

BAYESIAN MODELLING FOR ASYMMETRIC MULTI-MODAL CIRCULAR
DATA

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MUHAMMET BURAK KILIÇ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
STATISTICS

AUGUST 2015

Approval of the thesis:

**BAYESIAN MODELLING FOR ASYMMETRIC MULTI-MODAL CIRCULAR
DATA**

submitted by **MUHAMMET BURAK KILIÇ** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Statistics Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Ayşen Dener Akkaya
Head of Department, **Statistics**

Assoc. Prof. Dr. Zeynep Kalaylıoğlu
Supervisor, **Statistics Department, METU**

Prof. Dr. Ashis SenGupta
Co-supervisor, **Indian Statistical Institute, Kolkata, India.**

Examining Committee Members:

Prof. Dr. Yılmaz Akdi
Statistics Department, Ankara University

Assoc. Prof. Dr. Zeynep Kalaylıoğlu
Statistics Department, METU

Prof. Dr. Birdal Şenoğlu
Statistics Department, Ankara University

Assoc. Prof. Dr. Barış Sürücü
Statistics Department, METU

Assoc. Prof. Dr. Oğuz Uzol
Aerospace Engineering Department, METU

Date:

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: MUHAMMET BURAK KILIÇ

Signature :

ABSTRACT

BAYESIAN MODELLING FOR ASYMMETRIC MULTI-MODAL CIRCULAR DATA

Kılıç, Muhammet Burak

Ph.D., Department of Statistics

Supervisor : Assoc. Prof. Dr. Zeynep Kalaylıođlu

Co-Supervisor : Prof. Dr. Ashis SenGupta

August 2015, 107 pages

In this thesis, we propose a Bayesian methodology based on sampling importance re-sampling for asymmetric and bimodal circular data analysis. We adopt Dirichlet process (DP) mixture model approach to analyse multi-modal circular data where the number of components is not known. For the analysis of temporal circular data, such as hourly measured wind directions, we join DP mixture model approach with circular times series modelling. The approaches are illustrated with both simulated and real-life data sets. Our Bayesian methodologies have been shown to have good statistical properties in multi-modal circular data analysis. Computational codes for DP mixture models are constructed in OpenBUGS and R.

Keywords: Directional Data, Dirichlet Process Mixture model, Asymmetry, Circular Time Series

ÖZ

ASİMETRİK ÇOKMODLU DAİRESEL VERİLER İÇİN BAYESÇİ MODELLEME

Kılıç, Muhammet Burak

Doktora, İstatistik Bölümü

Tez Yöneticisi : Doç. Dr. Zeynep Kalaylıođlu

Ortak Tez Yöneticisi : Prof. Dr. Ashis SenGupta

Ađustos 2015 , 107 sayfa

Bu tezde, asimetrik ve iki modlu veri analizi için, önem örneklemesine dayalı Bayesci bir yaklaşım önerdik. Karma sayısı bilinmeyen, çok modlu dairesel veriler için, Dirichlet süreç (DS) karma model yaklaşımını adapte ettik. Zamana bađlı olarak deđişen dairesel veri analizi için, örneđin rüzgar yönü, DS karma model yaklaşımını, dairesel zaman serileri modeli ile birleřtirdik. Bu yaklaşımlar, simulasyon ve gerçek veriler ile gösterildi. Sonuç olarak, önerdiğimiz yöntemler, çok modlu dairesel veri analizinde, iyi istatistiksel özelliklere sahip olduđu gösterilmiştir. Dirichlet süreci karma modeller için sayısal kodlar R ve OpenBUGS da yapıldı.

Anahtar Kelimeler: Dairesel Veri, Dirichlet Süreci Karma Model, Asimetri, Dairesel Zaman Serisi.

To my family

ACKNOWLEDGMENTS

I would like to thank my supervisor Assoc. Professor Zeynep Kalaylıođlu for her constant support, guidance and friendship. It was a great honour to work with her for the last four years and our cooperation influenced my academical view highly.

I would also like to thank Professor Ashis SenGupta for his support and guidance on my stay at both Riverside, USA, and Kolkata, India. While away from my home, he not only supported me on my research but also provided that I feel welcome and personally attended for my needs and problems. He also motivated and influenced me highly in scientific context.

A lot of people influenced and supported this work scientifically and their contribution were most valuable for me. Members of my dissertation exam committee Prof. Yılmaz Akdi, Prof. Birdal Őenođlu, Assoc. Prof. BarıŐ Sürücü and Assoc. Prof. Ođuz Uzol always gave valuable feedback for the progress of this work, and were not hesitant to warn me of the shortcomings or risks of my work. I would also like to thank Professor Daniel Jeske to give me opportunity for my research in Riverside, USA and Indian Statistical Institute, (ISI) Kolkata, India, respectively.

I would also like to thank The Scientific and Technological Research Council of Turkey (TÜBİTAK) and Faculty Development Programme (ÖYP) for supporting and funding my visits to USA and India during my Ph.D study.

I would like to thank specially to my wife Nuriye Őeyda. She always make me feel loved and cared. Finally, sincerest thanks to each of my family members for supporting and believing in me all the way through my academic life.

TABLE OF CONTENTS

ABSTRACT	v
ÖZ	vi
ACKNOWLEDGMENTS	viii
TABLE OF CONTENTS	ix
LIST OF TABLES	xiv
LIST OF FIGURES	xv
LIST OF ALGORITHMS	xviii
LIST OF ABBREVIATIONS	xix
CHAPTERS	
1 INTRODUCTION	1
1.1 Motivating examples	2
1.2 Outline	4
2 CIRCULAR DATA	7
2.1 Basic descriptive statistics for circular data	7
2.2 Properties of circular distributions	9
2.3 Review of common circular distributions	10

2.4	Multi-modal models	13
3	A BAYESIAN ANALYSIS FOR ASYMMETRIC AND BIMODAL CIRCULAR DATA	15
3.1	Introduction	15
3.2	Some properties of two sub-models of generalisations of von Mises distribution and their joint conjugate and constrained priors	17
3.2.1	Generalised von Mises distribution	18
3.2.1.1	Conjugate prior distribution for GvM	19
3.2.1.2	Constrained joint prior distribution for GvM	19
3.2.2	Asymmetric generalised von Mises distribution	19
3.2.2.1	Conjugate Prior distribution	20
3.2.2.2	Constrained joint prior distributions for AGvM	20
3.3	Bayesian analysis for GvM and AGvM with SIR	21
3.4	Model selection	23
3.5	Real data examples	24
3.5.1	Spawning time of fish	24
3.5.2	Movement of turtle	28
3.6	Concluding remarks	30
4	BAYESIAN SEMI-PARAMETRIC MODELS FOR MULTI-MODAL CIRCULAR DATA	31
4.1	Introduction	31

4.2	DP mixture models	34
4.2.1	Stick breaking construction	35
4.3	DP mixture circular models with stick breaking construction .	36
4.3.1	DP mixture von Mises model	36
4.3.2	DP mixture wrapped Cauchy model	37
4.3.3	Inference via Gibbs sampler	37
4.4	Applications	39
4.4.1	Simulated data examples	39
4.4.2	Monte Carlo study	41
4.4.3	Real data examples	49
4.4.3.1	Turtle data	49
4.4.3.2	Ant data	53
4.5	Discussion	56
5	BAYESIAN SEMI-PARAMETRIC MODEL FOR MULTI-MODAL CIRCULAR TIME SERIES DATA	57
5.1	Introduction	57
5.2	Review of circular time series models	59
5.2.1	Linked process	59
5.2.2	Circular autoregressive process	59
5.2.3	Wrapped process	60
5.2.4	Projected Normal process	60
5.2.5	Möbius time series model	60

5.3	DP mixture model for circular time series	61
5.3.1	DP mixture Möbius model	62
5.4	Examples	64
5.4.1	Simulated data example	64
5.4.2	Real data examples	68
5.4.2.1	Wind directions in Australia	68
5.4.2.2	Wind directions in Turkey	70
5.5	Discussion	75
6	CONCLUSION	77
	REFERENCES	79
	APPENDICES	
A	APPENDIX FOR CHAPTER 3	85
A.1	Posterior distribution-Conjugacy for Generalised von Mises distribution	85
A.1.1	Posterior distribution	86
A.2	Posterior distribution-Conjugacy for Asymmetric Generalised von Mises distribution	86
A.2.1	Posterior distribution	87
A.3	Constrained joint prior distribution for dependent parameters of GvM	87
A.3.1	Bivariate exponential conditionals distribution	87
A.4	Constrained joint prior distributions of dependent parame- ters for AGvM	88

A.4.1	Bivariate beta distribution	88
A.4.2	Bivariate Dirichlet distribution	89
A.4.3	Bivariate beta conditionals distribution	89
B	APPENDIX FOR CHAPTER 4	93
B.1	Posterior computation for DP mixture von Mises model	93
B.2	Posterior computation for DP mixture wrapped Cauchy model	94
B.3	OpenBUGS codes	95
C	APPENDIX FOR CHAPTER 5	99
C.1	Circular-Circular association	99
C.2	Posterior computation for DP mixture Möbius model	100
C.3	OpenBUGS codes	101
	CURRICULUM VITAE	105

LIST OF TABLES

TABLES

Table 3.1	Five SIR runs and posterior mean estimates of both AGvM and GvM	26
Table 3.2	Prior selection for turtle data	29
Table 3.3	Comparison of the models for turtle data	29
Table 4.1	Posterior means of the mixing probabilities and parameters of the mixture for simulated data from three mixture vM distribution	40
Table 4.2	Posterior means of the mixing probabilities and parameters of the mixture for simulated data from three mixture wC distribution	41
Table 4.3	Monte Carlo study results for DP mixture vM model ($C = 2$)	46
Table 4.4	Monte Carlo study results for DP mixture vM model ($C = 3$)	46
Table 4.5	DP mixture vM model fits for Monte Carlo study	47
Table 4.6	Monte Carlo study results for DP mixture wC model ($C = 2$)	50
Table 4.7	Monte Carlo study results for DP mixture wC model ($C = 3$)	50
Table 4.8	DP mixture wC model fits for Monte Carlo study	51
Table 4.9	Estimates of parameters for turtle data	52
Table 4.10	Posterior means of the mixing probabilities and parameters of turtle data	54
Table 4.11	Posterior means of the mixing probabilities and parameters of ant data	55
Table 5.1	Comparison with model selection criterion	75

LIST OF FIGURES

FIGURES

Figure 1.1 Rose diagram of turtle data	2
Figure 1.2 (a) Rose diagram of wind direction data. (b) Circular observed time series of a hourly wind direction data from Turkey.	3
Figure 1.3 Dihedral angles of the backbone a protein. <i>Figure is adopted from "http://www.bioinf.org.uk/teaching/bbk/molstruc/practical2/peptide.html"</i>	4
Figure 2.1 von Mises densities with mean direction $\mu = 0^\circ$ and different concentration parameters $\kappa = 1, 2, 7$ and 10	11
Figure 3.1 Rose diagram of time of low tide	16
Figure 3.2 Bayesian estimation of AGvM distribution: the vertical red line shows maximum likelihood estimates, the blue line shows Bayesian estimates of the posterior means	27
Figure 3.3 Bayesian estimation of GvM distribution: the vertical red line shows maximum likelihood estimates, the blue line shows Bayesian estimates of the posterior means	27
Figure 3.4 (a) Comparison of model fits for turtle data (b) Posterior distribution for Bayesian model averaged for turtle data	29
Figure 4.1 Rose diagram of turtle data	32
Figure 4.2 Comparison of mixture DP vM model and kernel density estimation and true density for simulated data from three mixture vM distribution	42
Figure 4.3 Comparison of mixture DP wC model and kernel density estimation and true density for simulated data from three mixture wC distribution	42
Figure 4.4 Rose diagram of two mixture vM data	43

Figure 4.5	Boxplots of estimated circular mean directions for two mixture vM distributions	44
Figure 4.6	Boxplots of estimated concentration parameters for two mixture vM distributions	44
Figure 4.7	Boxplots of estimated weight parameters for two mixture vM distributions	45
Figure 4.8	Rose diagram of two mixture wC data	47
Figure 4.9	Boxplots of estimated circular mean directions for two mixture wC distributions	48
Figure 4.10	Boxplots of estimated concentration parameters for two mixture wC distributions	48
Figure 4.11	Boxplots of estimated weight parameters for two mixture wC distributions	49
Figure 4.12	(a) Comparison of mixture DP vM model and kernel density estimation for turtle data. (b) Identified clusters for turtle data	51
Figure 4.13	Comparison of mixture DP and kernel density estimation for ant data.	53
Figure 5.1	Rose diagram of a hourly wind direction data from Turkey	58
Figure 5.2	(a) Rose diagram of simulated Möbius time series data. (b) Plot of simulated Möbius time series data	65
Figure 5.3	(a) Plot of direction of the simulated circular time series data (b) Plot of direction of the predicted circular time series data	66
Figure 5.4	Posterior densities of all parameter of DP mixture Möbius model for simulated data	67
Figure 5.5	(a) Sample circular autocorrelations for the time series of wind directions in Australia (b) Plot of observed circular time series data in Australia	68
Figure 5.6	(a) Rose diagram of wind direction data. (b) Rose diagram of predicted wind direction data.	70
Figure 5.7	(a) Sample circular autocorrelations for the time series of wind directions in Turkey (b) Plot of observed circular time series data in Turkey	70

Figure 5.8	Posterior densities of parameters of DP Möbius model for wind direction data from Australia.	71
Figure 5.9	Posterior density of K latent variables which belong to second cluster	72
Figure 5.10	(a) Rose diagram of wind direction data in Turkey (b) Rose diagram of predicted wind direction data in Turkey	74
Figure 5.11	(a) Plot of direction of the observed circular time series in Turkey (b) Plot of direction of the predicted circular time series in Turkey	74

LIST OF ALGORITHMS

ALGORITHMS

Algorithm 1	Simulation of bivariate exponential conditionals distribution . .	88
Algorithm 2	Simulation of bivariate beta distribution	91
Algorithm 3	Simulation of bivariate Dirichlet distribution	91
Algorithm 4	Simulation of bivariate beta conditionals distribution	91

LIST OF ABBREVIATIONS

AIC	Akaike Information Criteria
AGvM	Asymmetric Generalised von Mises
BIC	Bayesian Information Criteria
CAR(1)	Circular Autoregressive Order One
DP	Dirichlet Process
GvM	Generalised von Mises
HMM	Hidden Markov Model
KL	Kullback Leibler Divergence
MACE	Mean Absolute Cosine Error
MAP	Maximum at Posterior
MCDE	Mean Cosine Difference Error
MCMC	Markov Chain Monte Carlo
MCSE	Monte Carlo Standard Error
R.Bias	Relative Bias
SE	Standard Error
SIR	Sampling Importance Re-sampling
WAR	Wrapped Autoregressive Process
vM	von Mises
wC	wrapped Cauchy
wN	wrapped Normal

CHAPTER 1

INTRODUCTION

Circular data exist in many scientific contexts such as medicine, ecology, meteorology and biology. In many environmental and medicine applications, there arise multi-modal circular data. To address multi-modality seen in various different applications changing from finance to astrophysics, in recent years, there has been an increasing interest in developing statistical analyses for asymmetric and multi-modal circular distributions. In general there are two solutions for multi-modal circular data problems as follows

- One possible solution is to use asymmetric and multi-modal circular distributions. Main challenge for these distributions is the normalizing constants not having a closed form expression.
- Another solution is to use mixture circular distributions. The main problem of mixture circular distributions is that the number of modes may not be precisely determined based on the available data sets.

