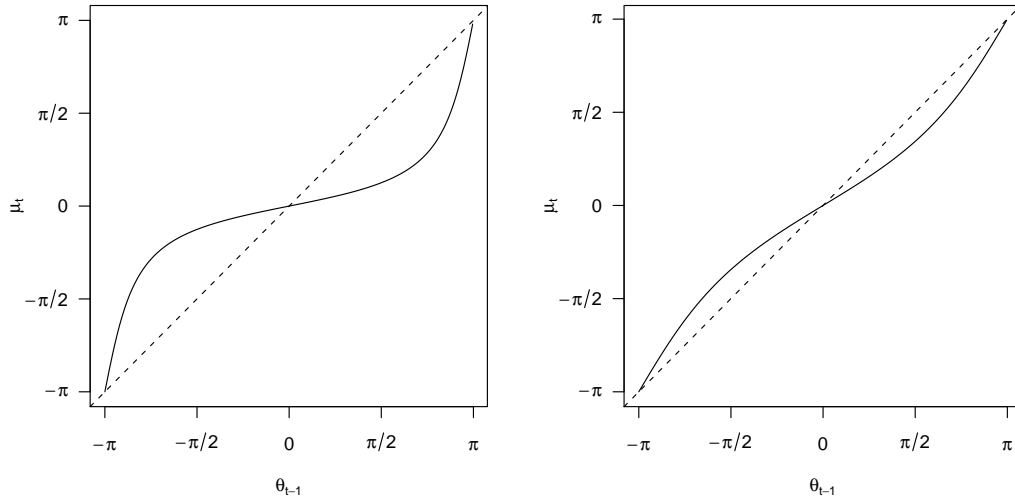


Figure 7.1: Plots of  $\theta_t$  versus  $\theta_{t-1}$  based on Equation (7.4) and for  $\alpha = 0$  (both plots),  $\omega = 0.2$  (left) and  $\omega = 0.6$  (right).

continuous and differentiable.

## 7.6 Simulating data

In order to investigate the Möbius time series model (7.6) data sets are simulated and their behaviour related both to the observations in the previous section regarding the deterministic component of the model and to the inclusions of a random component.

In Section 7.8 we will consider the equilibrium distribution of the process, from which (if an expression for the density of this distribution were obtainable) we could simulate a first value for a data set. In the present section we will proceed as follows in order to obtain a first and subsequent values. With starting value  $y_1 = \alpha$  we simulate a series of values  $y_2, \dots, y_{n_1}$ , say, using Equation (7.3). We then use  $y_{n_1}$  as the first value of our simulated data set, so that  $y_{n_1} = x_1$ , say. A series of  $x$  values  $x_2, \dots, x_{n_2}$ , say, are simulated, again

using (7.3), and we take  $x_1, \dots, x_{n_2}$  to be our simulated data set.

As mentioned in Section 7.2, numerical errors can result if angles and their differences are not taken in the interval  $[-\pi, \pi)$ . To see this, consider the situation  $\theta_{t-1} = \pi - \tau$ ,  $\alpha = -\pi + \tau$  for  $\tau \approx 0$ . Then  $\theta_{t-1} - \alpha = 2\pi - 2\tau$ , and if this difference is not converted to the interval  $[-\pi, \pi)$  then  $(\theta_{t-1} - \alpha)/2 = \pi - \tau$ , which is not what we require. Taking  $\theta_{t-1} - \alpha$  in the interval  $[-\pi, \pi)$ , on the other hand, gives  $(\theta_{t-1} - \alpha)/2 = -\tau$ , which is the correct angle. In the simulation procedure, therefore, it is important to take angles and their differences in the interval  $[-\pi, \pi)$ , which for angle  $\chi$  can be achieved via  $\chi \rightarrow (\chi - \pi) \bmod(2\pi) - \pi$ .

Figure 7.2 shows plots of simulated data for the parameter configurations  $(\omega, \kappa) = (0.7, 2)$ ,  $(-0.2, 2)$ ,  $(0.7, 4)$  and  $(-0.2, 4)$ . The value of  $\alpha$  in each case is taken to be zero. Comparing these plots, the effects of keeping one of  $\omega$  and  $\kappa$  fixed whilst changing the other can be observed. In particular, keeping  $\kappa$  fixed and increasing the absolute value of  $\omega$ , the deterministic component of the model has less of an effect on the transformation of values (see discussion of Figure 7.1) and this is observable in the plots as the series appear more “random”. Keeping  $\omega$  fixed whilst decreasing the value of  $\kappa$  simply induces more randomness into the model through the  $\epsilon_t$  terms in Equation (7.3). A negative value of  $\omega$ , as discussed in Section 7.5, has the effect of switching the sign of  $\theta_{t-1}$  (for  $\alpha = 0$ ) and then rotating the negated value towards zero based on the absolute value of  $\omega$ . In these cases we therefore obtain series that cross the line  $\alpha(=0)$  frequently.

For those parameter configurations considered in the plots, the interpretation of parameters is seen to be simple. More care is needed when we come to simulating data using large negative values of  $\omega$ . As discussed in Section 7.4, when  $\omega = -1$ , the maximum likelihood estimate of  $\omega$  is a mean axis. This property is exhibited by simulated data not only for  $\omega = -1$  but also in data for which  $\omega$  is less than approximately 0.5. Figure 7.3 shows two examples. In each of the plots of Figure 7.3, we have taken  $\alpha = \pi/2$  and  $\kappa = 4$ . For the top two plots  $\omega = -0.6$  whilst for the bottom two  $\omega = -0.9$ .

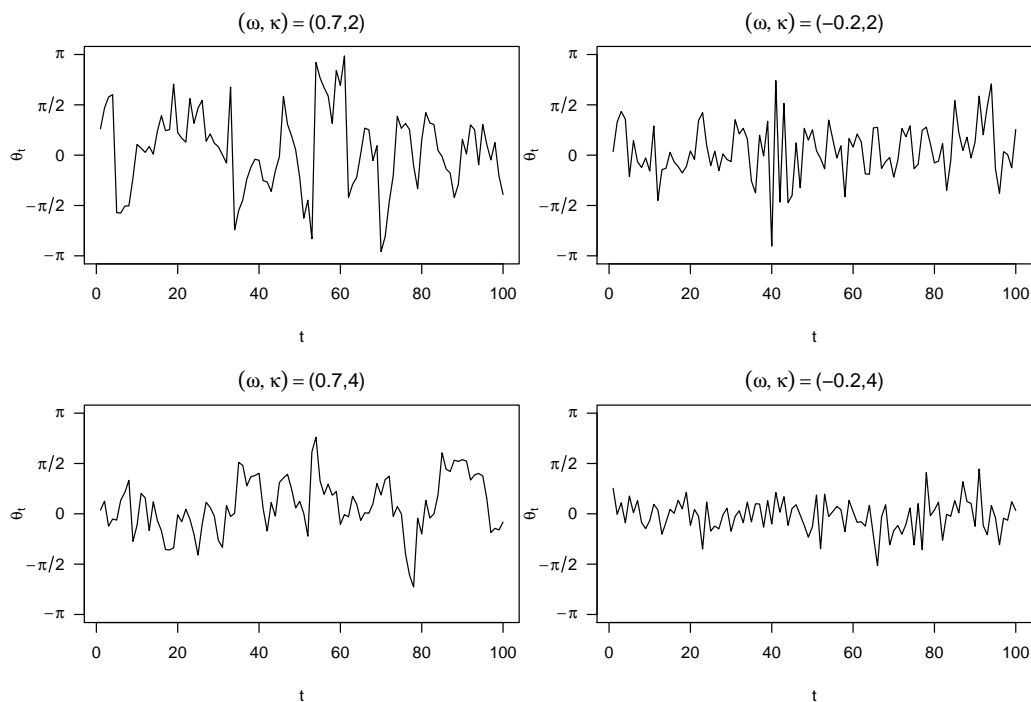


Figure 7.2: Plots of simulated data for  $\alpha = 0$  and those values of  $\omega$  and  $\kappa$  given above each plot.

From the left hand plots, in which the data points are connected with a line, it can clearly be seen that the overall series seems to switch between two “mini series” at a distance of approximately  $\pi$  apart. This is more apparent the closer  $\omega$  gets to  $-1$ . At first glance of these connected plots it may appear that the data are bimodal. The right hand plots, however, which show the same data as points, show that the overall distribution of values appears to tend to uniformity over the interval  $[-\pi, \pi)$  as  $\omega$  tends to  $-1$ .

The equilibrium distribution of the Möbius time series model is discussed in more detail in Section 7.8, before which time-reversibility is investigated.



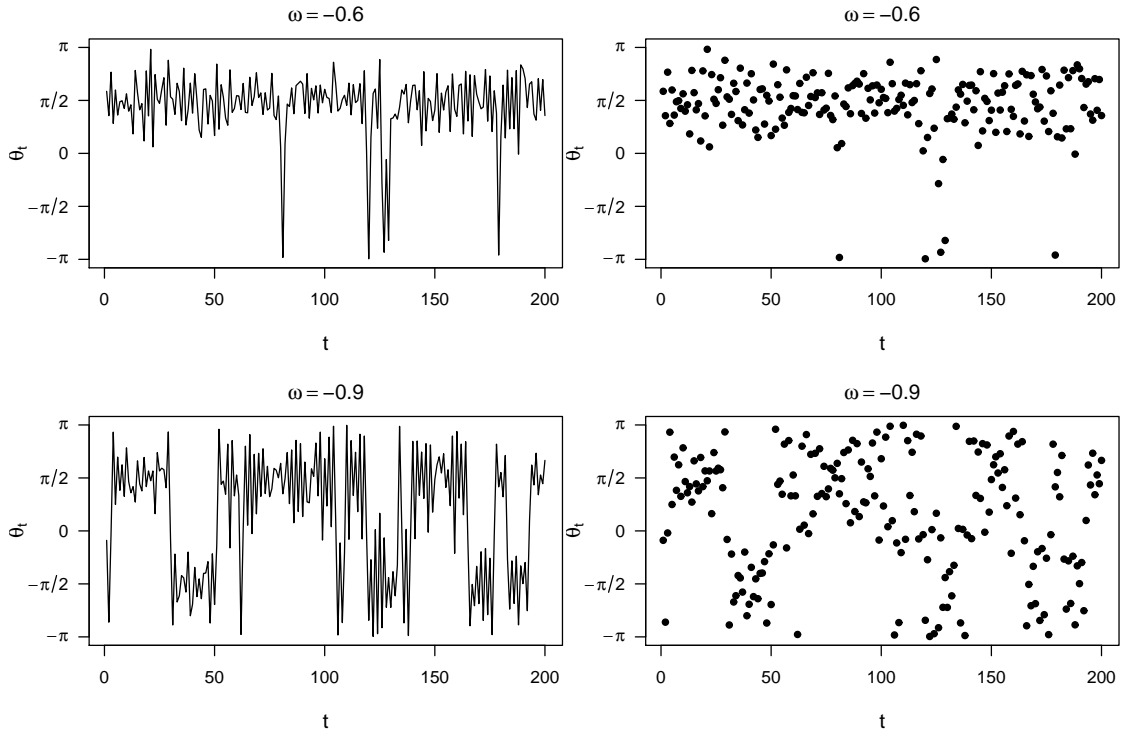


Figure 7.3: Plots of simulated data for  $\alpha = \pi/2$ ,  $\kappa = 4$  and those values of  $\omega$  above each plot.