Circular data analysis is more challenging than linear data analysis due to the restriction of the support on the unit circle $[0, 2\pi)$ or $[-\pi, \pi)$ and to the sensitivity of descriptive and inferential statistics on the unit circle. However, there are various substantial methods and techniques for analysing circular data. (see, e.g. [Mardia \(1972\)](#); [Fisher \(1993\)](#); [Mardia and Jupp \(1999\)](#); [Jammalamadaka and SenGupta \(2001\)](#)). These methods are broadly represented for simple circular univariate models. On the other hand, on Bayesian circular data modelling, there has been small literature. (see, e.g. [Coles, \(1988\)](#); [Damien and Walker \(1999\)](#); [George and Ghosh \(2006\)](#); [Bhattacharya](#)

and SenGupta (2009); Lasinio, et al. (2012); Wang and Gelfand (2013); Antonio et al. (2014)).

In scope of this dissertation, we mainly focus on multi-modal circular data and develop the flexible novel methods in analysing multi-modal circular data. In the following section, we illustrate multi-modal circular data examples from ecology (study of animal movement), meteorology (wind direction) and bio-informatics (dihedral angles).

1.1 Motivating examples

In this section, we illustrate circular real data examples to motivating our study.

Turtle data

A particular example is a study of animal movement. For illustration, we consider a turtle data by Gould's cited by Stephens, (1969). Fig 1.1 shows that most of turtles move in one main direction, while small part of them moved in different directions. Here, the main problem is the unknown number of modes for this kind of data types. One may consider fitting models with different number of modes and assessing the goodness of the fit of these models. However, this is hindered by the fact that there is lack of goodness of fit test for multi-modal circular data.

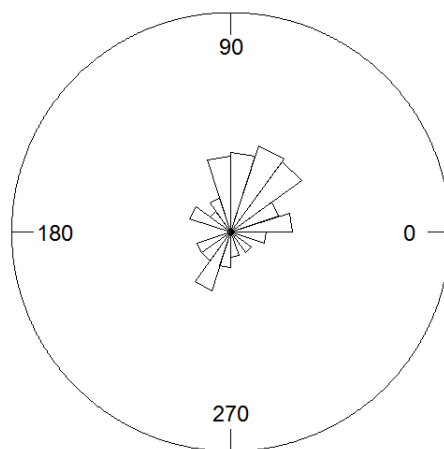


Figure 1.1: Rose diagram of turtle data

Wind direction data

Another particular example, Fig 1.2 displays a circular time series plot corresponding to hourly wind direction collected on three days in Turkey. Here, there arise uncertainty problem associated with number of modes in this time series modelling. For multi-modal circular data depended on time, we considered a new class of circular time series model.

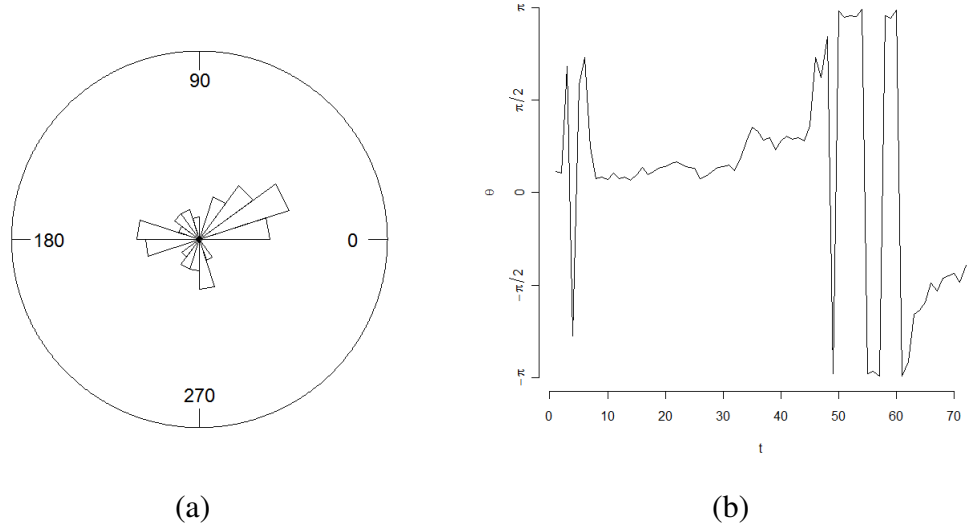
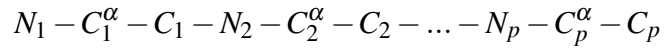


Figure 1.2: (a) Rose diagram of wind direction data. (b) Circular observed time series of a hourly wind direction data from Turkey.

Protein data

Another particular example consists of protein backbone data set from bio-chemical processes. Proteins play an important role in living organisms. A protein molecule consist of a chain or sequence of amino acids. Dihedral angles define the backbone of a protein. The protein is a polypeptide chain comprise of amino acids. The backbone of a polypeptide chain consist of a sequence of atoms.



The protein backbone can be explained as three dihedral angles, namely, ϕ , ψ and ω angles. ϕ angle is the angle around the $-N - C_\alpha$ bond. ψ angle is the angle around $-C_\alpha - C$ bond. ω angle is the angle around the $-C - N-$ bond (see Fig. 1.3). In particular, ψ and ω angles can be observed as multi-modal data (see [Hughes \(2007\)](#); [Durán and Domínguez \(2014\)](#)).

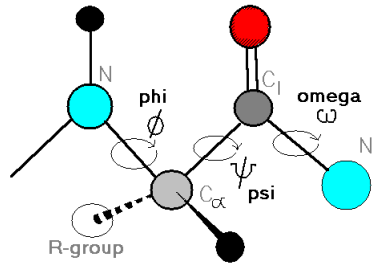


Figure 1.3: Dihedral angles of the backbone a protein. *Figure is adopted from* "<http://www.bioinf.org.uk/teaching/bbk/molstruc/practical2/peptide.html>"

In following section, we provide a summary of the whole chapters of the dissertation and emphasize the most important components of it.

1.2 Outline

In Chapter 2, we give a brief summary of circular data and explain the basic properties of them such as measurements of location, concentration and the other important descriptive statistics. Then, we discuss the modelling approaches for circular data. At the end of this chapter, I present asymmetric and bimodal circular distributions which are based on the generalisations of von Mises distribution.

In Chapter 3, we facilitate a Bayesian analysis of bimodal distributions based on the extension of von Mises distribution. The most important challenge of these distributions in terms of Bayesian analysis is the complex normalizing constants. Another challenge is to draw the samples from their complex posterior distributions. In order to overcome these challenges, we provide a general way to facilitate their Bayesian analysis with *sampling importance re-sampling*, SIR. Meanwhile, real data examples are used to illustrate the usefulness of the proposed approach.

In Chapter 4, we propose to adopt Dirichlet process (DP) for independent identical distributed (i.i.d) circular observations. Here, the main challenge is determining the number of modes. In many studies, the number of the modes is unknown a priori. This leads to an uncertainty about the true number of modes. In order to handle this challenge, we use circular Bayesian non parametric models. These models detect the number of the modes when it is unknown. Finally, simulated and real data examples are presented to illustrate the flexibility of the proposed models.

In Chapter 5, we first give a summary of the existing methods. One of these existing models refers to circular autoregressive model with order (1), CAR(1) or Möbius model. First drawback of these models is the potential identifying problem in model parameters when multi-modality or changing concentration over time are observed. Second drawback is that these models give poor fit to multi-modal circular data. In order to handle these drawbacks, we introduce a new class of circular time series model based on DP mixture models. Potential identifying problem is handled with a constraint in Bayesian panorama. In addition, our model assumes that the conditional distribution of the model is the mixture von Mises distribution. This provides the flexibility of the proposed approach in terms of its error distribution. Numerical and real data examples are provided to illustrate the plausibility of the proposed model.

Computational details and technical results are explained in chapter appendices at the end of thesis. In all of the computational implementations, we used R language and environment and OpenBUGS software which is a useful and efficient tool for Bayesian computations.

CHAPTER 2

CIRCULAR DATA

Circular statistics is a special branch of statistics that is used to analyse data which can be mapped onto the circumference of a unit circle such as directional observations in the study of wind directions and orientations in turtles. In circular data analysis, standard linear statistical methods are not appropriate because of the geometric shape underlying the data. For illustration, let 1° and 359° be two independent circular observations. It is obviously not appropriate to use the standard arithmetic mean that is equal to 180° . The circular mean direction is equal to 0° . In this chapter, we summarize circular statistics and distributions.

2.1 Basic descriptive statistics for circular data

Corresponding to the usual descriptive statistics for linear variables, there exist descriptive statistics for circular variables. Most basic descriptive statistics that are fundamental to all the subsequent circular data analysis and modelling are sample circular mean, measure of concentration of the data, and sample circular variance. Circular mean is the mean of the directions and a simple formalization is provided below. Measure of concentration and sample circular variance are measures of concentration and dispersion respectively and given below. Interested readers should refer to [Jammalamadaka and SenGupta \(2001\)](#) (Topics in circular statistics) as well as [Mardia and Jupp \(1999\)](#) (Directional statistics) for more.

To find sample circular mean direction, we need to use vector summation properties. For instance, let $\theta_1, \theta_2, \dots, \theta_n$ be a set of circular observations represented as points

on the circumference of the unit circle. We transform each data point from the polar coordinate to the Cartesian coordinate to obtain $(\cos\theta_i, \sin\theta_i)$, $i = 1, \dots, n$ and sum them up to obtain the resultant vector R shown below

$$R = \left(\sum_{i=1}^n \cos\theta_i, \sum_{i=1}^n \sin\theta_i \right) = (C, S)$$

Then length of resultant vector is given by

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

The direction of this resultant vector R is the sample circular mean direction denoted by $\bar{\theta}$ which is shown by:

$$\bar{\theta} = \arctan^*(S/C) = \begin{cases} \arctan(S/C), & \text{if } C > 0, S \geq 0, \\ \pi/2, & \text{if } C = 0, S > 0, \\ \arctan(S/C) + \pi, & \text{if } C < 0, \\ \arctan(S/C) + 2\pi, & \text{if } C \geq 0, S < 0, \\ \text{undefined}, & \text{if } C = 0, S = 0. \end{cases}$$

given in [Jammalamadaka and SenGupta, \(2001\)](#).

Circular concentration is given by the L2 norm of the mean resultant vector $\bar{R} = (\bar{C}, \bar{S})$ where $\bar{C} = \frac{1}{n} \sum_{i=1}^n \cos\theta_i$ and $\bar{S} = \frac{1}{n} \sum_{i=1}^n \sin\theta_i$ as

$$\bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2}, \quad 0 \leq \bar{R} \leq 1 \quad (2.1)$$

\bar{R} being close to 1 indicates that vectors (i.e. the directional data) are concentrated around their mean vector (circular mean). A sample median direction denoted by $\tilde{\theta}$ is defined as any angle ψ for which half of data points lie in the arc $[\psi, \psi + \pi)$ and the majority of the points are nearer to ψ than to $\psi + \pi$. From this definition, it is clear that the median direction need not be unique. When n is odd, a median direction will correspond to one of the data points. When n is even, it is usually taken to be the mean of those data points. Formally, a median direction can be identified by minimizing the dispersion measure

$$d(\psi) = \frac{1}{n} \sum_{i=1}^n (\pi - |\pi - |\theta_i - \psi||).$$

Simplest form of circular sample variance is defined as $V = 1 - \bar{R}$ and sample circular standard deviation is $\hat{\sigma} = \{-2\log(1 - V)\}^{1/2}$. The higher order sample moments is defined by taking higher power of $\{e^{i\theta_j}\}$ and averaging these. We can write as

$$\begin{aligned}\frac{1}{n} \sum_{j=1}^n (e^{i\theta_j})^p &= \frac{1}{n} \sum_{j=1}^n e^{ip\theta_j} \\ &= \frac{1}{n} \sum_{j=1}^n \cos p\theta_j + i \frac{1}{n} \sum_{j=1}^n \sin p\theta_j \\ &= \bar{C}_p + i\bar{S}_p, \quad p = 0, 1, 2, \dots,\end{aligned}$$

where (\bar{C}_p, \bar{S}_p) are called as the p^{th} order trigonometric moments based on the sample. These calculations are needed to for computing posterior means and standard deviations from Markov Chain Monte Carlo (MCMC) output.

2.2 Properties of circular distributions

A circular probability density function is similar to a continuous probability density function on real line. The range of a circular random variable θ which is measured in radians is taken as $[0, 2\pi)$ or $[-\pi, \pi)$. A circular probability density function satisfies the following properties

- (a) $f(\theta) \geq 0$;
- (b) $\int_0^{2\pi} f(\theta)d\theta = 1$ or $\int_{-\pi}^{\pi} f(\theta)d\theta = 1$;
- (c) $f(\theta) = f(\theta + 2\pi k)$ for any integer k .

The characteristic function of a circular random variable θ is given by

$$\varphi_{\theta}(p) = E(e^{ip\theta}) = \int_0^{2\pi} e^{ip\theta} dF(\theta) = C_p + iS_p = \rho_p e^{i\mu_p}, \quad p = 0, \pm 1, \pm 2, \dots$$

where $C_p = E(\cos p\theta)$, $S_p = E(\sin p\theta)$, $\rho_p = \sqrt{C_p^2 + S_p^2}$ and $\mu_p = \arctan^*(S_p/C_p)$. The value of the characteristic function at an integer p is called the p^{th} trigonometric moment of θ . In particular, consider the first trigonometric moment as

$$\varphi_1 = C_1 + iS_1 = \rho_1 e^{i\mu_1},$$

where μ_1 is defined as mean direction denoted by μ and ρ_1 is defined as mean resultant length denoted by ρ . The length of ρ lies between 0 and 1 due to the inequality of expectation and the characteristic complex number, $0 \leq \|E(e^{i\theta})\| \leq E\|e^{i\theta}\| = 1$. This first trigonometric moment φ_1 are used to provide population measures of the mean direction and the concentration of θ , respectively. Again the sample analogues of μ and ρ are $\bar{\theta}$ and \bar{R} respectively, described in Section 2.1.

2.3 Review of common circular distributions

There are several circular distributions. Most common are uniform, von Mises, wrapped normal family, projected normal and generalisations of von Mises distributions.

Uniform distribution on circle is defined as follows

$$f(\theta) = \frac{1}{2\pi}, \quad 0 \leq \theta \leq 2\pi \quad (2.2)$$

Length of the first trigonometric moment of this distribution is 0, therefore there is no preferred mean direction. This corresponds to a situation where all directions are equally likely. This distribution is used to test the hypothesis about the uniformity of directions. For instance, testing the null hypothesis that the orientation of a newly born turtle has no particular direction is equivalent to testing that distribution of the orientation is circular uniform.

Most commonly used model for circular data is von Mises distribution denoted by $vM(\mu, \kappa)$ (also known as circular normal distribution) and this distribution is symmetric and uni-modal. Its pdf is defined as shown below

$$f(\theta) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta - \mu)}, \quad 0 \leq \mu \leq 2\pi, \quad \kappa > 0 \quad (2.3)$$

where μ is mean direction and κ is the concentration parameter. $I_0(\kappa)$ is modified Bessel function of the first kind and order zero. The mean resultant length have shown to have the expression $I_1(\kappa)/I_0(\kappa)$, where I_p is the modified Bessel function of first kind of order p which is given by

$$I_p(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \cos p\theta e^{\kappa \cos\theta} d\theta$$

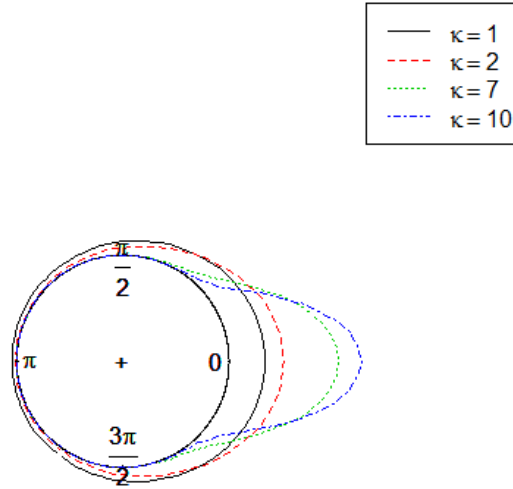


Figure 2.1: von Mises densities with mean direction $\mu = 0^\circ$ and different concentration parameters $\kappa = 1, 2, 7$ and 10 .

For $0 \leq \theta < 2\pi$, values of the distribution function of the von Mises distribution are given by

$$F(\theta) = \frac{1}{2\pi I_0(\kappa)} \int_0^\theta e^{\kappa \cos(\phi - \mu)} d\phi$$

the computation of both the Bessel function and the integral requiring quadrature. Main problem with the use of von Mises distribution is modified Bessel function which is not available in closed form. [Abramowitz and Stegun \(1970\)](#) give the expression of the function as shown below

$$I_0(\kappa) = \sum_{r=0}^{\infty} (r!)^{-2} \left(\frac{1}{2} \kappa\right)^{2r}. \quad (2.4)$$

They also give polynomial approximations to $I_0(\kappa)$ as :

$$I_0(\kappa) \simeq 1 + 3.5156229t^2 + 3.0899424t^4 + 1.2067492t^6 + 0.2659732t^8 + 0.0360768t^{10} + 0.0045813t^{12}, \quad 0 \leq \kappa \leq 3.75, \quad t = \kappa/3.75$$

For large κ , this Bessel function is approximated by

$$I_0(\kappa) \sim \frac{e^\kappa}{\sqrt{2\pi\kappa}}$$

and we use the above approximations to evaluate the value of Bessel function in the preceding sections.

Another commonly used models for the analysis of circular data are wrapped family distributions. Wrapped family distributions are obtained by wrapping the line around

the circumference of the circle. For illustration, let Y be a random variable on \mathbb{R} with probability density function $g(y)$, then corresponding random variable on the circle is

$$\theta = Y \bmod 2\pi \quad (2.5)$$

Probability density function of wrapped distribution is

$$f(\theta) = \sum_{k=-\infty}^{\infty} g(\theta + 2\pi k)$$

Below, we briefly give the two important class of wrapped family distribution as follows

A wrapped normal distribution $\text{WN}(\mu, \rho)$ is obtained by wrapping a $\text{N}(\mu, \sigma^2)$ distribution around the circle.

$$f(\theta) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} \exp\left\{-\frac{(\theta - \mu + 2\pi k)^2}{2\sigma^2}\right\}, \quad 0 \leq \theta < 2\pi$$

where $\sigma^2 = -2\log\rho$. This distribution $\text{WN}(\mu, \rho)$ is also uni-modal and symmetric as $\text{N}(\mu, \sigma^2)$.

Wrapped Cauchy (WC) distribution is defined by [Lévy \(1939\)](#) and the probability density function of $\text{WC}(\mu, \rho)$ is as follows

$$\begin{aligned} f(\theta) &= \frac{1}{2\pi} \left(1 + 2 \sum_{k=1}^{\infty} \rho^k \cos k(\theta - \mu)\right) \\ &= \frac{1}{2\pi} \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(\theta - \mu)} \end{aligned}$$

where $0 \leq \mu < 2\pi$ and $0 \leq \rho < 1$. This distribution is both uni-modal and symmetric and is used for circular distributions with heavy peaks.

Another family of circular distributions can be obtained by radial projection of bivariate distributions on the plane. Let \mathbf{Y} be a two dimensional random vector with $P(\mathbf{Y} = \mathbf{0}) = 0$. Then obviously, $\|\mathbf{Y}\|^{-1}\mathbf{Y}$ is a random point on the unit circle. If, \mathbf{Y} has a bivariate normal distribution $N_2(\mu, \Sigma)$, then $\|\mathbf{Y}\|^{-1}\mathbf{Y}$ has a projected normal distribution denoted by $PN_2(\mu, \Sigma)$. This distribution is known as offset normal distribution. (see e.g. [Mardia \(1972\)](#); [Jammalamadaka and SenGupta 2001](#)). Probability density function of a general projected normal distribution is defined as shown below

$$f(\theta) = \frac{1}{C(\theta)} \left\{ \phi(\mu_1, \mu_2, 0, \Sigma) + aD(\theta)\Phi[D(\theta)]\phi\left[\frac{a(\mu_1 \sin\theta - \mu_2 \cos\theta)}{\sqrt{C(\theta)}}\right] \right\} \quad (2.6)$$

where

$$\begin{aligned}
a &= \{\sigma_1 \sigma_2 \sqrt{1 - \rho^2}\}^{-1} \\
C(\theta) &= a^2(\sigma_2^2 \cos^2(\theta) - \rho \sigma_1 \sigma_2 \sin 2\theta + \sigma_1^2 \sin^2 \theta), \\
D(\theta) &= \frac{a^2}{\sqrt{C(\theta)}} [\mu_1 \sigma_2 (\sigma_2 \cos \theta - \rho \sigma_1 \sin \theta) + \mu_2 \sigma_1 (\sigma_1 \sin \theta - \rho \sigma_2 \cos \theta)] \quad (2.7)
\end{aligned}$$

and ϕ and Φ are the pdf and cdf of $N(0, 1)$, respectively.

A special case of a general projected normal distribution with zero mean and variance denoted by $PN_2(0, \Sigma)$ leads to

$$f(\theta) = \frac{\sqrt{(1 - \rho^2)}}{2\pi(1 - \rho \sin 2\theta)} \quad (2.8)$$

There is a relationship between wrapped Cauchy and projected normal distributions. [Kent and Tyler \(1988\)](#); [Mardia\(1972\)](#) showed that

$$\theta \sim PN_2(0, \Sigma) \Rightarrow 2\theta \sim WC(\mu, \rho) \quad (2.9)$$

General projected normal distribution has not been in common use for circular data modelling because of complicated and unwieldy expression. However, in recent years there are Bayesian developments for using general projected normal for the analysis of circular data (see [Wang and Gelfand 2012](#)).

2.4 Multi-modal models

General approach to modelling multi-modal circular data is to use finite mixtures of any uni-modal distributions considered in Section 2.3. One of the important features of mixture distribution is that their parameters are generally easy to interpret. Most commonly used models are finite mixtures of von Mises distributions. (see e.g. [Mardia and Sutton \(1975\)](#); [Spurr \(1981\)](#) etc.)

Another useful way is generalisations of von Mises distribution which can be obtained by expanding in Fourier series form. These distributions have an extensive history and some of the most relevant references are [Maksimov \(1967\)](#); [Cox \(1975\)](#); [Yfantis and Borgman \(1982\)](#); [Gatto and Jammalamadaka \(2007\)](#); [Kim and SenGupta \(2013\)](#). For

illustration, we consider the natural extension of von Mises distribution introduced by [Cox, D.R \(1975\)](#) for two components as follows

$$\propto \exp(a_1 \cos \theta + b_1 \sin \theta + a_2 \cos 2\theta + b_2 \sin 2\theta)$$

Taking $a_2 = 0$, $b_2 = 0$ give von Mises density as

$$\propto \exp(\kappa \cos(\theta - \mu))$$

where $a_1 = \kappa \cos \mu$ and $b_1 = \kappa \sin \mu$. Additionally, the generalisations of von Mises density for two components can be used to represent symmetric or asymmetric, uni-modal or bi-modal shapes depending on the choice of parameters. In the following chapter, we will examine Bayesian analysis of two important sub-models of generalisations of von Mises distributions for two components.

CHAPTER 3

A BAYESIAN ANALYSIS FOR ASYMMETRIC AND BIMODAL CIRCULAR DATA

Many circular data, such as the ones encountered in astrophysics, bio-informatics, geosciences, environmental sciences, meteorology, etc. have the properties of asymmetry and bi-modality simultaneously. In this chapter, we present a Bayesian analysis of two elegant asymmetric and possibly bimodal distributions, which can be considered as generalisations of von Mises distribution, which are difficult to analyse by the frequentist approach since their normalizing constants are not available in closed forms. In order to obtain samples from their posterior distributions, we use a *sampling importance re-sampling* (SIR) method. Because of the weights involved therein are discrete, we advocate the use of Shannon entropy. Additionally, we construct constrained joint prior distributions for the bi-modal cases. For model selection, we encounter an interesting situation where model averaging procedures become necessary, and use approximate Bayes factor and Bayesian information criteria. Our approaches are illustrated with real data examples from biology and ecology.

3.1 Introduction

In many biological and environmental real life research, there arise asymmetric and bi-modal circular data. For illustration, one of the most important research topic in marine biology is the spawning time of a particular fish. The spawning time is affected by tidal characteristics in fish biology. One of tidal characteristics is time of low tide. Figure 3.1 displays the rose diagram of time of low tide and shows that the

distribution is asymmetric.

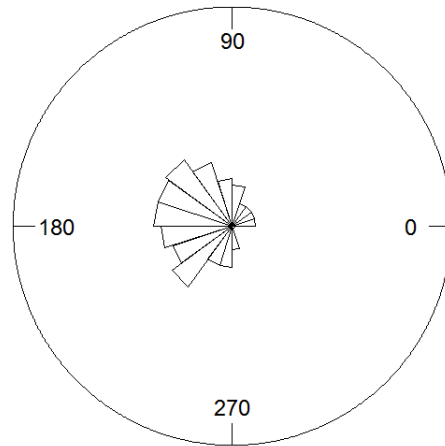


Figure 3.1: Rose diagram of time of low tide

In the context of Bayesian inference for von Mises distribution, the earliest attempt was presented by [Mardia and El-Atoum \(1976\)](#). The conjugate prior distribution of von Mises distribution was introduced by [Guttorp and Lockhart \(1987\)](#). A full Bayesian analysis of von Mises distribution was given by [Damien and Walker \(1999\)](#) who suggested Gibbs sampler, using auxiliary variables, to draw samples from posterior distribution. Sampling importance resampling (SIR) method was used by [Antonia and Peña \(2005\)](#) who generated the samples from posterior distribution for von Mises distribution using conjugate prior distributions. Noting the hurdles associated with the latent variable approach of Damien and Walker, SIR procedure was introduced by [SenGupta and Laha \(2008\)](#) to determine a change point with von Mises distribution.