## 7.7 Checking time-reversibility

An important property in time series modeling is that of time-reversibility. Obtaining the equilibrium distribution of a process can be simplified considerably if the process is time-reversible. In the present section we consider the time-reversibility of the Möbius time series model, and in Section 7.8 we investigate the equilibrium distribution of the model. It should be noted that the time-reversibility property is not one that would necessarily benefit a model for protein data, as polypeptides have a definite start and end, and are not “reversible”.

We proceed by constructing a transition matrix as follows. The interval  $[-\pi, \pi]$  is discretised into a vector  $\boldsymbol{\theta}$  of  $p$ , say, equally spaced values ( $\boldsymbol{\theta}^{(1)} = -\pi, \dots, \boldsymbol{\theta}^{(p)} = \pi$ ).

Each pair  $(\theta^{(i)}, \theta^{(j)})$  of values,  $1 \leq i, j \leq p$  gives rise to an element of a transition matrix  $\mathbf{P}$  via

$$P_{(i,j)} = f_1(\theta_2 = \theta^{(i)} | \theta_1 = \theta^{(j)})$$

where  $P_{(i,j)}$  is the  $(i, j)$ th element of  $\mathbf{P}$  and  $f_1$  is given by Equation (7.10). The transition matrix is then normalised so that its rows sum to one.

Denote by  $\mathbf{p}$  the first eigenvector solution of

$$\mathbf{p}\mathbf{P} = \mathbf{P}.$$

In this case, if

$$p_i P_{i,j} = p_j P_{j,i} \quad (7.19)$$

for all  $(i, j)$  then the model is time-reversible (Feller, 1968). With  $p = 501$  and various starting values for  $\omega$  and  $\kappa$ , solutions are obtained that indicate the model is not time-reversible. Figure 7.4 shows plots of the left hand side minus the right hand side of Equation (7.19) for the parameter values  $(\omega, \kappa) = (0.3, 6)$  (left) and  $(\omega, \kappa) = (0.6, 3)$  (right). In each case, the value of  $j$  used is that value maximising  $|p_i P_{i,j} - p_j P_{j,i}|$  over all values of  $i$ .

As can be seen from the plots, the apparent difference between the left hand side and the right hand side of Equation (7.19) for the Möbius time series model, although small, would indicate that the model is not time-reversible. For the special cases  $|\omega| = 1$  or  $0$  and/or  $\kappa = 0$  the model is seen to be time-reversible.

We next consider the equilibrium distribution of the process.

## 7.8 The equilibrium distribution

The equilibrium distribution of a process describes its statistical behaviour at a point in time a long way from the origin. For a continuous state, discrete time process  $X_1, X_2, \dots$

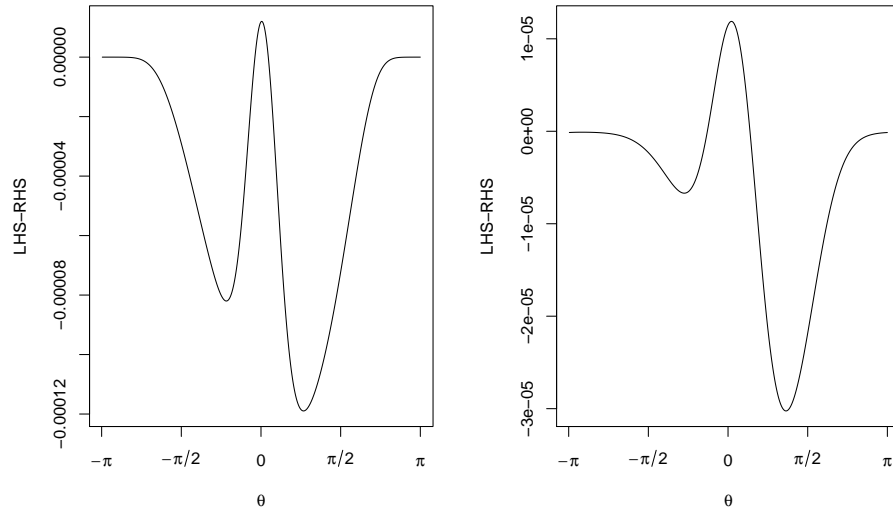


Figure 7.4: Plots of  $LHS - RHS$  of Equation (7.19) versus  $\theta = (-\pi, \pi)$  for  $(\omega, \kappa) = (0.3, 6)$  (left) and  $(\omega, \kappa) = (0.6, 3)$  (right). Plot is for that value of  $j$  leading to the largest absolute value on the  $y$ -axis of the plot.

in which  $f(y, x) = f(X_t = x | X_{t-1} = y)$  is the transition probability,  $t = 2, 3, \dots$ , this equilibrium distribution is given by Cox and Miller (1980, p.135) as

$$f_2(x) = \int_{-\infty}^{\infty} f_1(y, x) f_2(y) dy. \quad (7.20)$$

In Section 7.10 a context is given in which the Möbius model is applied as a continuous state, discrete ‘time’ process  $\theta_1, \theta_2, \dots$ . In this case, if  $y = \theta_{t-1}$  and  $x = \theta_t$ , then the transition probability  $f_1(y, x)$  is given by Equation (7.10), whilst  $f_2(\cdot)$  is the density of the equilibrium distribution. As noted by Cox and Miller (1980), an analytic solution to this equation is not always obtainable, and numerical methods may be required. As mentioned in Section 7.7, the equilibrium distribution may be easier to obtain if the model is time-reversible. Since we appear to have evidence to suggest that this is not the case for the Möbius model, we investigate the equilibrium distribution as follows. In Section 7.8.1, a numerical approach to obtaining an equilibrium distribution is given, whilst Section 7.8.2

outlines an approximate analytic solution under high concentration.

### 7.8.1 Numerical approach

Since time-reversibility cannot be used to simplify the calculation of a theoretical solution to the equilibrium distribution, a numerical solution is here sought. This is obtained by discretising the interval  $[-\pi, \pi]$ , calculating a transition matrix and using an iterative method to carry out a numerical integration. The numerical solutions (NS) found (for various values of  $\omega$  and  $\kappa$ ) are compared with von Mises (VM) distributions and kernel density estimates (KDE) based on samples from the equilibrium distribution, in such a way as to minimise  $(\text{NS} - \text{VM})^2$  and  $(\text{NS} - \text{KDE})^2$ .

Since the transition probabilities are assumed to be stationary,  $\theta_2$  and  $\theta_1$  will be used in place of  $\theta_t$  and  $\theta_{t-1}$ , respectively. Additionally, the mean of all von Mises distributions will be set to zero.

#### Numerical integration

In order to obtain a numerical solution to the equation

$$f_2(\theta_2) = \int f_1(\theta_2|\theta_1)f_2(\theta_1)d\theta_1 \quad (7.21)$$

we obtain the transition matrix  $\mathbf{P}$  as described in Section 7.7. Since the initial distribution (that of  $\theta_1$ ) is unknown,  $f_2(\theta_1)$  is initialised as a vector  $(\mathbf{\Pi}^{(0)})$  of length  $p$  with all elements equal to  $1/p$ . The numerical integration is then performed using the iteration

$$\mathbf{\Pi}^{(k+1)} = \mathbf{P}\mathbf{\Pi}^{(k)}, \quad k = 0, 1, \dots$$

until

$$\|(\mathbf{\Pi}_t^{(k+1)} - \mathbf{\Pi}_t^{(k)})^2\| < \epsilon,$$

where  $\Pi_l^{(k)}$  is the  $l$ 'th element of  $\Pi^{(k)}$ . The squared Euclidean distance is used with  $\epsilon = 10^{-6}$ .

With the solution of the numerical integration denoted by  $\Pi^{(\infty)}$ , a plot of  $\Pi^{(\infty)}$  against  $\theta$  is informative in showing the shape of the numerical solution obtained. Figure 7.5 shows such a plot, in which the values  $\omega = 0.3$  and  $\kappa_1 = 4$  are chosen for the calculation of  $f_1(\theta_2|\theta_1)$  in the transition matrix, where  $\kappa_1$  is the concentration parameter of the conditional pdf  $f_1(\theta_2|\theta_1)$ . For comparison, a plot of the von Mises pdf with parameters

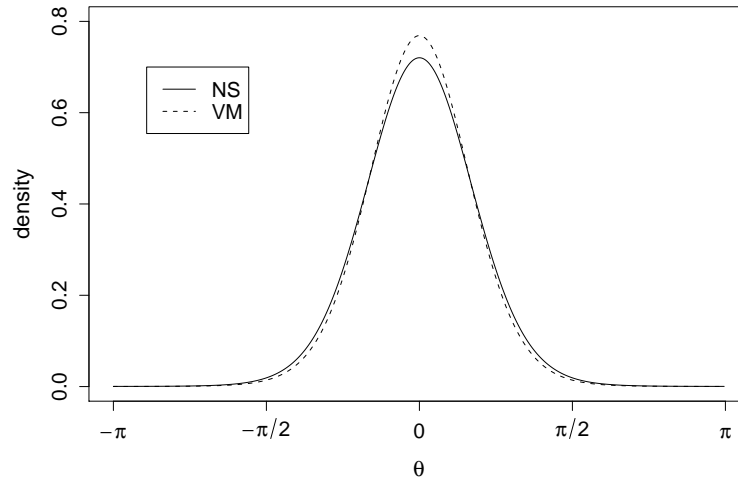


Figure 7.5: Comparison of numerical solution to the equilibrium distribution (solid line) based on initial parameter values  $\omega = 0.3$ ,  $\kappa_1 = 4$  and the von Mises density (dashed line) with  $\mu = 0$  and  $\kappa = 4$ .

$\mu = 0$  and  $\kappa = 4$  is superimposed on the same plot with a dashed line. As can be seen, the plots are very similar in shape, suggesting that the equilibrium distribution is at least close to a von Mises distribution, even if not actually von Mises. The comparison between the numerical equilibrium distribution and the von Mises distribution is further explored later in the section.

We observed in Section 7.6 that for  $\omega$  close to  $-1$  the equilibrium distribution appears to

be approximately uniform. The numerically calculated equilibrium distribution concurs with this, and a circular uniform distribution is obtained when  $|\omega| = 1$  and/or  $\kappa = 0$ . If  $\omega = 0$  then it is easily seen that the equilibrium distribution is von Mises with concentration  $\kappa$ .

As an alternative to numerical integration, the eigen decomposition of Section 7.7 could be used in place of the above approach in order to obtain a numerical solution to the equilibrium distribution. Comparison between the two approaches in the current context shows negligible disagreement between the solutions obtained.

### **Comparing the numerical solution with a supposed sample from the equilibrium distribution**

In order to obtain a sample from the equilibrium distribution of the model, data are simulated as described in Section 7.6 for chosen values of  $\omega$  and  $\kappa_1$ , and values from these data are sampled sufficiently far apart to ensure that the correlation structure imposed by  $\omega$  is no longer influential. In this way, a comparison can be made between the numerical solution to the equilibrium distribution and this supposed sample. Indeed, a kernel density estimate (kde) based on the sample can be computed, and the `nlm` function in R used to choose a bandwidth for the kde to minimise

$$\sum_{\theta} [(f_2(\omega, \kappa_1) - \hat{f}_2(b))^2]. \quad (7.22)$$

The function  $f_2(\omega, \kappa_1)$  in (7.22) is the vectorial solution to the equilibrium distribution. That is,  $f_2(\omega, \kappa_1)$  is the numerical solution to the equilibrium distribution calculated for the  $p$  values  $(-\pi, \dots, \pi)$  and based on the initial choices of  $\omega$  and  $\kappa_1$ . The function  $\hat{f}_2(b)$  is the kernel density estimate of the sample from the equilibrium distribution, calculated over the same  $p$  values from  $-\pi$  to  $\pi$ , and dependent on the choice  $b$  of bandwidth. As with the Sine and Cosine time series models, the appropriate adjustment for density estimation from circular data is made before the density is estimated.