To overcome the problems encountered for Bayesian inferences on circular distributions with unknown normalising constant, [Bhattacharya and SenGupta \(2009\)](#) proposed to combine importance sampling with MCMC (IS-MCMC) for some circular distributions.

The aim of this chapter is to model and develop Bayesian analyses procedures for recently emerging possibly asymmetric bi-modal circular distributions for such data as are often encountered in environmental and biological real-life data sets as will be exemplified later in this chapter. Two important models used in analysing such data

are based on generalisations of von Mises distribution. These distributions have uni-modality, asymmetry, as well as bi-modality properties. Main problem for analysing such distributions is their complex normalizing constants which are not available in closed forms. These constants lead to unknown normalising constants of their posterior distributions. To overcome this problem, first, we suggest that a sample from their posterior distributions can be generated using SIR, thereby eliminating the need to deal with the complex normalizing constants of the posterior distributions. Additionally, another feature of these distributions is that their uni-modality or bi-modality are dictated by certain constraints on the parameters. We overcome this problem by constructing new constrained joint prior distributions dependent parameters.

The rest of this chapter is organized as five sections. In Section 3.2, we explain briefly the definitions and some properties for two important sub-models of generalisations of von Mises distribution. Then, first we define conjugate prior distribution for them and construct joint prior distributions under the functional constraint on the parameters. In Section 3.3 , we explain how to apply SIR method and how to draw samples from their posterior distribution. In Section 3.4, we present model selection procedures. In Section 3.5, SIR methodology is illustrated with real life examples from biology and ecology. Section 3.6 includes some concluding remarks.

3.2 Some properties of two sub-models of generalisations of von Mises distribution and their joint conjugate and constrained priors

A wide class of absolutely continuous circular distributions that have an exponential family was introduced by [Maksimov \(1967\)](#) as follows

$$f(\theta) \propto \exp \left(\sum_{j=1}^k (a_j \cos j\theta + b_j \sin j\theta) \right) \quad (3.1)$$

which we we will rewrite as

$$f(\theta) \propto \exp \left(\sum_{j=1}^k \kappa_j \cos j(\theta - \mu_j) \right) \quad (3.2)$$

where $\theta \in [0, 2\pi)$, $a_j = \kappa_j \cos j\mu_j$ and $b_j = \kappa_j \sin j\mu_j$, $j = 1, \dots, k$. We consider Bayesian analysis of two important sub-models where $k = 2$ which leads to general-

isations of von Mises distribution, namely generalised von Mises (GvM) distribution which was introduced by [Cox \(1975\)](#) and a three parameter asymmetric generalised von Mises distribution (AGvM) which was introduced by [Kim and SenGupta \(2013\)](#). These sub-models can allow a great deal of flexibility in terms of asymmetry and bi-modality compared to von Mises distribution (vM).

We now review some definitions and define conjugate prior distributions and construct joint prior distributions of two important sub-models for their Bayesian inferences, respectively.

3.2.1 Generalised von Mises distribution

We note that a special case of (3.2), where $k = 2$, $a_j = \kappa_j \cos \mu_j$ and $b_j = \kappa_j \sin \mu_j$, $j = 1, 2$ commonly referred to as GvM, has received special attention, e.g. [Cox \(1975\)](#); [Yfantis and Borgman \(1982\)](#) etc.

GvM distribution has probability density function given by

$$f(\theta) = \frac{1}{2\pi c(\delta, \kappa_1, \kappa_2)} \exp(\kappa_1 \cos(\theta - \mu_1) + \kappa_2 \cos 2(\theta - \mu_2)) \quad (3.3)$$

where $\mu_1 \in [0, 2\pi)$, $\mu_2 \in [0, \pi)$, $\delta = \mu_1 - \mu_2$ and $\kappa_1, \kappa_2 > 0$. The normalizing constant is defined as

$$c(\delta, \kappa_1, \kappa_2) = \frac{1}{2\pi} \int_0^{2\pi} \exp(\kappa_1 \cos(\theta) + \kappa_2 \cos 2(\theta + \delta)) d\theta$$

An infinite series form of the normalizing constant can be written as

$$c(\delta, \kappa_1, \kappa_2) = I_0(\kappa_1)I_0(\kappa_2) + 2 \sum_{i=1}^{\infty} I_{2i}(\kappa_1)I_i(\kappa_2) \cos 2i\delta \quad (3.4)$$

where $I_p(\cdot)$ is the modified Bessel function of first kind and order p . Some important inferential features for GvM distribution is presented in e.g. [Yfantis and Borgman \(1982\)](#).

Note that GvM distribution is uni-modal if $\kappa_1 \geq 4\kappa_2$ otherwise bimodal. Under $H_0: \mu_1 = \mu_2$ hypothesis, by differentiation of the pdf we obtain, $-\kappa_1 \sin(\theta - \mu) -$

$$4\kappa_2 \sin(\theta - \mu) \cos(\theta - \mu) = 0. \text{ The solution of this equation is}$$

$$\mu = 0 \text{ or } \mu = \arccos\left(\frac{-\kappa_1}{4\kappa_2}\right), \quad \kappa_2 \neq 0 \quad (3.5)$$

from this solution, the constraint follows.

3.2.1.1 Conjugate prior distribution for GvM

Result 1 : A conjugate prior for GvM distribution is given by

$$\{c(\delta, \kappa_1, \kappa_2)\}^{-r} \exp(\kappa_1 R_{01} \cos(\mu_1 - \mu_{01}) + \kappa_2 R_{02} \cos 2(\mu_2 - \mu_{02}))$$

where r is an integer which shows the number of realizations from the joint prior distribution ($r = 1$). $\{c(\delta, \kappa_1, \kappa_2)\}^{-1}$ is the normalizing constant. $\mu_{01}, \mu_{02}, R_{01}$ and R_{02} can be considered the vector hyper-parameters of the prior.

Proof: see Appendix A.1

3.2.1.2 Constrained joint prior distribution for GvM

Joint prior distribution $p(\mu_1, \mu_2, \kappa_1, \kappa_2)$ of GvM is taken to be the product of the following three prior distributions as follows

$$p(\mu_1, \mu_2, \kappa_1, \kappa_2) = vM(\mu_1 | \mu, \kappa) \times \text{Unif}(\mu_2 | 0, \pi) \times f(\kappa_1, \kappa_2) \quad (3.6)$$

where $f(\kappa_1, \kappa_2)$ is a constrained joint prior distribution of dependent parameters of GvM as follows

Result 2: The boundary conditions of bi-modal case are considered as $0 < \kappa_1 < 4\kappa_2$, $0 < \kappa_2 < \infty$. Here, we consider truncated bivariate exponential conditionals distribution as constrained joint prior distribution $f(\kappa_1, \kappa_2)$ in explained Appendix A.3

3.2.2 Asymmetric generalised von Mises distribution

As a special case of (3.2), where $k = 2$, $\mu_j = \mu$, $a_j = \kappa_j \cos \mu$ and $b_j = \kappa_j \sin \mu$, $j = 1, 2$ referred to as AGvM introduced by [Kim and SenGupta \(2013\)](#). Here, assume

that we are given a sample of data $(\theta_1, \theta_2, \dots, \theta_n)$ from a AGvM distribution defined as follows

$$f(\theta) = \frac{1}{2\pi c(\frac{\pi}{4}, \kappa_1, \kappa_2)} \exp(\kappa_1 \cos(\theta - \mu) + \kappa_2 \sin 2(\theta - \mu)) \quad (3.7)$$

where $\mu \in [0, 2\pi)$ is location parameter, and the concentration parameter is $\kappa_1 > 0$ and $\kappa_2 \in [-1, 1]$ is a skewness parameter and the normalizing constant is

$$c(\frac{\pi}{4}, \kappa_1, \kappa_2) = \frac{1}{2\pi} \int_0^{2\pi} \exp(\kappa_1 \cos(\theta) + \kappa_2 \sin 2(\theta + \frac{\pi}{4})) d\theta.$$

The infinite series form of the normalizing constant can be obtained by selecting $\delta = \pi/4$ in equation (3.4). Note that this distribution is uni-modal if $\kappa_1 \geq |2\kappa_2|$ otherwise bimodal. The defining equation for modes and anti-modes for AGvM distribution is given by $\kappa_1 \sin \theta + 2\kappa_2 \cos 2\theta = 0$, or

$$\arcsin \left(\frac{-\kappa_1 \pm \sqrt{\kappa_1^2 + 32\kappa_2^2}}{8\kappa_2} \right), \quad \kappa_2 \neq 0 \quad (3.8)$$

This solution (3.8) (see, proof [Kim and SenGupta \(2013\)](#)) yields the constraint.

3.2.2.1 Conjugate Prior distribution

Result 3: A conjugate prior for AGvM distribution can be defined as shown below

$$\{c(\delta, \kappa_1, \kappa_2)\}^{-r} \exp(\kappa_1 R_{01} \cos(\mu - \mu_0) + \kappa_2 R_{02} \sin 2(\mu - \mu_0))$$

where r is the number of realisations from the joint prior distribution, $\{c(\delta, \kappa_1, \kappa_2)\}^{-1}$ is the normalizing constant. μ_0 , R_{01} and R_{02} are hyper-parameters.

Proof 2: see Appendix A.2

3.2.2.2 Constrained joint prior distributions for AGvM

The joint prior distribution of μ , κ_1 , κ_2 for AGvM can be taken to be the product of the two prior distributions. The prior distribution of μ is taken as von Mises distribution.

we consider constrained joint prior distribution of κ_1 and κ_2 as a bivariate distribution. The joint prior distribution $p^*(\mu, \kappa_1, \kappa_2)$ can then be written as shown below

$$p^*(\mu, \kappa_1, \kappa_2) = \text{vM}(\mu|\mu, \kappa) \times f(\kappa_1, \kappa_2) \quad (3.9)$$

Based on the dependent parameters which are the shape parameter κ_1 and the scale parameter κ_2 of AGvM distribution, the distribution may be uni-modal or bi-modal. As a result of this we have to define a constrained joint prior distributions for $f(\kappa_1, \kappa_2)$ as follows

Result 4: The boundary conditions of bi-modal case are considered as $0 < \kappa_1 < 2|\kappa_2| < 2$, and they can be written as $0 < \kappa'_1 = \frac{\kappa_1}{2|\kappa_2|} < 1$ and $0 < \kappa'_2 = \frac{\kappa_2+1}{2} < 1$. In order to obtain a sample from $f(\kappa'_1, \kappa'_2)$ distribution, we consider three constrained joint prior distributions, specifically, bivariate beta distribution, bivariate Dirichlet distribution and bivariate beta conditionals distribution and then retain only these simulated values which obey $\kappa_1 = 2\kappa'_1|\kappa_2|$ and $\kappa_2 = 2\kappa'_2 - 1$.

Some technical and computational details for these constrained joint prior distributions are explained in Appendix A.4.

3.3 Bayesian analysis for GvM and AGvM with SIR

Suppose that a sample of random variates is easily generated from continuous density $g(\boldsymbol{\varphi})$, but that what is really required is a sample from density as follows

$$h(\boldsymbol{\varphi}) = \frac{f(\boldsymbol{\varphi})}{\int f(\boldsymbol{\varphi})d\boldsymbol{\varphi}}$$

More generally, given positive function $f(\boldsymbol{\varphi})$, then how can we obtain a sample from given only a sample from $g(\boldsymbol{\varphi})$ and functional form $f(\boldsymbol{\varphi})$? One of the resulting sampling procedures is known as sampling importance re-sampling, (SIR) . (see, [Rubin \(1987\)](#) ; [Smith and Gelfand \(1992\)](#)). SIR methodology has two steps:

a) Draw a sample $\boldsymbol{\varphi}_i$, $i = 1, 2, \dots, M$ i.i.d from $g(\boldsymbol{\varphi})$ which includes the support of $f(\boldsymbol{\varphi})$

b) Compute sample weights $w(\boldsymbol{\varphi}_i) = f(\boldsymbol{\varphi}_i)/g(\boldsymbol{\varphi}_i), i = 1, \dots, M$, and calculate

$$q_i = w(\boldsymbol{\varphi}_i) / \sum_{j=1}^M w(\boldsymbol{\varphi}_j) \quad (3.10)$$

then draw $\boldsymbol{\varphi}^*$ from discrete distribution over the $\boldsymbol{\varphi}_1, \dots, \boldsymbol{\varphi}_M$ placing mass q_i on $\boldsymbol{\varphi}_i$. The new sample $\boldsymbol{\varphi}^*$ is approximately distributed according to $f(\boldsymbol{\varphi})$. This approximation will be improved by increasing M .

The sets of parameters $\boldsymbol{\varphi}$ are independent from each other. In other words, the two sub-models of interest namely GvM and AGvM may be unimodal or bi-modal. To obtain a sample from posterior distributions of GvM and AGvM as described Appendix A1 and A2, we suggest the following proposal densities:

GvM proposal densities $g_1(\boldsymbol{\varphi})$ and $g_2(\boldsymbol{\varphi})$ are as follows

$$\begin{aligned} g_1(\boldsymbol{\varphi}) &= \text{vM}(\mu_1 | \hat{\mu}, \hat{\kappa}) \times \text{Unif}(\mu_2 | 0, \pi) \times \text{Gamma}(\kappa_1 | \alpha_{gvm}, \beta_{gvm}) \\ &\quad \times \text{Gamma}(\kappa_2 | \alpha_{2gvm}, \beta_{2gvm}) \\ g_2(\boldsymbol{\varphi}) &= \text{Unif}(\mu_1 | 0, 2\pi) \times \text{Unif}(\mu_2 | 0, \pi) \times \\ &\quad \text{Gamma}(\kappa_1 | \alpha_{gvm}, \beta_{gvm}) \times \text{Gamma}(\kappa_2 | \alpha_{2gvm}, \beta_{2gvm}) \end{aligned}$$

AGvM proposal densities $g_1^*(\boldsymbol{\varphi})$ and $g_2^*(\boldsymbol{\varphi})$ are given below

$$\begin{aligned} g_1^*(\boldsymbol{\varphi}) &= \text{vM}(\mu | \hat{\mu}, \hat{\kappa}) \times \text{Gamma}(\kappa_1 | \alpha_{agvm}, \beta_{agvm}) \times \text{Unif}(\kappa_2 | -1, 1) \\ g_2^*(\boldsymbol{\varphi}) &= \text{Unif}(\mu | 0, 2\pi) \times \text{Gamma}(\kappa_1 | \alpha_{agvm}, \beta_{agvm}) \times \text{Unif}(\kappa_2 | -1, 1) \end{aligned}$$

where $\hat{\mu}$ and $\hat{\kappa}$ are maximum likelihood estimates for von Mises distribution. The parameters of the proposal gamma distributions can be selected around the center of maximum likelihood estimates of the two sub-models. A useful way of controlling accuracy of the proposal density $g(\boldsymbol{\varphi})$ is Shannon entropy H for a discrete random variable q_i , $H = -\sum_{i=1}^M q_i \log q_i$. The smaller values of H for the proposal densities would be preferable.

Another sampling procedure of SIR is *prior to posterior* from which the likelihood function plays an important role as re-sampling probability q_i . Since the posterior density can be written as $f(\boldsymbol{\varphi} | \theta_i) \propto L(\boldsymbol{\varphi}, \theta_i) \times p(\boldsymbol{\varphi})$, q_i is given by

$$q_i = L(\boldsymbol{\varphi}_i, \theta_i) / \sum_{j=1}^M L(\boldsymbol{\varphi}_j, \theta_i) \quad (3.11)$$

In other words, a prior (joint) distribution for unknown parameters may be determined $\boldsymbol{\varphi}$. Samples are then drawn from this prior distribution and likelihood calculated for each sample. The prior is re-sampled using likelihood as weights. This procedure is used to obtain a sample the posterior distribution of the parameters using constrained joint prior distributions of GvM and AGvM distributions in analysing bi-modal directional data. Here, these constrained joint distributions have complex form, but are easy to simulate.

3.4 Model selection

In order to compare Bayesian modelling with AGvM, M_{agvm} and Bayesian modelling with GvM, M_{gvm} , we consider Bayes factor formula as follows

$$\frac{p(\boldsymbol{\theta}|M_{agvm})}{p(\boldsymbol{\theta}|M_{gvm})} = \frac{\int_{\boldsymbol{\varphi}_1} f_1(\boldsymbol{\theta}|\boldsymbol{\varphi}_1)\pi_1(\boldsymbol{\varphi}_1)d\boldsymbol{\varphi}_1}{\int_{\boldsymbol{\varphi}_2} f_2(\boldsymbol{\theta}|\boldsymbol{\varphi}_2)\pi_2(\boldsymbol{\varphi}_2)d\boldsymbol{\varphi}_2} = B_{12} \quad (3.12)$$

where $\boldsymbol{\theta}$ denote observed data points, and $\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2$ denote the parameters of each model. Bayes factor which is shown by B_{12} is used to compare M_{agvm}, M_{gvm} models. Each of the integral is known as a marginal likelihood, and the calculation of each marginal likelihood is too difficult so we use approximate Bayes factor

$$\widehat{B}_{12} = \frac{\frac{1}{n_1} \sum_{i=1}^{n_1} f_1(\boldsymbol{\theta}|\boldsymbol{\varphi}_1^{(i)})\pi_1(\boldsymbol{\varphi}_1^{(i)})/g_1(\boldsymbol{\varphi}_1^{(i)})}{\frac{1}{n_2} \sum_{i=1}^{n_2} f_2(\boldsymbol{\theta}|\boldsymbol{\varphi}_2^{(i)})\pi_2(\boldsymbol{\varphi}_2^{(i)})/g_2(\boldsymbol{\varphi}_2^{(i)})}. \quad (3.13)$$

where $g_1(\boldsymbol{\varphi}_1), g_2(\boldsymbol{\varphi}_2)$ are importance functions and $\pi_1(\boldsymbol{\varphi}_1), \pi_2(\boldsymbol{\varphi}_2)$ are conjugate prior distributions for each model.

For *prior to posterior* implementation, we also present more traditional methods for selection of the joint prior distribution or the constrained joint prior distribution of dependent parameters of AGvM distribution, namely, the Akaike information criterion (AIC), Bayesian information criterion (BIC) and Bayesian model averaging (BMA). We compute AIC and BIC as $AIC = -2\log(\boldsymbol{\theta}|\hat{\boldsymbol{\varphi}}) + 2 \times (\#number\ of\ parameters)$ and $BIC = -2\log(\boldsymbol{\theta}|\hat{\boldsymbol{\varphi}}) + (\#number\ of\ parameters) \times \log(n)$. For BMA, the posterior model probabilities are defined as $p(M_k|\boldsymbol{\theta}) = \frac{p(\boldsymbol{\theta}|M_k)p(M_k)}{\sum_{l=1}^K p(\boldsymbol{\theta}|M_l)p(M_l)}$. To compute model probabilities, we use the simple BIC approximation introduced by [Raftery](#)

(1995) is given by $P(M_k|\theta) \approx \exp(-\text{BIC}_k/2) / \sum_{l=1}^K \exp(-\text{BIC}_l/2)$ where K is number of models.

3.5 Real data examples

Our methods proposed in this section can be used for the analysis of (i) symmetric and uni-modal, (ii) symmetric and bimodal, (iii) asymmetric and uni-modal, (iv) asymmetric and bimodal data. Here we illustrate the use of our Bayesian approach with SIR through the analysis of asymmetric-uni-modal fish data, asymmetric-bimodal turtle data. At first, we consider time of low tide of a particular fish. In order to draw samples from posterior distribution, we applied SIR described in the equation (3.10). As second example, we consider a study of movement turtle data. Here, we applied *prior to posterior* implementation in SIR as given by (3.11). For each case we simulate random samples $M=1,000,000$ from the corresponding proposal densities and we re-sample 10,000 samples to produce a sample from the desired distribution. The parameters of the proposal densities are selected as around the center of maximum likelihood estimates of two sub- models.

3.5.1 Spawning time of fish

To illustrate the use of our Bayesian method, we use the data that were collected on the spawning time of a particular fish by Robert R. Warner at the University of California, Santa-Barbara. These data are analysed by [Kim and SenGupta \(2013\)](#) using maximum likelihood approach. They conclude that AGvM or GvM fits data set better than vM for asymmetry data. We also re-analyse the data set in Bayesian perspective. For this data, $\hat{\mu} = 2.90$, $\hat{\kappa} = 1.76$, $R_{n_1} = 56.37$ and $R_{n_2} = 15.55$. We take hyper-parameters $r = 0$, $R_{01} = 0$, $R_{02} = 0$ as vague prior and select $\alpha_{agvm}=4$, $\beta_{agvm}=2$, $\alpha_{gvm} = 4$, $\beta_{gvm} = 2$, $\alpha_{2gvm} = 1$, $\beta_{2gvm} = 2$ as around center of maximum likelihood estimates of two sub-models.

In order to obtain a sample from posterior distribution both AGvM and GvM, we use the following proposal densities, respectively.

AGvM proposal density $g_1^*(\mu, \kappa_1, \kappa_2)$ is

$$vM(\mu|\hat{\mu}, \hat{\kappa}) \times \text{Gamma}(\kappa_1|\alpha_{agvm}, \beta_{agvm}) \times \text{Unif}(\kappa_2|-1, 1)$$

GvM proposal density $g_1(\mu_1, \mu_2, \kappa_1, \kappa_2)$ is

$$vM(\mu_1|\hat{\mu}, \hat{\kappa}) \times \text{Unif}(\mu_2|0, \pi) \times \text{Gamma}(\kappa_1|\alpha_{gvm}, \beta_{gvm}) \times \text{Gamma}(\kappa_2|\alpha_{2gvm}, \beta_{2gvm}).$$

In Table 3.1, the posterior means and the 95% credible intervals for AGvM and GvM are summarized for five independent SIR algorithms. We also used Shannon entropy H measure to determine the accuracy of the proposal densities. According to H measure results as shown in Table 3.1, the proposal densities are suitable for AGvM and GvM distribution. Fig 3.2 shows the maximum likelihood results with the red vertical line close to the posterior means with blue line for AGvM distribution. Fig 3.3 shows the posterior means of GvM distribution with a vertical blue line.

Table 3.1: Five SIR runs and posterior mean estimates of both AGvM and GvM

AGvM	$\hat{\mu}$ 95%CI	$\hat{\kappa}_1$ 95%CI	$\hat{\kappa}_2$ 95%CI			H	
	2.90 (2.67, 3.12)	1.77 (1.16,2.27)	-0.003 (-0.34,0.34)			9.81	
	2.90 (2.67, 3.12)	1.77 (1.32,2.26)	-0.001 (-0.33,0.35)			9.80	
	2.90 (2.67, 3.13)	1.76 (1.31,2.28)	0.002 (-0.35,0.33)			9.82	
	2.90 (2.68, 3.12)	1.76 (1.33,2.27)	0.002 (-0.34,0.34)			9.82	
	2.90 (2.68, 3.12)	1.77 (1.32,2.29)	0.003 (-0.34,0.34)			9.81	
GvM	$\hat{\mu}_1$ 95%CI	$\hat{\mu}_2$ 95%CI	$\hat{\kappa}_1$ 95%CI	$\hat{\kappa}_2$ 95%CI			H
	2.75 (2.57, 2.94)	0.97 (0.65,1.24)	2.40 (1.69,3.27)	0.71 (0.24,1.18)			8.21
	2.76 (2.58, 2.94)	0.97 (0.66,1.25)	2.42 (1.68,3.29)	0.71 (0.24,1.19)			8.21
	2.75 (2.58, 2.94)	0.96 (0.66,1.26)	2.41 (1.69,3.27)	0.70 (0.24,1.19)			8.19
	2.75 (2.58, 2.94)	0.96 (0.67,1.25)	2.42 (1.70,3.21)	0.72 (0.24,1.19)			8.19
	2.75 (2.58, 2.94)	0.96 (0.66,1.25)	2.41 (1.71,3.23)	0.71 (0.25,1.21)			8.21

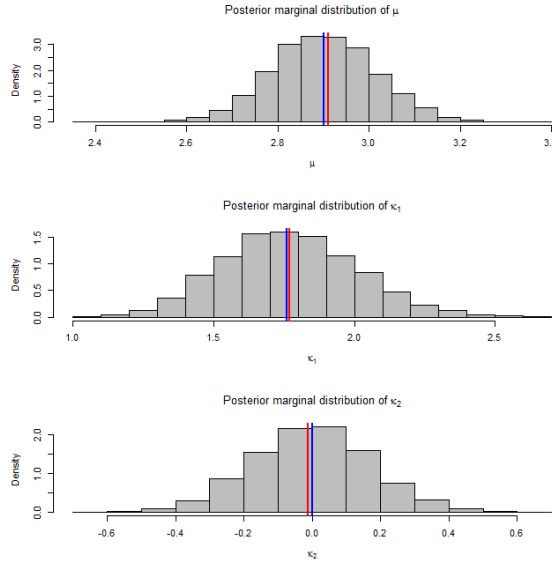


Figure 3.2: Bayesian estimation of AGvM distribution: the vertical red line shows maximum likelihood estimates, the blue line shows Bayesian estimates of the posterior means

In order to compare models, AGvM and GvM, we compute Bayes factor as $\widehat{B}_{12} = 3.46$. Then, following [Jeffreys \(1961\)](#)¹, we propose the asymmetric generalised von Mises distribution for the spawning time data.

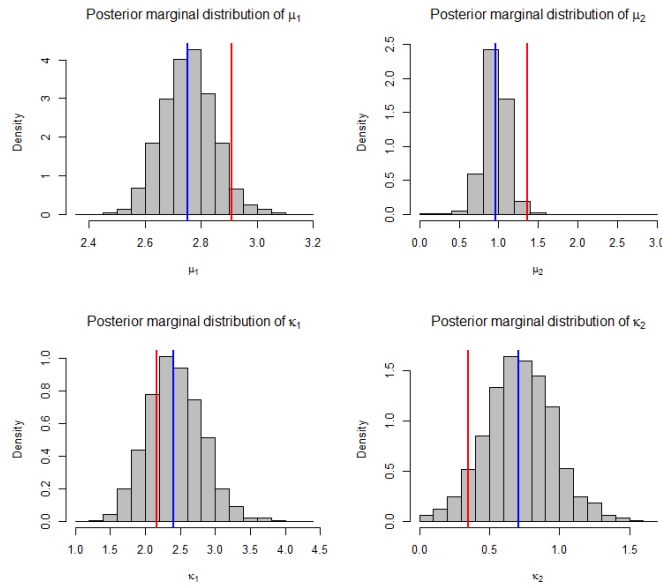


Figure 3.3: Bayesian estimation of GvM distribution: the vertical red line shows maximum likelihood estimates, the blue line shows Bayesian estimates of the posterior means

¹ Jeffreys (1961) suggests that there is substantial evidence about model 1 if $3 < BF < 10$

3.5.2 Movement of turtle

For the illustration of the bimodal case of AGvM, we consider turtle data of Gould cited by Stephens, (1969). The data consists of orientations of 76 turtles after laying eggs. Previously, turtle data was analysed by using two mixture von Mises distribution. (see, e.g Stephens, (1969); Mardia (1975)) We re-analyse this data and use three constrained joint prior distributions for dependent parameters of AGvM, namely bivariate beta Olkin $f_{bo}(\cdot)$, bivariate Dirichlet $f_{bd}(\cdot)$ and bivariate beta conditionals $f_{bc}(\cdot)$. We take the hyper parameters as shape parameters, $a = 750$, $b = 0.01$, $c = 1000$ both bivariate beta Olkin and bivariate beta Dirichlet distribution. The hyper-parameters of bivariate beta conditionals distribution are taken as $m_{01} = 100$, $m_{11} = 0.03$, $m_{02} = 10000$, $m_{20} = 1000$, $m_{10} = 1000$. Here, again, the hyper-parameters of three constrained joint prior distributions are centred at maximum likelihood estimates of AGvM distribution.