Figure 7.6 shows two plots making the comparison explained above, the first of which is based on the values  $(\omega = 0.6, \kappa_1 = 3)$  and the second on the values  $(\omega = -0.1, \kappa_1 = 5)$ .

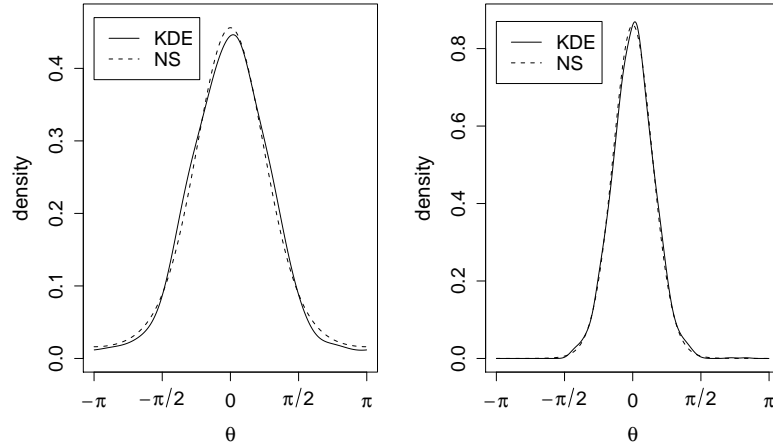


Figure 7.6: Comparisons of numerical solutions (NS) to the equilibrium distribution and kde's based on samples from the equilibrium distribution, based on initial parameters  $(\omega = 0.6, \kappa_1 = 3)$  (left) and  $(\omega = -0.1, \kappa_1 = 5)$  (right).

In each case, the simulated data are sampled forty apart to obtain the sample from the equilibrium distribution. Each sample is of size 400. As can be seen in both cases, there is a close resemblance between the plot of the numerical solution and the kernel density estimate.

Figure 7.7 shows a polar plot corresponding to the right hand plot of Figure 7.6. The plots, in cartesian coordinates, are those of  $((1 + g) \cos \theta, (1 + g) \sin \theta)$ , where the function  $g$  is given by  $f_2(\omega, \kappa_1)$  for the numerical solution (dashed line) and by  $\hat{f}_2(b)$  for the kernel density estimate (solid line). For comparison with these plots, the unit circle is also displayed. As with the plot in Figure 7.6, there is seen to be a close resemblance between the numerical solution and the kernel density estimate.

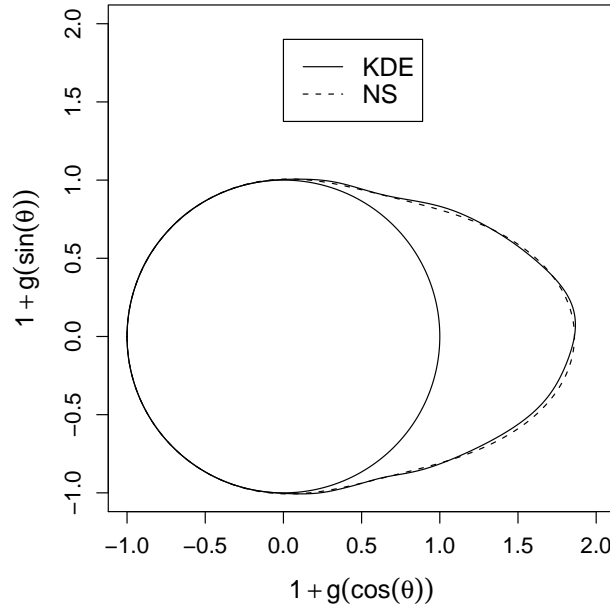


Figure 7.7: Polar plot based on the right hand plot of Figure 7.6 comparing the numerical solution (NS) to the equilibrium distribution, based on initial values  $\omega = -0.1, \kappa_1 = 5$ , with a kernel density estimate (KDE) based on a sample from the equilibrium distribution with these parameter values.

### Comparing the numerical solution with a von Mises distribution

Having obtained a numerical solution to the equilibrium distribution that looks as though it may be at least approximated by a von Mises distribution, attempts are made to compare von Mises distributions, with various values of  $\kappa = \kappa_2$ , say, to the numerical solution obtained, given the initial values of  $\omega$  and  $\kappa_1$  that were used in obtaining the numerical solution. In particular, if we denote by  $f_2(\omega, \kappa_1)$  the numerical solution as calculated with the values  $\omega$  and  $\kappa_1$ , and by  $f(\kappa_2)$  the pdf of the von Mises distribution with mean direction equal to zero and concentration parameter  $\kappa_2$ , then we wish to choose  $\kappa_2$  to



minimise

$$\sum_{\theta} \{[f_2(\omega, \kappa_1) - f(\kappa_2)]^2\}. \quad (7.23)$$

The sum in Equation (7.23) is taken over the values in the vector  $\theta$ , and is minimised using the `nlm` function in R. Of additional interest is the relationship between  $\kappa_2^*$  and  $(\omega, \kappa_1)$ , where  $\kappa_2^*$  is that value of  $\kappa_2$  which minimises (7.23).

Figure 7.8(a) shows a plot comparing  $f_2(\omega = 0.6, \kappa_1 = 4)$  and  $f(\kappa_2^*)$ , where  $\kappa_2^*$  has been calculated to be 2.40. As can be seen, there is a reasonably close resemblance between the two, suggesting that a von Mises distribution can, at least, be a good approximation to the equilibrium distribution. Similar experimentation with other starting values of  $\omega$  and  $\kappa_1$  lead to the same conclusion.

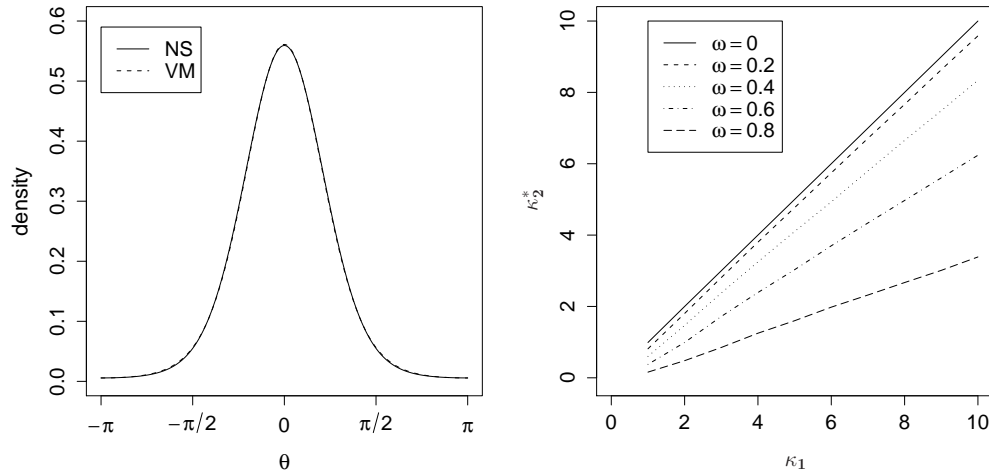


Figure 7.8: Left: comparison of numerical solution (NS) for equilibrium distribution based on the values  $(\omega = 0.6, \kappa_1 = 4)$  and the von Mises pdf (VM) with parameters  $(\mu = 0, \kappa_2 = 2.4)$ . Right: Plots of  $\kappa_1$  versus  $\kappa_2^*$  for  $\omega = 0, 0.2, 0.4, 0.6, 0.8$  (top line to bottom line).

In order to investigate the relationship between  $\kappa_2^*$  and  $(\omega, \kappa_1)$ ,  $\kappa_2^*$  is computed for all combinations of  $1 \leq \kappa_1 \leq 10$  and  $\omega = 0, 0.1, 0.2, \dots, 0.8$ . (Negative values of  $\omega$  give the

same results as positive values.) Some of the results are displayed (with points connected for ease of interpretation) on part (b) of Figure 7.8, with the lines, from top to bottom, representing  $\omega = 0, 0.2, 0.4, 0.6$  and  $0.8$  respectively. As can be seen, each line is virtually straight as  $\kappa_1$  increases. As expected, the line representing  $\omega = 0$  has a zero intercept and a slope equal to unity, since in this case  $f_1(\theta_2|\theta_1)$  in Equation (7.21) is the von Mises pdf with parameters (mean) zero and (concentration)  $\kappa_1$ . Also apparent from the plot is that  $\omega$  and  $\kappa_1$  are negatively correlated in their effect on  $\kappa_2^*$ . This can be seen on consideration of the fact that increasing (the absolute value of)  $\omega$  decreases  $\kappa_2^*$  (for fixed  $\kappa_1$ ), whereas increasing the value of  $\kappa_1$  increases  $\kappa_2^*$  (for fixed  $\omega$ ).

To further investigate the relationship between  $\kappa_2^*$  and  $(\kappa_1, \omega)$  regression models are fitted that are consistent with the right hand plot of Figure 7.8. That is, regression models in which  $\kappa_1$  acts multiplicatively on  $\kappa_2^*$  (since  $\kappa_1 = 0 \Rightarrow \kappa_2^* = 0$ ) and for which

- $\omega = 0 \Rightarrow \kappa_1 = \kappa_2^*$
- $\kappa_2$  depends on either the absolute or squared (or both) value of  $\omega$ ,

are fitted. The models fitted are as follows:

$$\kappa_2^* = \alpha\kappa_1 + \beta\kappa_1\omega^2 \quad (7.24)$$

$$\kappa_2^* = \alpha\kappa_1 + \beta\kappa_1|\omega|$$

$$\kappa_2^* = \alpha\kappa_1 + \beta\kappa_1\omega^2 + \gamma\kappa_1|\omega|$$

The adjusted  $R^2$  values for the three models were respectively 0.9996, 0.994 and 0.9996. The only parameter not significantly different from zero at the 1% level was  $\gamma$ , with a  $p$ -value of 0.14. The model selected for further investigation is therefore that given by Equation (7.24), for which  $\hat{\alpha} = 0.997$  and  $\hat{\beta} = -1.048$ , so that

$$\hat{\kappa}_2^* \approx \kappa_1(1 - \omega^2).$$

The left hand plot of Figure 7.9 shows a plot of the fitted values ( $\hat{\kappa}_2^* = \hat{\alpha}\kappa_1 + \hat{\beta}\kappa_1\omega^2$ ) versus the residuals, in which a different symbol type represents each different value of  $\kappa_1$ . As can be seen, there is a curvature to the plot overall, which indicates the inclusion

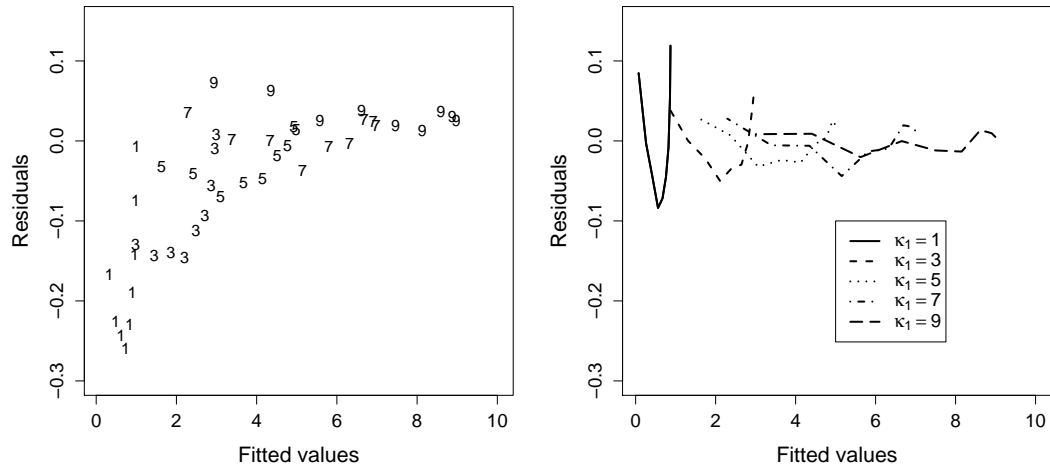


Figure 7.9: Left: Residuals versus fitted values for regression model  $\hat{\kappa}_2^* = 0.997 - 1.048\kappa_1\omega^2$ , with a symbol for each  $\kappa = 1, 3, 5, 7, 9$ . Right: Residuals versus fitted values for five models of the form  $\kappa_2^* = \alpha\kappa_1 + \beta\kappa_1\omega^2$ , with  $\kappa_1 = 1, 3, 5, 7, 9$  (increasing from left to right).

of a higher order term in  $\kappa_1$  or  $\omega$ . Also apparent is a decrease in curvature and range of residuals as  $\kappa_1$  increases, as can be seen on consideration of each of the different sets of symbols. This suggests that the fit of the model improves as  $\kappa_1$  increases. This point is emphasised by the right hand plot of Figure 7.9. For this plot, five separate regression models of the form (7.24) have been fitted, one for each of the values  $\kappa_1 = 1, 3, 5, 7, 9$ , and the residuals plotted against the fitted values. It is immediately apparent that the fit of the model improves as  $\kappa_1$  increases.

### 7.8.2 Analytic approach

In Downs and Mardia (2002, p.56), an alternative form for Equation (7.10) is shown to be (given here in terms of  $x$  and  $y$ )

$$f_1(x|y) = [2\pi I_0(\kappa_1)e^{-\kappa_1}]^{-1} \exp\{-2\kappa_1 \mathbf{y}^T \boldsymbol{\eta} \boldsymbol{\eta}^T \mathbf{y} / (1 + 2\omega_p \omega_q \cos y)\} \quad (7.25)$$

where  $f$  and  $\kappa$  have been replaced by  $f_1$  and  $\kappa_1$  respectively, as suggested by the current context, and where  $y$ ,  $\eta$ ,  $\omega_p$  and  $\omega_q$  are defined as follows:

$$\begin{aligned} \mathbf{y}^T &= [\sin 1/2(x+y), -\cos 1/2(x+y), \sin 1/2(x-y), -\cos 1/2(x-y)]; \\ \boldsymbol{\eta}^T &= [\omega_p \cos \alpha, \omega_p \sin \alpha, \omega_q, 0]; \\ \omega_p^2 + \omega_q^2 &= 1; \quad 2\omega_p \omega_q = \frac{1 - \omega^2}{1 + \omega^2}. \end{aligned} \quad (7.26)$$

Setting  $\alpha$  equal to zero and for ‘large’  $\kappa$ , we have

$$\begin{aligned} \mathbf{y}^T &\approx [(x+y)/2, -1, (x-y)/2, -1]; \\ \boldsymbol{\eta}^T &= [\omega_p, 0, \omega_q, 0]; \\ 1 + 2\omega_p \omega_q \cos y &\approx 1 + 2\omega_p \omega_q, \end{aligned} \quad (7.27)$$

substitution of which into (7.25) yields, upon simplification,

$$f_1(x|y) = [2\pi I_0(\kappa_1)e^{-\kappa_1}]^{-1} \exp\left\{\frac{-2\kappa_1[\omega_p(x+y)/2 + \omega_q(x-y)/2]^2}{1 + 2\omega_p \omega_q}\right\}. \quad (7.28)$$

Multiplying Equation (7.28) by  $f_2(y)$ , the von Mises pdf with mean direction equal to 0, gives an expression in which a function  $C(\kappa_1, \kappa_2)$  of  $\kappa_1$  and  $\kappa_2$  multiplies an exponential term. This exponential term is a quadratic, in  $y$ , of the form  $ay^2 + by + c$ , which can be written equivalently as  $a(y + b/2a)^2 + c - b^2/4a$ . In this case, Equation (7.20) becomes

$$[2\pi I_0(\kappa_2)]^{-1} e^{\kappa_2 \cos x} = C(\kappa_1, \kappa_2) \exp(c - b^2/4a) \int_{-\infty}^{\infty} \exp[a(y + b/2a)^2] dy. \quad (7.29)$$

The integrand in Equation (7.29) is proportional to the pdf of a normal distribution. Using this fact and the approximation  $\cos x \approx 1 - x^2/2$ , we have, from (7.29),

$$\exp[\kappa_2(1 - x^2/2)] \propto \exp[c - b^2/4a]. \quad (7.30)$$

Using the expressions in Equation (7.26) for  $\omega_p$  and  $\omega_q$  in terms of  $\omega$ , and some algebra, we can calculate

$$c - b^2/4a = -\frac{\kappa_1 \kappa_2 x^2}{2(\kappa_1 \omega^2 + \kappa_2)}. \quad (7.31)$$

Putting (7.31) in (7.29), we see that both the left and right hand sides of (7.29) have a normalising constant and an exponential term involving  $x^2$ , in which case, since the left hand side is a von Mises pdf, the right hand side must also be the pdf of a von Mises distribution. Using (7.31) and equating terms in  $x^2$  in Equation (7.30) gives

$$\kappa_2 = \kappa_1(1 - \omega^2). \quad (7.32)$$

Thus, for ‘large’  $\kappa$ , the equilibrium distribution is approximately von Mises distributed with concentration parameter  $\kappa_2 = \kappa_1(1 - \omega^2)$ . This result is both simple and intuitively appropriate, as  $\omega = 0 \Rightarrow \kappa_2 = \kappa_1$ , and  $\kappa_1 = 0 \Rightarrow \kappa_2 = 0$ .

Returning to the regression model of Section 7.8.1, we see that there is agreement between the theoretical relationship of Equation (7.32) obtained above and the regression model suggested by a numerical approach to the equilibrium distribution. Moreover, the approximation obtained above explains the patterns of the residual plots of Figure 7.9. That is, the curvature in the first plot can be attributed to taking only the first two terms in the expansion of  $\cos x$  with the approximation  $\cos x \approx 1 - x^2/2$ , whilst clearly the improvement in fit for increasing  $\kappa_1$  evident in the second plot is due to the fact that  $\kappa_2 = \kappa_1(1 - \omega^2)$  is an approximation for large  $\kappa_1$ . Of additional interest is the rate of this improvement. In particular, there appears to be a considerable improvement between  $\kappa_1 = 1$  and  $\kappa_1 = 3$ , after which the rate of improvement decreases. This plot therefore gives a graphical indication of how ‘large’  $\kappa_1$  must be in order for the approximation to be ‘acceptable’ in some sense.

To investigate this further, a grid of values is calculated for  $1 \leq \kappa_1 \leq 10$  and  $\omega = 0, 0.1, 0.2, \dots, 0.8$ , in which each cell is the absolute difference

$$|\kappa_2^* - \kappa_1(1 - \omega^2)|, \quad (7.33)$$

where  $\kappa_2^*$ , as in Section 7.8.1, is that value of the concentration parameter of a von Mises distribution which minimises the difference between the numerical equilibrium distribution and a von Mises distribution, as determined by Equation (7.23). The results are displayed on the left hand plot of Figure 7.10. The profiles at  $\kappa_1 = 0$  and  $\omega = 0$

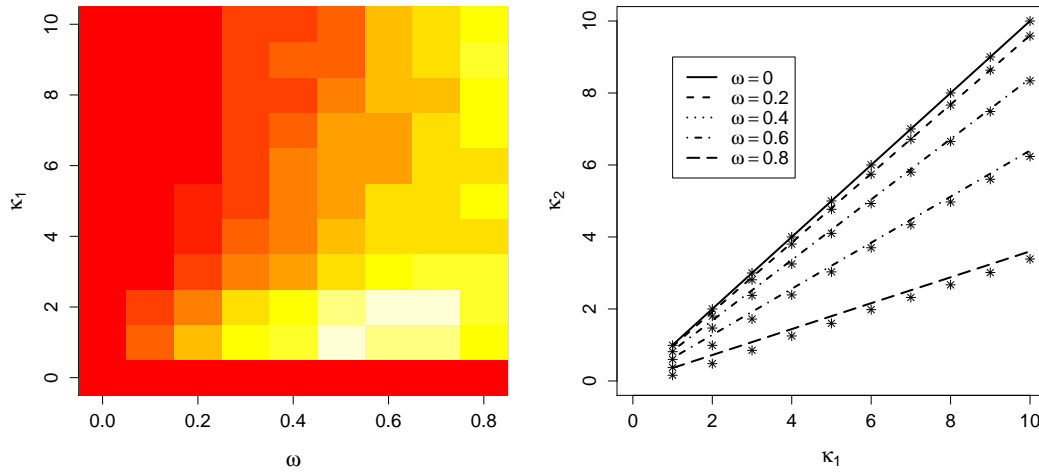


Figure 7.10: Left: Image plot of Equation (7.33) for  $\kappa_1 = 1, \dots, 10$  and  $\omega = 0, 0.1, \dots, 0.8$ . Right: Lines representing  $\kappa_2 = \kappa_1(1 - \omega^2)$  for  $\kappa_1 = 1, \dots, 10$  (x-axis) and  $\omega = 0, 0.2, 0.4, 0.6, 0.8$  (top line to bottom line), with corresponding  $\kappa_2^*$  superimposed as points.

represent a perfect fit. In other areas of the image, the improvement in fit is evident as  $\kappa_1$  increases, but more so for smaller values of  $\omega$ . This point is emphasised by the right hand plot of Figure 7.10, in which the lines  $\kappa_2 = \kappa_1(1 - \omega^2)$  (for  $1 \leq \kappa_1 \leq 10$  and  $\omega = 0, 0.2, 0.4, 0.6, 0.8$ ) are displayed alongside points which are the corresponding values of  $\kappa_2^*$ . The lines, from top to bottom, correspond to  $\omega = 0, 0.2, 0.4, 0.6$  and  $0.8$ , and

are approximated more closely by the points for smaller values of  $\omega$ , the approximation within each line improving (with increasing  $\kappa$ ) less as  $\omega$  increases.

## 7.9 Estimating model parameters

In this section we consider parameter estimation for the Möbius time series model. Firstly, a grid search method is used in order to estimate the parameters  $\alpha$  and  $\omega$  based on Equation (7.11), and the behaviour of the function (7.11) is investigated over this grid of values. Second, a numerical approach is considered using the R function `nlm`. For each, the method outlined in Section 7.6 is employed for data simulation, and the parameters estimated based on the simulated data.

### 7.9.1 Grid search and likelihood plots

Equation (7.11) can be evaluated for a grid of values in  $(\alpha, \omega)$  space which contains the true parameter values  $(\alpha_0, \omega_0)$ , say, used in the simulation of data. Searching this grid for values of  $\alpha$  and  $\omega$  that maximise (7.11) is then a means of estimating the model parameters, and more specifically a way to compare these estimates with the  $(\alpha_0, \omega_0)$  that generated the data. Evaluating a grid of values in this way also enables one to investigate the behaviour of the conditional likelihood by plotting Equation (7.11) either as a 2-dimensional perspective plot, a contour plot, or as profile plots for one of  $\alpha$  or  $\omega$ , given a fixed value (for example  $\alpha_0$  or  $\omega_0$ ) of the other.