The joint prior distributions of AGvM are given by

$$p_1^*(\mu, \kappa_1, \kappa_2) = vM(\mu | \hat{\mu}, \hat{\kappa}) \times f_{bo}(\kappa_1, \kappa_2)$$

$$p_2^*(\mu, \kappa_1, \kappa_2) = vM(\mu | \hat{\mu}, \hat{\kappa}) \times f_{bd}(\kappa_1, \kappa_2)$$

$$p_3^*(\mu, \kappa_1, \kappa_2) = vM(\mu | \hat{\mu}, \hat{\kappa}) \times f_{bc}(\kappa_1, \kappa_2)$$

where $\hat{\mu} = 1.12$ and $\hat{\kappa} = 1.14$ are maximum likelihood estimates of von Mises distribution for turtle data. In this example, we compare three different joint prior distributions of AGvM model by the model selection criteria. In Table 3.2, Akaike information criteria, (AIC) and Bayesian information criteria (BIC) for AGvM model are summarized for three different joint prior distributions. The smallest criteria denotes the best model. Based on the results of model selection criteria, it is useful to use $p_1^*(\mu, \kappa_1, \kappa_2)$ as the joint prior for AGvM. This is an expected result, as the number of hyper-parameters is increased, SIR technique may not be suitable. For instance, the constrained joint prior distribution which is bivariate beta conditionals distribution has six hyper-parameters and $p_3^*(\mu, \kappa_1, \kappa_2)$ has a slightly higher AIC and BIC compared to the others. Others may consider flexible bivariate beta distribution which is introduced by Arnold and Tony (2011) as the constrained joint prior distribution for dependent parameters, but, we eliminate this because of the number of parameters and it not being available in closed form. However, this distribution may be consid-

ered under the positive and the negative correlations. We propose three constrained joint prior distributions for Bayesian analysis of the AGvM distribution for dependent parameters. Among these, the bivariate beta distribution can be proposed as the constrained joint prior distribution for AGvM distribution.

Table 3.2: Prior selection for turtle data

AGvM	$p_1^*(\mu, \kappa_1, \kappa_2)$	$p_2^*(\mu, \kappa_1, \kappa_2)$	$p_3^*(\mu, \kappa_1, \kappa_2)$
AIC	224.76	224.82	225.07
BIC	231.75	231.81	232.06

From these results, we get these estimates as $\hat{\mu} = 1.74$, $\hat{\kappa}_1 = 0.87$ and $\hat{\kappa}_2 = -1.00$ using joint prior $p_1^*(\mu, \kappa_1, \kappa_2)$. To compare two mixture von Mises model and AGvM model, we use BIC and BMA. For two mixtures von Mises distribution, we take parameter estimates of [Mardia' \(1975\)](#).

Table 3.3: Comparison of the models for turtle data

Model	BIC	$P(M_k \theta)$
Two mixture von Mises	232.48	0.41
AGvM	231.75	0.59

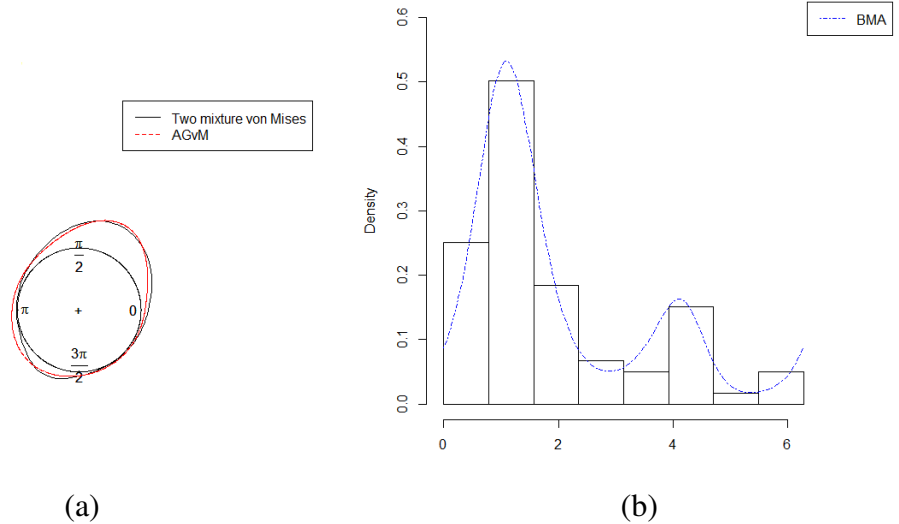


Figure 3.4: (a) Comparison of model fits for turtle data (b) Posterior distribution for Bayesian model averaged for turtle data

From Table 3.3 , AGvM model is more utility than two mixture von Mises distribution in terms of comparison criteria. Fig. 3.4.a shows the fits of two mixture von Mises

and AGvM models. Furthermore, the posterior model probabilities were found to be 0.41 for two mixture von Mises model, 0.59 for AGvM model. With these values, we plot BMA posterior distribution for turtle data displayed in Fig. 3.4.b.

3.6 Concluding remarks

In this study, we have provided a fully Bayesian analysis of two sub-models of generalisations of von Mises distribution in analysing asymmetric and bi-modal circular data. SIR method presents both a good alternative and a simple form in Bayesian analysis of circular distributions which have complex normalising constants. To handle their complex normalising constants, we applied SIR method and consider two approaches, firstly, we assume that the parameters are independent from each other, then we use SIR algorithm to obtain a sample from their posterior distributions. Secondly, if the parameters are functionally dependent, then we propose joint prior distributions using likelihood principle in SIR. The main property of the second approach is that their complex normalizing constants can be ignored. Moreover, we also propose constrained joint prior distributions under re-parametrization for depended parameters of AGvM and GvM. Additionally, for the determination of hyper-parameters, maximum likelihood estimation provides global maximum for our case and hence SIR method is not further needed here for that purpose.

We would like to emphasize that these sub-models cover uni-modality, asymmetry as well as bi-modality.

CHAPTER 4

BAYESIAN SEMI-PARAMETRIC MODELS FOR MULTI-MODAL CIRCULAR DATA

In many environmental and ecological data analysis such as wind directions, dihedral angles and orientation of a specific bird, the empirical distribution displays a multi-modal structure. One way to deal with the analysis of such data sets is to consider k -mixture distribution where k is the number of mixing components which is often unknown. In this chapter, we aim at addressing this problem and adopting Dirichlet process (DP) mixture model with mixtures of von Mises (vM) and mixtures of wrapped Cauchy (wC) distributions. In fact, the main problem about model uncertainty is to choose an appropriate model via a suitable probability distribution. Recently, there has been an increasing interest in the use of Bayesian non-parametric models based on probability distributions over spaces of distributions. These models are not commonly used in analysing circular data due to the difficulty of obtaining a sample from the posterior distribution of the parameters of the component distributions. Our proposed models overcome this difficulty and we present a simulation study and real data examples to illustrate the usefulness and flexibility of them.

4.1 Introduction

In many environmental and ecological researches, data are directional such as wind directions, the orientations of turtles and spawning times of a particular fish. Exploratory data analyses reveal that some of such data are multi-modal, for example turtle data as seen in Fig 4.1 . For the analysis of multi-modal circular data, one may

consider using mixtures of von Mises (vM) distribution as follows

$$f(\theta_i; p_1, \dots, p_C, \mu_1, \dots, \mu_C, \kappa_1, \dots, \kappa_C) = \sum_{k=1}^C p_k f(\theta_i; \mu_k, \kappa_k), \text{ for } i = 1, \dots, n \quad (4.1)$$

where p_k 's are unknown mixing probabilities, $\sum_{k=1}^C p_k = 1$, μ_k and κ_k are circular mean and concentration parameters of the k th mixing distribution respectively, $f(\theta_i; \mu_k, \kappa_k), k = 1, \dots, C$ are vM probability density function of participating in the mixture.

The number of modes is generally unknown and the challenge with which the analyst is faced is determining the number of modes. To overcome this problem, we adopt a Bayesian approach based on Dirichlet process (DP) mixture model.

DP mixture model approaches are commonly used in analysing linear data while these approaches in analysing circular data are limited. This is due to the fact that it is hard to deal with the complicated normalizing constant which is not available in their closed forms and some burdensome problems in Markov chain Monte Carlo (MCMC) methods.

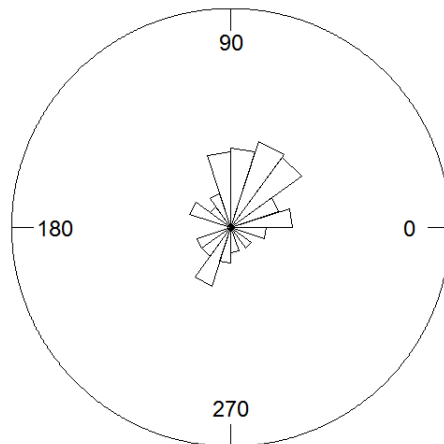


Figure 4.1: Rose diagram of turtle data

In the context of DP mixture for circular data modelling, [Ghosh et al \(2003\)](#) considered DP mixture for vM distribution for the problems of prediction and test of hypothesis, but they worked under the assumption of having same concentration pa-

parameter κ over the mixing distributions. [Bhattacharya and SenGupta \(2009\)](#) considered DP mixture vM model for determining an unknown number of parameters. Their approach works well under the assumption of a single concentration parameter κ for all the distributions participating to the mixture, but they do not pursue their approach for multiple concentration parameters κ_k due to burdensome problems in their MCMC applications. Recently, [Antonio et al. \(2014\)](#) have provided DP mixture circular models for projected and wrapped normal distributions due to the complex normalizing constant of vM probability density function.

In the context of kernel density estimation for circular data, [Hall et al \(1987\)](#) used cross-validation and minimization based on mean squared error loss and Kullback-Leibler loss for selecting the bandwidth. [Taylor \(2008\)](#) assumed the underlying population to be von Mises and used von Mises kernel. [Oliviera et al. \(2012\)](#) proposed a new selector based on finite mixture vM densities.

Focus of the current chapter is to detect the number of modes of both vM Mises and wrapped Cauchy (wC) distributions while relaxing the assumption for equal concentration parameters. Benefit of DP mixture model approach in the analysis of multimodal data is twofold: 1. offers an efficient method to analyse the dataset in the presence of unknown number of mixture components, 2. gives an estimation about the number of mixture components. Based on the latter result, one may re-analyse the dataset using a more appropriate distribution. For instance, if DP mixture method estimates the number of mixing components to be 2, then one may consider fitting a bimodal distribution.

The rest of this chapter is organized as five sections. Section [4.2](#) gives a brief summary to DP mixture model. Section [4.3](#) gives our proposed DP mixture approach for common mixture circular distributions with unequal concentration parameters. In Section [4.4](#), we evaluate the accuracy of our approach for circular data analysis using simulated data sets and apply it to turtle and ant data. Section [4.5](#) gives some discussion and in Appendix [B.3](#), we set forth to apply this process with our OpenBUGS codes.

4.2 DP mixture models

Modelling a distribution as a mixture of simpler distributions is useful both as a non-parametric density estimation method and as a way of identifying latent classes that can explain the dependencies observed between variables. Mixture models can easily be handled in a Bayesian framework by employing a prior distribution for mixing weights. In recent years, there has been a remarkable increase in the use of DP mixture model depending on the development of Markov Chain Monte Carlo methods for obtaining a sample from the posterior distribution of the parameters of the component distributions.

The earliest attempt with respect to Bayesian non-parametric studies was by [Ferguson \(1973\)](#). [Blackwell and MacQuenn \(1973\)](#) showed the marginal distribution of latent class variables that had Polya-Urn representation. This result leads to work on computational procedures for Bayesian non-parametric models.

DP is defined by [Ferguson \(1973\)](#) as follows

$$G \sim DP(G_0, \alpha)$$

where G_0 represents a base distribution, α is the concentration parameter which shows whether G_0 is in the close realisation of G . A distribution consist of all probabilities for partition of sample space Ω , that is, for all partitions denoted by (A_1, \dots, A_k) ,

$$(G(A_1), \dots, G(A_k)) \sim Dir(\alpha G_0(A_1), \dots, \alpha G_0(A_k)) \quad (4.2)$$

where *Dir* denotes Dirichlet distribution.

In DP mixture model, DP is used as a nonparametric prior by [Antoniak \(1974\)](#). A DP mixture model is shown as in the following hierarchical Bayesian specification

$$\begin{aligned} \theta_i | \varphi_i &\sim F(\varphi_i) \\ \varphi_i | G &\sim G \\ G &\sim DP(G_0, \alpha) \end{aligned} \quad (4.3)$$

where $\theta_1, \dots, \theta_n$ is a set of independent circular observations, and we model circular data from which θ_i , $i = 1, \dots, n$ are drawn from infinite mixture circular distribution $F(\boldsymbol{\varphi})$. The prior for infinite mixture distribution is DP.

Another form of writing in terms of finite mixture model with C components is shown as in the following (Neal (2000)):

$$\begin{aligned}
\theta_i | K_i, \boldsymbol{\varphi} &\sim F(\varphi_{K_i}) \\
K_i | \mathbf{p} &\sim \text{Discrete}(p_1, \dots, p_C) \\
\varphi_k &\sim G_0, \quad k = 1, \dots, C \\
\mathbf{p} &\sim \text{Dir}(\alpha/C, \dots, \alpha/C)
\end{aligned} \tag{4.4}$$

where K_i denotes the latent class to which the observation θ_i belongs. The parameters φ_k characterize the distribution of observations for each class k . The weights for classes, $\mathbf{p} = (p_1, \dots, p_C)$ are given by Dirichlet prior, with parameter α/C . Taking the limit as C goes to infinity of finite mixture models with C components can be obtained an equivalent model.