Two sets of data are investigated using the above methods. Firstly, the values  $(\alpha_0, \omega_0, \kappa_0) = (\pi/4, 0.6, 4)$  are used to simulate a data set of length  $n = 200$ . Using a 300 by 100 grid of values covering the intervals  $\alpha = [-\pi, \pi]$  and  $\omega = [-1, 1]$  respectively, the maximum value of (7.11) for the simulated data across the grid of values is found to be

that value for which  $(\tilde{\alpha}, \tilde{\kappa}) = (0.788, 0.556)$ , which are close to the true values  $\alpha_0$  and  $\omega_0$  ( $\pi/4 = 0.785$  to three decimal places).

Figure 7.11 shows plots based on the values of Equation (7.11) for various parts of the grid. The top left plot shows  $l(\alpha_0, \omega)$  for  $\omega$  between  $-1$  and  $1$ , whilst the top right plot shows  $l(\alpha, \omega_0)$  for  $\alpha$  between  $-\pi$  and  $\pi$ . The bottom left plot  $l(\alpha, -\omega_0)$  again for  $\alpha$  between  $-\pi$  and  $\pi$ . Finally, an image and contour plot is shown for  $\alpha \in [\alpha_0 - 0.4, \alpha_0 + 0.4]$  and  $\omega \in [\omega_0 - 0.1, \omega_0 + 0.1]$ . In each plot, dashed lines show the true values  $\alpha_0$  and  $\omega_0$ .

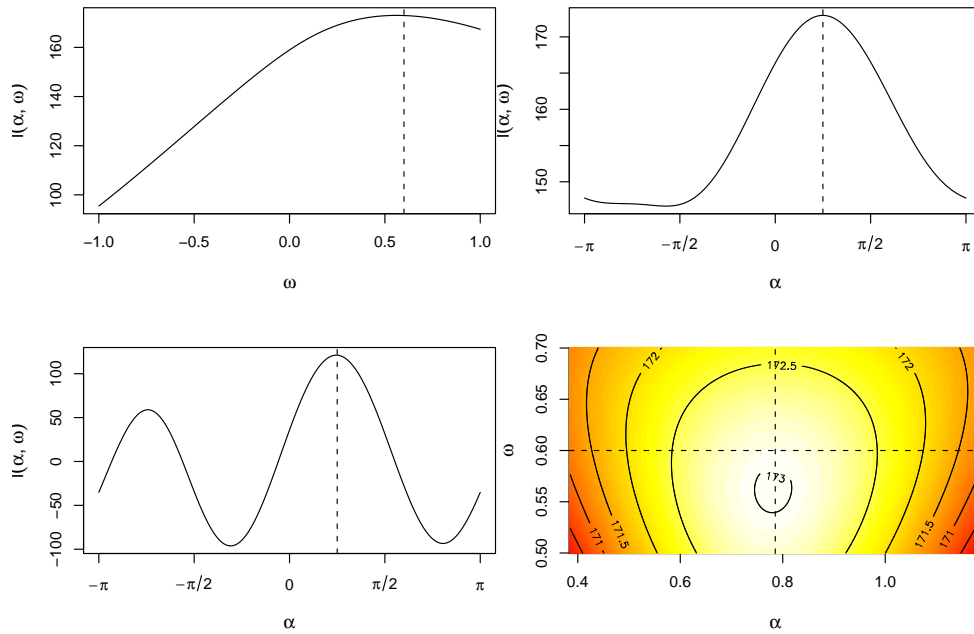


Figure 7.11: Likelihood plots based on simulated data set 1, in which  $\alpha_0 = \pi/4$ ,  $\omega_0 = 0.6$ ,  $\kappa = 4$ ,  $n = 200$ . (a) Profile in  $\omega$  with  $\alpha = \alpha_0$ . (b) Profile in  $\alpha$  with  $\omega = \omega_0$ . (c) Profile in  $\alpha$  with  $\omega = -\omega_0$ . (d) Contour likelihood plot.

The first profile in the figure, for  $\alpha$  fixed at  $\alpha_0$ , shows a maximum for the profile of the likelihood at a value close to  $\omega_0 = 0.6$ . Similarly, when  $\omega$  is fixed at  $\omega_0$ , the profile of (7.11) shows a maximum around  $\pi/4$ . In both these cases, the profile plots each exhibit a single maximum. In the third of the profile plots however, when  $\omega$  is fixed at  $-\omega_0$ ,



there are also local maxima at approximately  $\alpha_0 \pm \pi$ . The reason for this can be seen on consideration of the special case  $\omega = -1$  discussed in Section 7.4. In this case, the maximum likelihood estimate of  $\alpha$  is a mean axis. Although in the present case the maximum of the likelihood is considerably less in the profile exhibiting the local maxima, this may not necessarily be the case more generally (when  $\omega$  is very close to  $-1$ ).

The second set of data simulated are obtained using the values  $(\alpha_0, \omega_0, \kappa_0) = (2\pi/3, -0.4, 4)$  and  $n = 200$ . Using the grid search approach we obtain the parameter estimates  $(\tilde{\alpha}, \tilde{\omega}) = (2.049, -0.414)$ . These values are again close to the true values  $\alpha_0$  and  $\omega_0$  ( $2\pi/3 = 2.094$  to three decimal places). Figure 7.12 shows the plots produced in Figure 7.11 for the new data set. In this case, however, the image plot is plotted over the intervals  $\alpha \in [\alpha_0 - 0.2, \alpha_0 + 0.2]$  and  $\omega \in [\omega_0 - 0.2, \omega_0 + 0.2]$ .

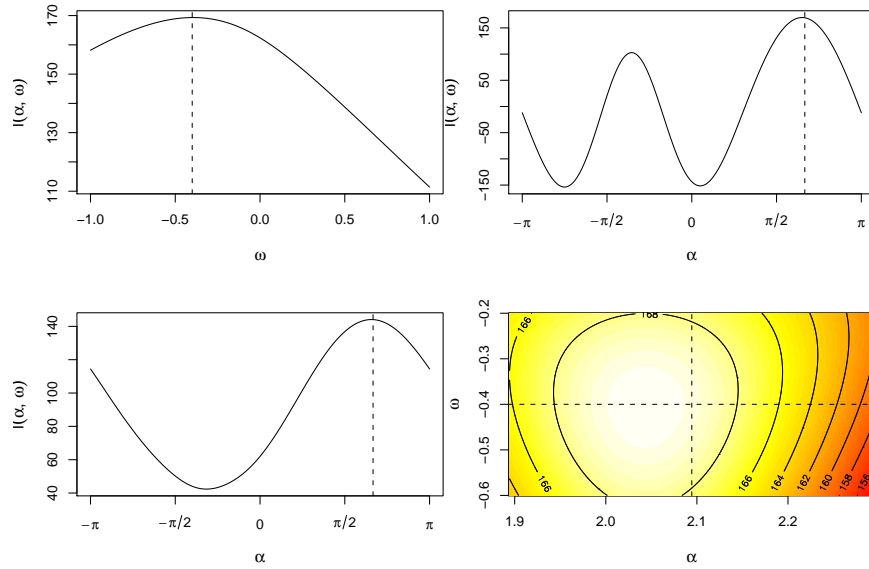


Figure 7.12: Likelihood plots based on simulated data set 2, in which  $\alpha_0 = 2\pi/3$ ,  $\omega_0 = -0.4$ ,  $\kappa = 4$ ,  $n = 500$ . (a) Profile in  $\omega$  with  $\alpha = \alpha_0$ . (b) Profile in  $\alpha$  with  $\omega = \omega_0$ . (c) Profile in  $\alpha$  with  $\omega = -\omega_0$ . (d) Contour likelihood plot.

As for data set one, we again observe that the maximum grid values are reasonably close

to the true parameter values. For data set one, the maximum of the function (7.11) is seen to lead to a more accurate estimate of  $\alpha$  than of  $\omega$ , whilst for data set two the opposite is true. For data set two we again observe the phenomenon of local maxima in the profile likelihood of  $\alpha$  when  $\omega < 0$ . This has the potential to be more of an issue when the true parameter value is negative, although the maximum is clearly visible and a value of  $\omega_0$  closer to  $-1$  may be needed to cause problems in estimation.

We next consider a numerical approach to maximising the likelihood function  $l(\alpha, \omega)$  by using the R function `nlm`.

## 7.9.2 Gradient descent

In this section we investigate the use of `nlm` to estimate model parameters for the Möbius time series model based on data sets simulated using various parameter values.

Experimentation with numerous data sets shows that, unless the value  $\omega_0$  that is used for data simulation is very close to  $-1$ , and/or  $\kappa$  is very small, the starting values  $(\alpha, \omega) = (\bar{\theta}, 0)$  result in reasonably good parameter estimates, where  $\bar{\theta}$  is the mean direction of the simulated data set. Indeed, the mean direction itself provides a reasonable estimate of  $\alpha$ . In the case  $\omega \approx -1$ ,  $\bar{\theta}$  occasionally correctly gives  $\alpha_0$ , but also occasionally gives  $\alpha_0 \pm \pi$ , for the reasons given in Section 7.4. On some occasions, the estimation method appears to fail, returning estimates of approximately zero for both  $\alpha$  and  $\omega$ , even when the true value of the former is  $\pi/2$  and the true value of the latter is close to  $-1$ .

For small values of  $\kappa$  (below about 1), a larger value of  $n$  enables better estimation of  $\alpha$  provided the value of  $\kappa$  is not “too” small.

Tables 7.1, 7.2 and 7.3 display the results of estimating  $\alpha$ ,  $\omega$  and  $\kappa$  using the `nlm` function for  $\alpha$  and  $\omega$  and using the profile likelihood (see Section 7.3) for  $\kappa$ . The latter is obtained

by minimising the absolute value of

$$\frac{I_1(\kappa)}{I_0(\kappa)} - \frac{l_{\hat{\alpha}, \hat{\omega}}}{(n-1)}$$

over a sequence of values for  $\kappa$  between 0.1 and 10, once  $l_{\hat{\alpha}, \hat{\omega}}$  (from Equation (7.11)) has been obtained via `nlm`.

For each combination of true parameter values  $(\alpha_0, \omega_0, \kappa_0)$ , 100 data sets, each comprising 400 observations, are simulated. The mean values and standard deviations of the parameter estimates over the 100 replications, within each  $(\alpha, \omega, \kappa)$  combination, are then calculated. In the case of  $\alpha$ , we calculate  $1 - \overline{R}$  instead of the standard deviation, where  $\overline{R}$  is the mean resultant length of the estimates of  $\alpha$ . The parameter values used in the simulation process are every combination of  $(\alpha_0, \omega_0, \kappa_0)$  from  $\alpha_0 = (-2\pi/3, 2\pi/3)$ ,  $\omega_0 = (-0.5, 0.5)$  and  $\kappa_0 = (1, 5)$ .

As can be seen from the tables, the means of the parameter estimates are close to the original parameter estimates that generated the data, whilst the standard deviations of estimates for  $\alpha$  and  $\omega$  are greater for  $\kappa = 1$  than for  $\kappa = 5$ .

The problems of parameter estimation induced by bimodality (when  $\omega$  is close to  $-1$ ) are not observed for the data simulated with the parameter values given in Table 7.3. For individual data sets when this is a problem, a visual analysis of the data (for example those data plotted in Figure 7.3) may give a better idea of the true parameter values.

## 7.10 Application to protein data

Having experimented with numerous simulated data sets, the model is applied to a real data set comprising the conformational angles of a protein.

In an attempt to categorise proteins by the properties and nature of their 3-D conformation, the Möbius time series model can be applied to data sets comprising the  $\phi$  and  $\psi$  angles

$\alpha_0 \rightarrow$		$-2\pi/3$		$2\pi/3$	
$\kappa_0 \rightarrow$		1	5	1	5
$\omega_0$	-0.5	-2.093 (0.001)	-2.094 (0.000)	2.089 (0.004)	2.096 (0.002)
	0.5	-2.084 (0.028)	-2.100 (0.001)	2.120 (0.050)	2.092 (0.003)

Table 7.1: Mean values (and one minus mean resultant length) of estimates of  $\alpha$  for 8 combinations of  $(\alpha_0, \kappa_0, \omega_0)$ , each based on 100 simulated data sets of 400 observations.

$\alpha_0 \rightarrow$		$-2\pi/3$		$2\pi/3$	
$\kappa_0 \rightarrow$		1	5	1	5
$\omega_0$	-0.5	-0.499 (0.052)	-0.497 (0.034)	-0.493 (0.083)	-0.496 (0.038)
	0.5	0.501 (0.057)	0.497 (0.037)	0.504 (0.062)	0.493 (0.035)

Table 7.2: Mean values (and standard deviations) of estimates of  $\omega$  for 8 combinations of  $(\alpha_0, \kappa_0, \omega_0)$ , each based on 100 simulated data sets of 400 observations.

$\alpha_0 \rightarrow$		$-2\pi/3$		$2\pi/3$	
$\kappa_0 \rightarrow$		1	5	1	5
$\omega_0$	-0.5	0.998 (0.081)	5.063 (0.329)	1.000 (0.066)	4.997 (0.295)
	0.5	0.994 (0.070)	5.034 (0.259)	0.992 (0.070)	4.971 (0.264)

Table 7.3: Mean values (and standard deviations) of estimates of  $\kappa$  for 8 combinations of  $(\alpha_0, \kappa_0, \omega_0)$ , each based on 100 simulated data sets of 400 observations.

that define their conformation. Ultimately, of course, the challenge is to incorporate both  $\phi$  and  $\psi$  jointly into the analysis. As a first stage, however, the model developed is applied to  $\phi$  and  $\psi$  individually.

The data to be considered comprise the 338  $\phi$  and  $\psi$  angles, as described above, of a particular peptide chain. Table 7.4 displays maximum likelihood estimates of the parameters  $\alpha$ ,  $\omega$  and  $\kappa$  for each of  $\phi$  and  $\psi$ . The table also shows estimated standard errors of the estimates of  $\alpha$  and  $\omega$ , obtained from the numerically calculated (through `nlm`) Hessian matrix. As a measure of goodness-of-fit, data are simulated based on the parameter estimates, the parameters of the resulting data are estimated and the simulated data are compared with the original protein data both in terms of maximum likelihood estimates and in terms of the data themselves. Table 7.4 also displays the maximum likelihood estimates of the parameters for the simulated data  $\phi^{\text{sim}}$  and  $\psi^{\text{sim}}$ .

	$\phi$	$\phi^{\text{sim}}$	$\psi$	$\psi^{\text{sim}}$
$\hat{\alpha}$	-1.420 (0.062)	-1.390 (0.061)	2.214 (0.164)	2.166 (0.162)
$\hat{\omega}$	-0.004 (0.012)	0.006 (0.068)	0.554 (0.116)	0.499 (0.063)
$\hat{\kappa}$	2.783	3.100	1.467	1.476
$\bar{\theta}$	-1.420	-1.389	-0.140	2.165
AR(1)	0.006	-0.002	0.057	-0.087

Table 7.4: Maximum likelihood estimates (and estimated standard errors) for protein data and for data simulated based on these estimates. Also given are mean directions  $\bar{\theta}$  for each data set, and AR(1) parameter estimates of  $Y_t = \tan(\theta_t - \hat{\alpha})/2$ , where  $\theta$  is to be replaced by the variable of interest.

Also given in Table 7.4 are the mean direction  $\bar{\theta}$  of each data set and an estimate of the AR(1) parameter of the data  $Y_t = \tan(\theta_t - \hat{\alpha})/2$ , where  $\theta$  is to be replaced by the variable ( $\phi$ ,  $\phi^{\text{sim}}$ ,  $\psi$  or  $\psi^{\text{sim}}$ ) of interest. For the latter, if  $\kappa$  is large then  $\epsilon_t$  in the time series model

(7.6) will be small, in which case the model can be approximated by

$$\tan \frac{1}{2}(\theta_t - \alpha) \approx \omega \tan \frac{1}{2}(\theta_{t-1} - \alpha). \quad (7.34)$$

The parameter  $\omega$  in (7.34) is then the AR(1) parameter of the transformed data  $Y_t = \tan \frac{1}{2}(\theta_t - \alpha)$ . Comparison of  $\hat{\omega}$  and the AR(1) estimates in Tables 7.4 indicates that the  $\psi$  and  $\psi^{\text{sim}}$  data are too dispersed for  $\hat{\omega}$  and the AR(1) estimate to be comparable. For the  $\phi$  and  $\phi^{\text{sim}}$  data, both  $\hat{\omega}$  and the AR(1) estimate are seen to be close to zero. When  $\omega = 0$  the data are simply a random sample from a univariate von Mises distribution with mean  $\alpha$  and concentration  $\kappa$ . The closeness of  $\omega$  to zero for the  $\phi$  values therefore indicates little or no dependence between  $\theta_t$  and  $\theta_{t-1}$ , assuming the model is appropriate.

Figure 7.13 shows circular plots and time series plots of both the protein data and the simulated data. As can be seen from the plots, the protein  $\psi$  data are very different from the simulated  $\psi^{\text{sim}}$  data, indicating a poorly fitting model for the  $\psi$  values. There is less discrepancy between the protein data and the simulated data for the  $\phi$  series, although for these series differences are again apparent. For example, the protein  $\phi$  data appear to exhibit changing concentration throughout the series. On the other hand, not only do the simulated data not exhibit this feature, the Möbius time series model in general has constant concentration. This can be contrasted with the Sine and Cosine time series models, for which the concentration parameter of the conditional distribution of  $\theta_t$  given  $\theta_{t-1}$  is time-dependent.

Figure 7.14 shows profile loglikelihood plots, as described in Section 7.9 for simulated data. The loglikelihood profile plots in  $\omega$  for the protein data are markedly different from those encountered when the data were simulated, with sharp turning points at  $\omega = 0$  in each case. In the case of the  $\psi$  angles a maximum is difficult to identify visually due to the small gradient in its vicinity. For all other estimates of  $\alpha$  and  $\omega$ , there is a close agreement between the nlm estimates and the profile plots.

For the  $\phi$  angles, the profile plots in  $\alpha$  are very similar for the protein and simulated

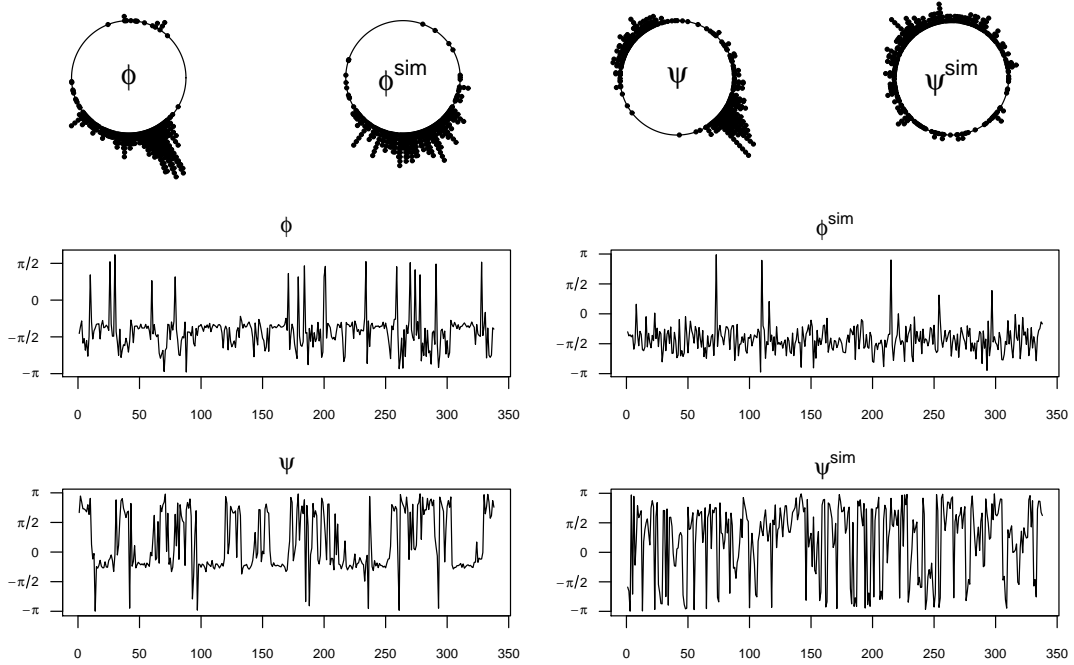


Figure 7.13: Circular and time series plots of the  $\phi$  and  $\psi$  angles of a polypeptide chain, and of data simulated using the parameter estimates obtained by fitting the Möbius model to the protein data.

data. For the  $\psi$  angles, we observe a local maximum at slightly less than  $\alpha = 0$  which corresponds to the mean direction as given in Table 7.4. From the circular plot of the  $\psi$  protein data in Figure 7.13, we clearly observe bimodality, which leads to the difficulty in identifying  $\alpha$  in this case. Again the large discrepancy between the plots for  $\psi$  and for  $\psi^{\text{sim}}$  would indicate a poorly fitting model for the  $\psi$  protein data. The discrepancy is again seen to be less for the  $\phi$  and  $\phi^{\text{sim}}$  series, although the profile in  $\omega$  for the protein data exhibits the aforementioned sharp turning point at  $\omega = 0$  that is not apparent for the simulated  $\phi^{\text{sim}}$  values.

Figures 7.15 and 7.16 show the goodness-of-fit plots developed in Chapter 6 for the

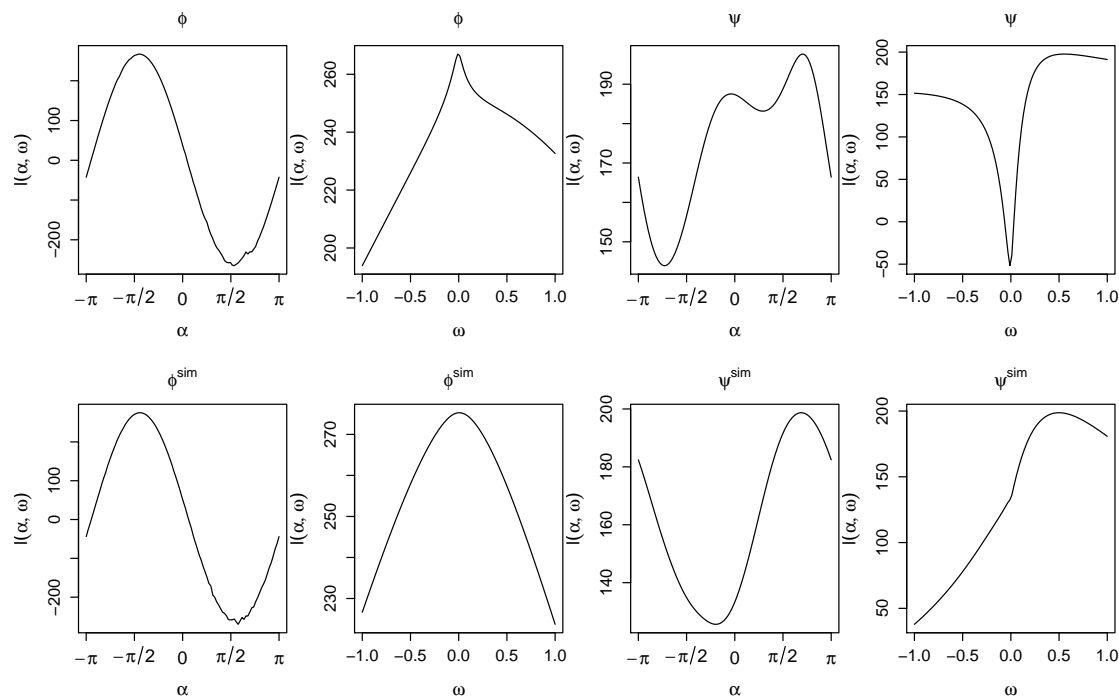


Figure 7.14: Profile loglikelihood plots for protein and simulated data. Profiles in each of  $\alpha$  and  $\omega$  taken at the maximum likelihood estimate of the other.