4.2.1 Stick breaking construction

Sethuraman (1994) defines the representation of DP in terms of stick breaking construction. Accordingly, for $q_i \sim \text{Beta}(1, \alpha)$, $i = 1, 2, \dots$,

$$\begin{aligned}
p_1 &= q_1 \\
p_2 &= (1 - q_1)q_2 \\
p_3 &= (1 - q_1)(1 - q_2)q_3 \\
&\dots
\end{aligned}$$

This recursive relation can be shown as $p_i = q_i \prod_{j=1}^{i-1} (1 - q_j)$, the stick breaking representation of G is given by

$$G = \sum_{i=1}^{\infty} p_i I_{\varphi_i}, \quad \varphi_i \sim G_0$$

where I_{φ_i} is an indicator function at φ_i . This representation of DP shows that G is discrete distribution with probability one.

Ishwaran and Zarepour (2000) and Ishwaran and James (2001) propose that this approach may be truncated at C components, that is, $\sum_{i=1}^C p_i = 1$ and truncated DP

(TDP) can be shown as follows:

$$\begin{aligned}\varphi_i|G &\sim G \\ G &\sim TDP(C, G_0, \alpha)\end{aligned}$$

By letting C to be the maximum number of components, the infinite series above can be approximated by its finite counterpart as shown below

$$\sum_{i=1}^{\infty} p_i I_{\varphi_i} \approx \sum_{i=1}^C p_i I_{\varphi_i} \quad (4.5)$$

Equation 4.5 shows that G converges almost surely to a DP with αG_0 . Finally, it can be written as $G \rightarrow DP(\alpha, G_0)$.

4.3 DP mixture circular models with stick breaking construction

In this section, we introduce two DP mixture circular models in analysing multi-modal circular data. The first modelling is DP mixture vM model. The second model, DP mixture wC model, considers for heavy peaks around on unit circle.

4.3.1 DP mixture von Mises model

Here, we consider a DP mixture vM model as follows

$$\begin{aligned}\theta_i|K_i, \boldsymbol{\varphi} &\sim \text{vM}(\mu_{K_i}, \kappa_{K_i}), \quad i = 1, \dots, n \\ K_i|\boldsymbol{p} &\sim \text{Discrete}(p_1, \dots, p_C) \\ \boldsymbol{\varphi}_k &= (\mu_k, \kappa_k) \sim G_0, \quad k = 1, \dots, C \\ \alpha &\sim \text{Gamma}(v_1, v_2)\end{aligned} \quad (4.6)$$

where G_0 is a bivariate distribution. We will consider $\text{vM}(\mu_0, \kappa_0) \otimes \text{Gamma}(a_0, b_0)$ for μ_k and κ_k . $I_a(b)$ denotes unit mass at $a = b$, $\boldsymbol{\varphi} = (\varphi_{K_1}, \dots, \varphi_{K_n})$, $\mathbf{K} = (K_1, \dots, K_n) \in (1, \dots, C)^n$ under G and the weights $\boldsymbol{p} = (p_1, \dots, p_C)$ are determined by stick breaking algorithm. For $q_k \sim \text{Beta}(1, \alpha)$, and the foregoing notation is shown as $p_k = (1 - q_{k-1})q_k p_{k-1} / q_{k-1}$. Note that this algorithm has been already truncated so that the stick is only broken C times, and G is defined as shown below:

$$G = \sum_{k=1}^C p_k I_{\varphi_k}, \quad \varphi_k \sim G_0,$$

A common choices for the parameter α of Dirichlet process is Gamma(2,2) distribution in which both shape and scale parameters are equal to 2. This prior is a good choice for high and low values of α . The high values of α denote the number of mixture components too high, while the low values of α denote the number of mixture components too low.

For κ_k component, instead of gamma prior, we can use a uniform prior as shown below

$$\kappa_k | a_0, b_0 \sim \text{Uniform}(a_0, b_0)$$

Hyper-parameters a_0, b_0 can be selected based on the value of Bessel functions as described in Chapter 2. In particular, for low concentration parameter κ , we may choose as $a_0 = 0, b_0 = 3.75$, respectively.

4.3.2 DP mixture wrapped Cauchy model

A DP mixture wC model is defined as follows

$$\begin{aligned} \theta_i | K_i, \boldsymbol{\varphi} &\sim \text{wC}(\mu_{K_i}, \rho_{K_i}), \quad i = 1, \dots, n \\ K_i | \boldsymbol{p} &\sim \text{Discrete}(p_1, \dots, p_C) \\ \boldsymbol{\varphi}_k = (\mu_k, \rho_k) &\sim G_0, \quad k = 1, \dots, C \\ \alpha &\sim \text{Gamma}(v_1, v_2) \end{aligned} \tag{4.7}$$

In this case, we select von Mises-Beta baseline prior G_0 , that is, the components of $\boldsymbol{\varphi}_k$ are independently distributed as $vM(\mu_0, \kappa_0) \otimes \text{Beta}(a_0, b_0)$ for μ_k and ρ_k simultaneously. α has gamma prior distribution with shape parameter v_1 and scale parameter v_2 .

4.3.3 Inference via Gibbs sampler

To obtain direct inference for G, we use blocked Gibbs sampling approach in described [Ishawran and James \(2002\)](#) for our model specifications . The posterior dis-

tribution of $G|\boldsymbol{\theta}$ can be written as shown below

$$\begin{aligned} [\boldsymbol{\varphi}, \mathbf{p}, \mathbf{K}, \boldsymbol{\alpha} | \boldsymbol{\theta}] &\propto \prod_{i=1}^n \text{vM}(\theta_i | \mu_{K_i}, \kappa_{K_i}) \times \text{vM}(\mu_{K_i} | \mu_0, \kappa_0) \times \text{Gamma}(\kappa_{K_i} | a_0, b_0) \\ &\times \prod_{i=1}^n \text{Discrete}(K_i | p_C) \times \pi(p_C = p_C(q_{C-1})) \times \text{Gamma}(\boldsymbol{\alpha} | \nu_1, \nu_2) \quad (4.8) \end{aligned}$$

π denotes the prior distribution of \mathbf{p} , which is obtained by stick breaking algorithm with Beta priors. The joint posterior distribution is not available in closed form. However, to draw a random sample, we can use Gibbs sampling using the full conditional distributions as follows

$$\begin{aligned} [\boldsymbol{\varphi} | \mathbf{K}, \mathbf{p}, \boldsymbol{\alpha}, \boldsymbol{\theta}] &= [\boldsymbol{\varphi} | \mathbf{K}, \boldsymbol{\theta}] \propto \prod_{i=1}^n \text{vM}(\theta_i | \mu_{K_i}, \kappa_{K_i}) \times \text{vM}(\mu_{K_i} | \mu_0, \kappa_0) \\ &\times \text{Gamma}(\kappa_{K_i} | a_0, b_0) \\ [\mathbf{K} | \mathbf{p}, \boldsymbol{\varphi}, \boldsymbol{\alpha}, \boldsymbol{\theta}] &= [\mathbf{K} | \mathbf{p}, \boldsymbol{\varphi}, \boldsymbol{\theta}] \propto \prod_{i=1}^n \text{vM}(\theta_i | \mu_{K_i}, \kappa_{K_i}) \times \text{Discrete}(K_i | p_C) \\ [\mathbf{p} | \mathbf{K}, \boldsymbol{\varphi}, \boldsymbol{\alpha}, \boldsymbol{\theta}] &= [\mathbf{p} | \mathbf{K}, \boldsymbol{\alpha}] \propto \prod_{i=1}^n \text{Discrete}(K_i | p_C) \times \pi(p_C = p_C(q_{C-1})) \\ &\times \text{Gamma}(\boldsymbol{\alpha} | \nu_1, \nu_2) \\ [\boldsymbol{\alpha} | \mathbf{p}, \boldsymbol{\varphi}, \mathbf{K}, \boldsymbol{\theta}] &= [\boldsymbol{\alpha} | \mathbf{p}] \propto \pi(p_C = p_C(q_{C-1})) \times \text{Gamma}(\boldsymbol{\alpha} | \nu_1, \nu_2) \quad (4.9) \end{aligned}$$

This procedure generates the samples from posterior distribution $[\boldsymbol{\varphi}, \mathbf{p}, \mathbf{K}, \boldsymbol{\alpha} | \boldsymbol{\theta}]$ and for each cycle of Gibbs sampler, we can oversee $(\boldsymbol{\varphi}^*, \mathbf{p}^*)$ which are drawn the samples of $(\boldsymbol{\varphi}, \mathbf{p})$. These samples generate a random probability measure as shown below:

$$G^*(\cdot) = \sum_{k=1}^C p_k^* I_{\varphi_k^*}(\cdot)$$

where G^* can be used to directly estimate posterior distribution $G|\boldsymbol{\theta}$. We may start from initial values $(\boldsymbol{\varphi}^{(0)}, \mathbf{p}^{(0)}, \mathbf{K}^{(0)}, \boldsymbol{\alpha}^{(0)})$, and we may moderately simulate $(\boldsymbol{\varphi}^{(t)}, \mathbf{p}^{(t)}, \mathbf{K}^{(t)}, \boldsymbol{\alpha}^{(t)})$ from the conditional distributions in the equation 4.9. In order to choose the initial values, we may run a trial MCMC algorithm, and then, we can use the final iteration of MCMC algorithm for inference. In addition, the full conditional distributions of DP mixture wC model are defined a very similar way to DP mixture vM model.

Finally, the derivations of the full conditional distributions for DP mixture vM and wC models are given in Appendix (B.1) and (B.2) respectively. Our OpenBUGS codes

for DP circular mixture models are also given in Appendix (B.3). In addition, the implementation of mixture Dirichlet process for linear data in WinBUGS or OpenBUGS can be found in Congdon, (2001) .

4.4 Applications

In this section, we use four simulated data examples, a Monte Carlo study and two real data examples to illustrate our proposed models. For circular data generation, we use circular package in R. On the other hand, all inferences is coded in OpenBUGS with same burn in (5000 iterations). In all cases, we save a posterior Monte Carlo sample of size 5000 iterations. Moreover, assessment of convergence, we monitored the dynamic traces of Gibbs sampling and used the value of the Brooks-Gelman-Rubin ratio. In addition, for comparison, we use circular kernel density estimation in R. For the bandwidth selection of circular kernel density approach can be used bw.nrd and bw.cv.ml functions in circular package of in R.

4.4.1 Simulated data examples

In order to assess the accuracy of the proposed mixture DP approach, we designed the following simulation study. The following four distinct models are considered for circular data generation. Then for each simulated data set, the proposed method is employed to estimate the model parameters. Resulting estimates are compared against the true parameters to evaluate the performance of the method.

$$\theta_i \sim 0.1vM(1, 1) + 0.2vM(2, 1) + 0.7vM(3, 2) \quad i = 1, \dots, n \quad (4.10)$$

where $n = 1000$ observations which have lower concentration were simulated from the mixture of three vM distributions.

$$\theta_i \sim 0.1vM(1, 4) + 0.2vM(3, 5) + 0.7vM(2, 5) \quad i = 1, \dots, n \quad (4.11)$$

where $n = 1000$ observations which have larger concentration were simulated from the mixture of three vM distributions.

$$\theta_i \sim 0.2wC(1, 0.2) + 0.3wC(0.5, 0.3) + 0.5wC(3, 0.4) \quad i = 1, \dots, n \quad (4.12)$$

where $n = 1000$ observations which have lower concentration were simulated from the mixture of three wC distributions. These true models encompass various different mixture scenarios as seen in Fig. 4.2 and Fig. 4.3

$$\theta_i \sim 0.2\text{wC}(1, 0.9) + 0.3\text{wC}(0.5, 0.8) + 0.5\text{wC}(3, 0.7) \quad i = 1, \dots, n \quad (4.13)$$

where $n = 1000$ observations which have larger concentration were simulated from the mixture of three wC distributions.

Table 4.1: Posterior means of the mixing probabilities and parameters of the mixture for simulated data from three mixture vM distribution

Lower κ (model 4.10)				Larger κ (model 4.11)		
c	\hat{p}_c	$\hat{\mu}_c$	$\hat{\kappa}_c$	\hat{p}_c	$\hat{\mu}_c$	$\hat{\kappa}_c$
1	0.84	2.95	1.85	0.83	2.00	3.86
2	0.13	0.98	2.10	0.13	3.20	6.00
3	0.01	0.15	2.01	0.04	0.60	7.68
4	0.001	0.03	1.83	0.005	0.20	7.04
5	0.0001	0.002	1.86	0.00009	0.07	6.98
6	0.00002	0.001	1.90	0.00003	0.03	6.90
7	0.000002	0.00002	1.87	0.00001	0.008	6.87
8	0.0000004	0.003	1.89	0.000005	0.01	6.82
9	0.00000005	0.002	1.87	0.000002	0.02	6.84
10	0.00000001	0.001	1.86	0.000003	0.002	6.90

We choose a maximum value for unknown modal number as $C=10$. To avoid convergence problems for lower κ data set, we take hyper-parameters as $\mu_0 = 0, \kappa_0 = 7, a_0 = 0, b_0 = 3.75$, we also take a fixed value as $\alpha = 0.5$. For larger κ , we take α and κ_k parameters as shown below

$$\alpha \sim \text{Gamma}(2, 2), \quad \kappa_k \sim \text{Uniform}(3.75, 10)$$

Posterior means of the mixing probabilities and parameters both lower concentration and larger concentration parameters κ for vM distribution are summarized in Table 4.1, respectively. In lower κ , 98 % of the simulated data set is drawn from three distinctive mixture clusters. In larger κ , there are three components with associated probabilities, 0.83, 0.13, 0.04, respectively. From these results, we observe that the finding cluster number is the same as true cluster size of simulated from three mixture vM distribution.

On the other hand, for wC distribution, we use the following steps:

$$\mu_k \sim \text{vM}(0, 7), \rho_k \sim \text{Beta}(0.5, 0.5) \quad \alpha \sim \text{Gamma}(1, 1)$$

Posterior means of the mixing probabilities and parameters both lower concentration and larger concentration parameter for wC distribution are summarized in Table 4.2, respectively. 98 % of the simulated data set is drawn from three distinctive clusters in lower ρ . Three components of the associated probabilities are 0.49, 0.22, 0.26, in larger ρ respectively. Consequently, our model is also working well for wC distribution, but it is observed that the results of the larger ρ parameters are slightly better than the results of lower ρ parameters.