Sine and Cosine time series models. That is, Figure 7.15 shows plots of  $\theta_t - \hat{\alpha}$  versus  $\theta_{t-1} - \hat{\alpha}$  where  $\theta$  is to be replaced by the variable of interest and  $\hat{\alpha}$  is the maximum likelihood estimate of  $\alpha$  for that variable. Solid black lines represent the function  $\hat{\mu}_t = 2 \tan^{-1} \{\hat{\omega} \tan(\theta_{t-1}/2)\}$  whilst dashed black lines represent  $\hat{\mu}_t \pm 1/\sqrt{\hat{k}}$ . Finally, solid red lines represent a smoothed version of  $\theta_t - \hat{\alpha}$  versus  $\theta_{t-1} - \hat{\alpha}$  where the smoothing is performed using the function `circ.smth` described in Section 6.2.5.

From these plots we clearly have further evidence that the model is a very poor fit to the  $\psi$  protein data. For the  $\phi$  values, the red line approximates the black line reasonably well, although there is a slight curvature apparent in the former. This kind of curvature was observed for the fitted Sine and Cosine models in Figures 6.5 and 6.13 respectively, indicating that they could be competing models for these data. Although the Sine and Cosine models were fitted to a different protein data set than is currently being analysed,



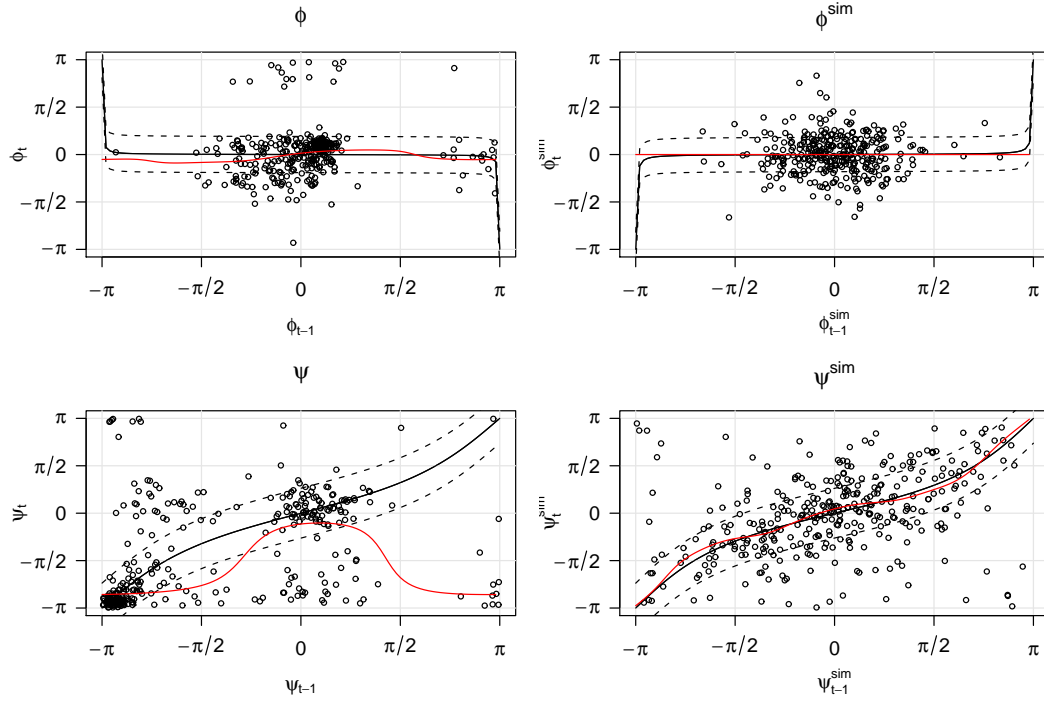


Figure 7.15: Plots of  $\theta_t$  versus  $\theta_{t-1}$ , a smoothed version (red line) and  $\hat{\mu}_t \pm 1/\sqrt{\hat{\kappa}_t}$ , where  $\theta$  is one of  $\phi, \phi^{\text{sim}}, \psi$  or  $\psi^{\text{sim}}$ .

comparing the plots in this section with those for the Sine and Cosine model reveal close similarities.

Figure 7.16 compares the (numerically calculated) equilibrium distributions based on the fitted models with kernel density estimates of the data themselves. Again we see evidence of a very poorly fitting model for the  $\psi$  values, with the kernel density estimate exhibiting the bimodality also apparent in the plots in Figure 7.13 for the  $\psi$  data. For the  $\phi$  values we observe a kernel density estimate that is slightly skewed and heavy in the tails, but which provides a reasonable approximation to the equilibrium distribution based on the fitted model.

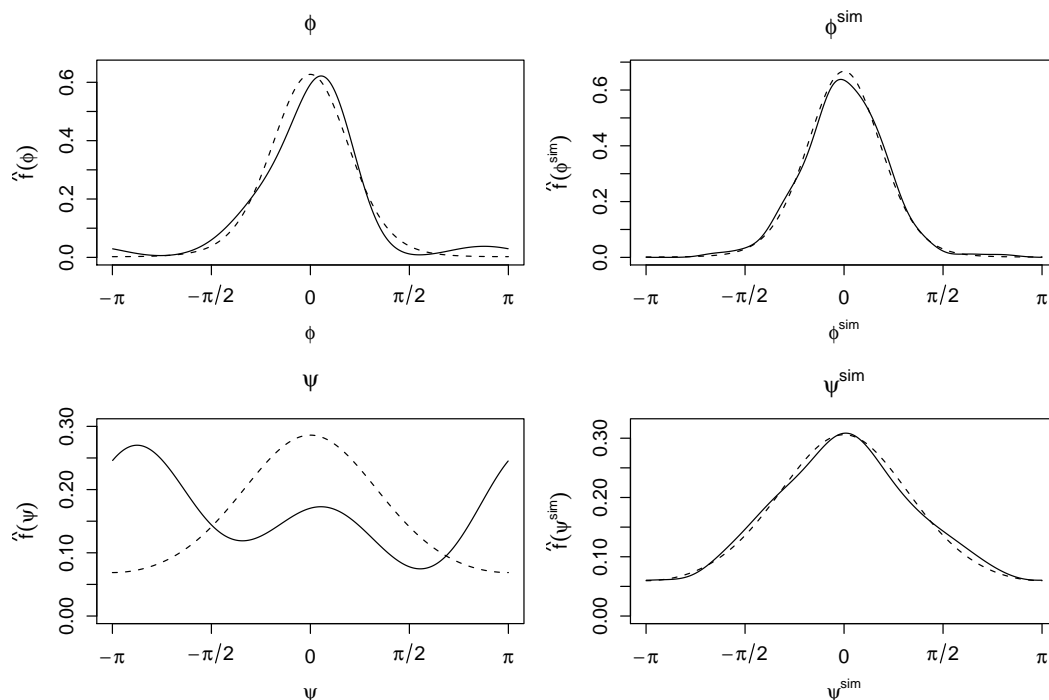


Figure 7.16: Numerically calculated equilibrium distribution based on parameter estimates (dashed lines) and kernel density estimates based on data (solid lines).

## 7.11 Conclusions

A circular regression model introduced by Downs and Mardia (2002) has been adapted to a circular time series context. The model properties have been investigated, including analysis of the deterministic component of the model. Data simulation and parameter estimation have been discussed, and successful methods have been developed in order to accomplish both. The potential problem of identifying  $\alpha$  when  $\omega$  is close to  $-1$  has been identified via analysis of the behaviour of the loglikelihood function. A thorough investigation into the equilibrium distribution has revealed similarities with a von Mises distribution. For the special cases  $|\omega| = 1$  and/or  $\kappa = 0$ , the equilibrium distribution is uniform on the circle, whilst if  $\omega = 0$  then it is von Mises with concentration  $\kappa$ .

The model has been fitted to the  $\phi$  and  $\psi$  angles of a protein data set, and has been shown

to be a very poor fit for the latter. For the  $\phi$  values, the fit appears reasonable and suggests that the dependence of  $\theta_t$  on  $\theta_{t-1}$  is at most very slight.

Having studied the Möbius model, the Sine time series model and the Cosine time series model, it is instructive to compare the properties of the three. This was done briefly in Sections 6.3.6 and 6.3.7 for the Sine and Cosine models, and we here compare the three models and their properties. These comparisons build on those of Hughes et al. (2005).

For comparison of the conditional means of the models, we will assume that  $\mu = 0$  for the Sine and Cosine models and that  $\alpha = 0$  for the Möbius model. We observed for the Sine model that the mean direction  $\mu_t$  of  $\theta_t$  given  $\theta_{t-1}$  is equal to  $\mu_t$  given  $\pi - \theta_{t-1}$ , which was considered an unusual property that may not be expected in an observed data set. A similar (but not exactly the same) thing was observed of  $\mu_t$  in the Cosine model when the values of the parameters  $a$  and  $b$  are such that  $b > |a|$ . On the other hand,  $\mu_t$  for the Möbius model was seen to be a monotonic (increasing or decreasing, depending on the sign of  $\omega$ ) function of  $\theta_{t-1}$ .

In terms of parameter interpretation, that for the Sine and Cosine models is made awkward with  $\lambda$  and  $\kappa$  for the Sine model and  $a$  and  $b$  for the Cosine model affecting both the deterministic and random components of the model. In contrast, the parameter  $\omega$  alone (taking  $\alpha = 0$ ) determines the behaviour of the deterministic component of the Möbius model, whilst  $\kappa$  determines the random component, making interpretation much simpler.

The equilibrium distributions of the Sine and Cosine models were seen to be symmetric around  $\mu$  and either unimodal or bimodal. For the numerically calculated equilibrium distribution for the Möbius model, only unimodal marginals were observed, although for values of  $\omega$  close to  $-1$  the overall series was seen to alternate between two “mini time series” at a distance of  $\pi$  apart.

One result of the relatively simple form of the Möbius model compared with the Sine and Cosine models is that the latter two result in a wider range of possible data, incorporating

the possibility of bimodality and varying concentration over time. The important question is whether or not these features are observed in real series. For the protein series, both changing concentration and bimodality were observed, but not in a way that was successfully captured by the models. For the protein data studied (which were seen to be reasonably similar), there isn't much to distinguish between the three models in terms of goodness-of-fit for the  $\phi$  values, but for the  $\psi$  values the Sine model was seen to be far superior to the other two.

It is clear that analysis of the individual models and comparison of the three can be extended by fitting the models to alternative data, as the protein data seem to exhibit behaviour that cannot be captured particularly well by any of the three time series models. The analysis of the three models in Part II of the thesis provides a strong foundation on which further analysis can be built.

Another obvious way in which the work in this chapter could be extended is by considering higher order models. The general  $\text{AR}(p)$  model then takes the form of Equation (5.4) with  $g(\cdot) = 2 \tan^{-1}(\cdot)$ . In particular, the  $\text{AR}(p)$  model is defined by the conditional distribution  $\theta_t | (\theta_{t-1}, \dots, \theta_{t-p}) \sim M(\mu_t, \kappa)$ ,  $t = p+1, \dots, n$ , where

$$\mu_t = \alpha + 2 \tan^{-1} \left\{ \omega_1 \tan \frac{1}{2}(\theta_{t-1} - \alpha) + \dots + \omega_p \tan \frac{1}{2}(\theta_{t-p} - \alpha) \right\}.$$

Again, the methods employed in this Chapter can be extended to investigate the behaviour of this higher order model.

## Chapter 8

### Conclusions

We have presented in each chapter of this thesis a section outlining the conclusions reached in that chapter and ways in which the work therein can be extended or improved. Just as it was insightful to consider in the introductory chapter material that was to recur throughout the thesis, so it is instructive to summarise retrospectively the conclusions drawn throughout the thesis that have a common theme.

In the abstract of the thesis the aims of both developing statistical methodologies for modeling circular data and applying these methodologies to protein conformational angles were presented. At this juncture it is appropriate to consider the relative success of these two aspects, and also to summarise the achievements of each. The above are the aims of this concluding chapter. We begin by considering the theoretical advancement of circular models, separate from any specific data and in terms of the thesis as a whole. Thereafter, the conclusions drawn concerning fitting multivariate and circular time series models to actual protein data are summarised.

## 8.1 Multivariate and time series models for circular data

As outlined in Chapters 2 and 5 respectively, there is a sparsity of multivariate circular models and circular time series models. There are various reasons for this, and some of them have been observed throughout the thesis. As previously noted, Kent and Mardia (2002) commented that intractability of at least one or other of the marginals or conditionals seems to be a universal law in directional models. We were unable to obtain an expression for the univariate marginal distribution when the trivariate von Mises distribution was considered. Even in the bivariate case, the marginals are not von Mises and can be unimodal or bimodal. The results of this fact were observed in time series modeling, with bimodal data obtained when simulating data from the Sine time series model for certain parameter values, and the same is true of the Cosine time series model. For the Möbius model the equilibrium distribution could not be obtained analytically, and numerical solutions were instead obtained.

Certain aspects of formulating multivariate and time series models were specific to the type of model being considered, since the former concentrated mainly on the joint density whilst the latter focused on the conditionals. For the multivariate von Mises model, parameter estimation provided the problem of an unknown normalising constant, and study of the pseudolikelihood provided valuable results regarding its efficiency both for the von Mises case and for the multivariate normal distribution. For the Sine and Cosine time series models, unusual behaviour of the mean direction of  $\theta_t$  given  $\theta_{t-1}$  was observed, and the features of time series models with changing concentration over time were considered. Of course, the features observed for the time series models are relevant to the understanding of the multivariate models (and vice versa), and, with the extension of the bivariate Cosine model to a multivariate context, the observations made in both parts of the thesis could contribute to a comparison of the multivariate Sine and multivariate Cosine models, extending the comparisons made by Mardia et al. (2007b)

for the bivariate case.

In spite of the difficulties associated with multivariate and time series circular models, progress has been made regarding the formulation of models. The multivariate von Mises distribution discussed has the appealing properties of following an approximate multivariate normal distribution under high concentration, and is a natural extension of the univariate von Mises distribution and bivariate Sine model. Circular time series models have been adapted from distributions in the literature, and for both the multivariate and time series models, the methods of parameter estimation presented were seen to be successful for simulated data. The important question remains regarding whether or not real data exhibit the features observed for these simulated data. Various goodness-of-fit measures have also been developed for the circular time series models, by comparing features of distributions based on the fitted model with those based on the data.

For all of the models developed and considered in the thesis, ways in which the analysis can be extended have been outlined. For the multivariate von Mises model, the alternative provided by the bivariate Cosine density for formulation of a multivariate von Mises model was discussed in Section 4.5. The development of this model would both provide the potential for the comparison with the multivariate model presented and also give a competing model. Each of the time series models presented has been discussed as a circular AR(1) model, and there is clearly scope for the extension of these models to AR( $p$ ) processes or even ARMA( $p, q$ ) processes. With these extensions the tools discussed in Section 5.2, such as the circular correlogram, could be employed in, for example, model selection.

## 8.2 Modeling protein data with the multivariate and time series models

Certain features of protein conformational angles data have caused difficulty in the application of the circular models formulated. The most notable of these is the multimodality or bimodality of data observed, for example, in Figures 4.8 and 6.4 respectively. In the case of the former, when the serine and valine data were considered, the problem was addressed by isolating a single mode of the data for analysis. For the latter (the  $\psi$  time series values), the fitted Sine and Cosine models were also bimodal, but were unable to capture the type of bimodality observed in the real data, which had different concentrations at each mode. The Möbius time series model, capturing only unimodal data, was less successful with regard to modeling the  $\psi$  angles. All models were a comparatively good fit to the  $\phi$  angles, which may be due in part to the fact that these data are less bimodal, and perhaps in part to the possibility that, as suggested by the fitted Möbius model, there is little association between  $\phi_t$  and  $\phi_{t-1}$ . A potential alternative could be to model certain regions of a polypeptide chain with the time series models, rather than the polypeptide in its entirety.

In respect of both the multivariate von Mises model and the circular time series models, just as the models themselves can be extended, so the existing models can be applied to different data sets to gauge their potential usefulness more generally, and in this regard there are numerous ways in which this work can be extended.



# Bibliography

- Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions*. Dover Publications.
- Arnold, B. C., Castillo, E., and Sarabia, J. M. (2001). Conditionally specified distributions: an introduction. *Statistical Science*, 16(3):249–265.
- Arnold, B. C. and Strauss, D. (1991). Pseudolikelihood estimation: some examples. *Sankhya series B*, 53(2):233–243.
- Baba, Y. (1981). Statistics of angular data: wrapped normal distribution model. *Proceedings of the Institute of Statistical Mathematics*, 28:41–54.
- Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician*, 24(3):179–195.
- Besag, J. (1977). Efficiency of pseudolikelihood estimation for simple gaussian fields. *Biometrika*, 64(3):616–618.
- Branden, C. and Tooze, J. (1998). *Introduction to Protein Structure*. Garland Publishing, New York.
- Breckling, J. (1989). *The analysis of directional time series: applications to wind speed and direction*. Springer Berlin.
- Chatfield, C. (1975). *The analysis of time series: theory and practice*. Chapman and Hall.

- Coles, S. (1998). Inference for circular distributions and processes. *Statistics and Computing*, 8(2):105–113.
- Cox, D. R. and Miller, H. D. (1980). *The Theory of Stochastic Processes*. Chapman and Hall.
- Cox, D. R. and Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika*, 91(3):729–737.
- Davison, A. C. (2003). *Statistical Models*. Cambridge University Press.
- Downs, T. D. and Mardia, K. V. (2002). Circular regression. *Biometrika*, 89(3):683–698.
- Dunbrack, R. L. and Cohen, F. E. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Science*, 6(8):1661–1681.
- Erwin, M. E., Cressie, N., and Johannesson, G. (2002). Spatial-temporal nonlinear filtering based on hierarchical statistical models. *Test*, 11(2):249–302.
- Feller, W. (1968). *An Introduction to Probability Theory and its Applications, Volume I*. Wiley.
- Fisher, N. and Lee, A. (1983). A correlation coefficient for circular data. *Biometrika*, 70(2):327–332.
- Fisher, N. and Lee, A. (1994). Time series analysis of circular data. *JRSS(B)*, 56(2):327–339.
- Fisher, N. I. (1993). *Statistical Analysis of Circular Data*. Cambridge University Press.
- Hughes, G., Mardia, K. V., and Taylor, C. C. (2005). Circular time series with application to protein conformations. In Barber, S., Baxter, P. D., Mardia, K. V., and Walls, R. E., editors, *Proceedings of the 24th LASR workshop: Quantitative Biology, Shape Analysis, and Wavelets*, pages 146–149.

- Hughes, G., Mardia, K. V., and Taylor, C. C. (2006). Modeling protein folds with a trivariate von Mises distribution. In Barber, S., Baxter, P. D., Mardia, K. V., and Walls, R. E., editors, *Proceedings of the 25th LASR workshop: Interdisciplinary Statistics and Bioinformatics*, pages 120–123.
- Jammalamadaka, S. R. and SenGupta, A. (2001). *Topics in Circular Statistics*. World Scientific.
- Kent, J. T. (1982). Robust properties of likelihood ratio tests. *Biometrika*, 69(1):19–27.
- Kent, J. T. and Mardia, K. V. (2002). (Discussion of) Spatial–temporal nonlinear filtering based on hierarchical statistical models. *Test*, 11(2):287–288.
- Lesk, A. M. (2000). *Introduction to Protein Architecture*. Oxford University Press.
- Lombard, F. (1988). Detecting change points by Fourier analysis. *Technometrics*, 30(3):305–310.
- Mardia, K. V. (1975). Statistics of directional data. *Journal of the Royal Statistical Society Series B*, 3(37):349–393.
- Mardia, K. V., Hughes, G., and Taylor, C. C. (2007a). Efficiency of the pseudolikelihood for multivariate normal and von Mises distributions. *Submitted to JRSS(B)*.
- Mardia, K. V. and Jupp, P. E. (1999). *Directional Statistics*. Wiley.
- Mardia, K. V. and Patrangenaru, V. (2005). Directions and projective shapes. *The Annals of Statistics*, 33(4):1666–1699.
- Mardia, K. V., Taylor, C. C., and Subramaniam, G. K. (2007b). Protein bioinformatics and mixtures of bivariate von Mises distributions for angular data. *Biometrics*, to appear.
- Porat, B. and Friedlander, B. (1986). Computation of the exact Information matrix of Gaussian time series with stationary random components. *IEEE transactions on acoustics, speech, and signal processing*, 34(1):118–130.

- Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *Molecular Biology*, 7:95–99.
- Rivest, L.-P. (1987). A distribution for dependent unit vectors. *Communications in Statistics*, A(17):461–483.
- Singh, H., Hnizdo, V., and Demchuk, E. (2002). Probabilistic model for two dependent circular variables. *Biometrika*, 89(3):719–723.
- Singh, H. and Mardia, K. V. (2004). A multivariate von Mises distribution. <http://www.maths.leeds.ac.uk/~sta6kvm/Singh.ps>.
- Singh, H., Mardia, K. V., Hughes, G., and Taylor, C. C. (2007). A multivariate von Mises distribution with applications to protein bioinformatics. *Submitted to Canadian Journal of Statistics*.
- Subramaniam, G. K. (2005). *Von Mises Distributions with Applications in Speech*. PhD thesis, University of Leeds.
- Wood, A. T. A. (1993). Estimation of the concentration parameters of the Fisher matrix distribution on  $SO(3)$  and the Bingham distribution on  $S_q$ ,  $q \geq 2$ . *Australian journal of statistics*, 35(1):69–79.