Table 4.2: Posterior means of the mixing probabilities and parameters of the mixture for simulated data from three mixture wC distribution

Lower ρ (model 4.12)				Larger ρ (model 4.13)		
c	\hat{p}_c	$\hat{\mu}_c$	$\hat{\rho}_c$	\hat{p}_c	$\hat{\mu}_c$	$\hat{\rho}_c$
1	0.68	2.93	0.31	0.49	3.02	0.69
2	0.19	0.32	0.43	0.22	0.99	0.90
3	0.07	0.13	0.44	0.26	0.50	0.81
4	0.03	0.1	0.49	0.02	0.07	0.54
5	0.01	0.05	0.49	0.005	0.01	0.49
6	0.005	0.001	0.50	0.002	0.008	0.50
7	0.003	0.002	0.50	0.0001	0.02	0.50
8	0.002	0.004	0.51	0.00006	0.003	0.51
9	0.001	0.009	0.49	0.00004	0.005	0.49
10	0.002	0.02	0.51	0.00007	0.009	0.51

We compare these with kernel density estimation, the predictive density estimation of mixture DP and true model for each simulated data set in Fig 4.2 and Fig 4.3, respectively. The predictive density estimations of mixture DP approach are closer to true models.

4.4.2 Monte Carlo study

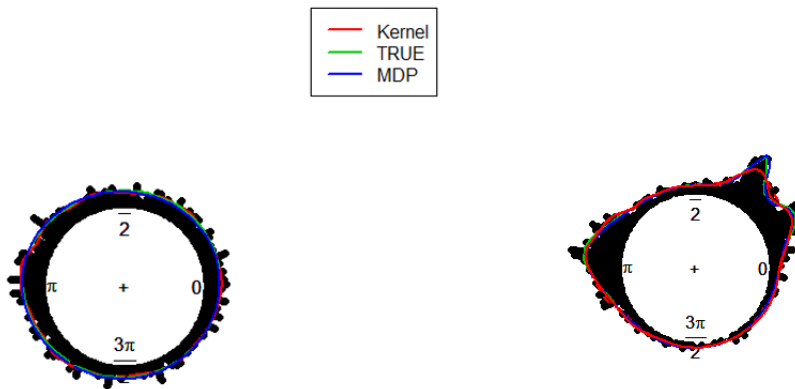
In this section, we conducted a Monte Carlo study to investigate the performance of the proposed approach for a number of two mixture vM and wC distributions.



(a) Lower κ (Model 4.10)

(b) Larger κ (Model 4.11)

Figure 4.2: Comparison of mixture DP vM model and kernel density estimation and true density for simulated data from three mixture vM distribution



(a) Lower ρ (Model 4.12)

(b) Larger ρ (Model 4.13)

Figure 4.3: Comparison of mixture DP wC model and kernel density estimation and true density for simulated data from three mixture wC distribution

The sample size was chosen as 100 and 500, respectively and 250 replicates were performed in each simulation design.

First, we examine a simulated series with parameters as $\boldsymbol{\mu} = (\mu_1, \mu_2) = (1, 3)$, $\boldsymbol{\kappa} = (\kappa_1, \kappa_2) = (5, 8)$, $\boldsymbol{p} = (p_1, p_2) = (0.75, 0.25)$ from two mixture vM distribution. The rose diagram of this simulated data clearly shows bi-modality in Fig. 4.4.

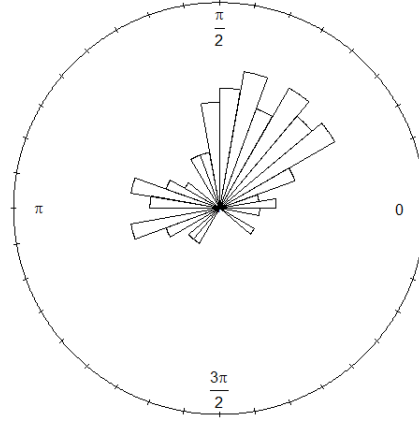


Figure 4.4: Rose diagram of two mixture vM data

In proposed models, first, we choose both as $C = 2$ for a number of latent class. (same as the number of cluster size with simulated data) and choose $C = 3$ for each simulation design. Additionally, for DP mixture vM model parameters, we set as $\mu_k \sim \text{vM}(0.1, 0.1)$, $\kappa_k \sim \text{Gamma}(0.01, 0.01)$, $k = 1, 2$ and $\alpha \sim \text{Uniform}(0.5, 10)$ while DP mixture wC model parameters, we set as $\mu_k \sim \text{vM}(0.1, 0.1)$, $\rho_k \sim \text{Beta}(1, 1)$, $k = 1, 2$ and $\alpha \sim \text{Uniform}(0.5, 10)$

The resulting estimators and their Monte Carlo properties and true values of estimators given in parenthesis for both DP mixture two vM model ($C = 2$) and DP mixture three vM model ($C = 3$) are presented in Table 4.3 and in Table 4.4. Also, box-plots for estimators obtained from the Monte Carlo experiment for $C = 2$ are given in Fig 4.5, Fig 4.6 and Fig. 4.7.

Performance of our estimation method is evaluated through relative bias (R. Bias), Monte Carlo standard error (MCSE), and standard error (SE). These performance

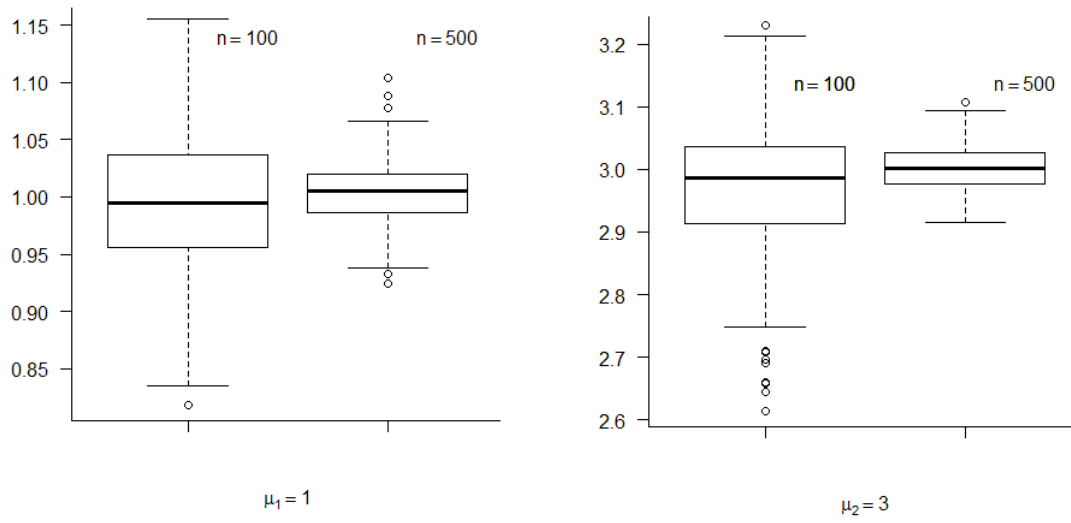


Figure 4.5: Boxplots of estimated circular mean directions for two mixture vM distributions

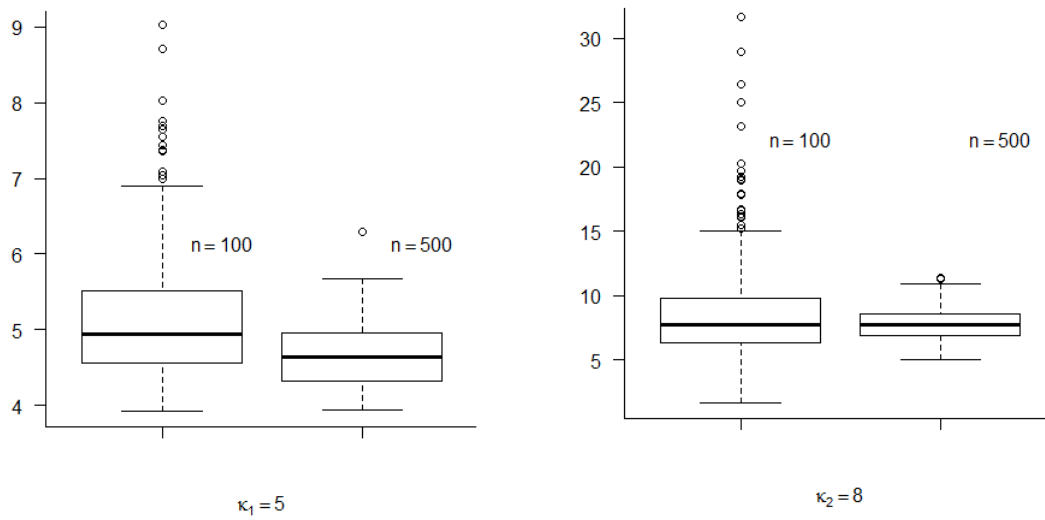


Figure 4.6: Boxplots of estimated concentration parameters for two mixture vM distributions

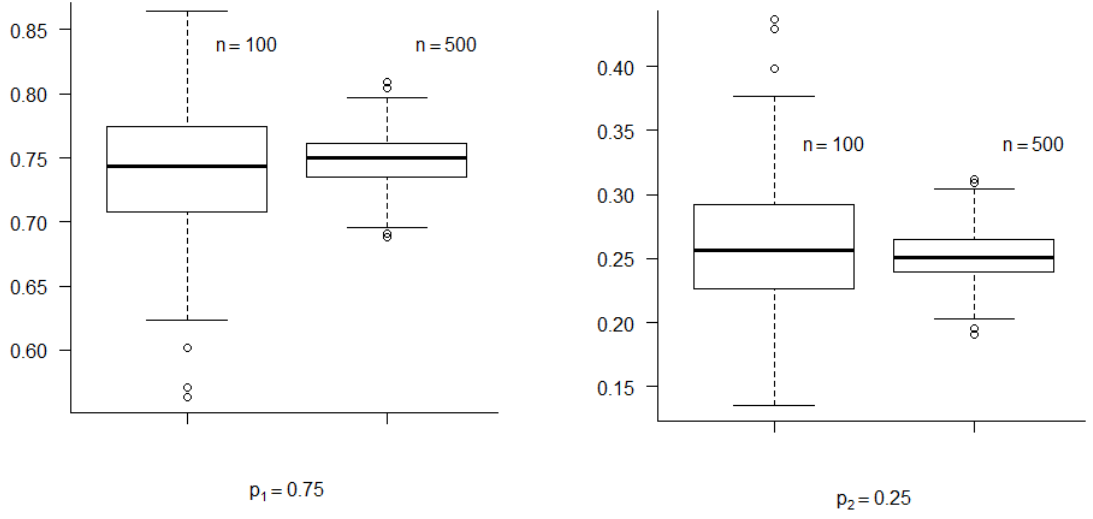


Figure 4.7: Boxplots of estimated weight parameters for two mixture vM distributions measures are computed as

$$R.Bias = \left(\frac{\bar{\hat{\beta}} - \beta}{\beta} \right)$$

where β is true value for estimate of interest, $\bar{\hat{\beta}} = \sum_{i=1}^B \hat{\beta}_i / B$, B is the number of replicates performed, $\hat{\beta}_i$ is the estimate of interest within each of the $i = 1, 2, \dots, B$. In order to determine an assessment of the uncertainty in estimate of interest between simulations, we use MCSE, which is calculated as the standard deviation of the estimates of interest from all simulations, $\sqrt{[1/(B-1)] \sum_{i=1}^B (\hat{\beta}_i - \bar{\hat{\beta}})^2}$. Alternatively, the average of the estimated within simulation SE for the estimate of interest is computed as $\sum_{i=1}^B SE(\hat{\beta}_i) / B$. If the estimates are unbiased, then, MCSE should be close to the average of the estimated within simulation SE (see; [Schafer and Graham \(2002\)](#)).

In order to compare $C = 2$ and $C = 3$ models, we use Deviance and BIC model criteria. We compute these criteria using the sets of parameters, among MCMC draws, that maximize the posterior distribution denoted by MAP (Maximum at Posterior). Let $\hat{\psi}$ the MAP estimators, then we compute Deviance $= -2\log(\theta | \hat{\psi})$ and BIC $= -2\log(\theta | \hat{\psi}) + (\#number\ of\ parameters) \times \log(n)$. The lowest criteria indicate the best model and their standard errors are given in parenthesis.

From these results we observe that concentration parameters κ_1 , κ_2 appear to have

small relative bias. Other all of five parameters are approximately unbiased for $n = 100$. For $n = 500$, we observe that same results but less Monte Carlo standard error (MCSE) and standard error (SE). To compare DP mixture vM models for $C = 2$ and $C = 3$, we compute Deviance and BIC criteria. These criteria give a slight to the true two group model ($C = 2$) and they show clearly favours true model in Table 4.5.

Table 4.3: Monte Carlo study results for DP mixture vM model ($C = 2$)

$n = 100$	Est.	R.Bias	MCSE	SE	$n = 500$	Est.	R.Bias	MCSE	SE
$\hat{\mu}_1$ (1)	1.00	0	0.06	0.06	$\hat{\mu}_1$	1.00	0	0.03	0.03
$\hat{\mu}_2$ (3)	2.98	-0.006	0.09	0.01	$\hat{\mu}_2$	3.00	0	0.04	0.04
$\hat{\kappa}_1$ (5)	4.99	-0.004	0.78	0.85	$\hat{\kappa}_1$	4.66	-0.07	0.41	0.39
$\hat{\kappa}_2$ (8)	8.02	0.003	3.33	3.04	$\hat{\kappa}_2$	7.79	-0.03	1.29	1.26
\hat{p}_1 (0.75)	0.74	-0.01	0.05	0.05	\hat{p}_1	0.75	0	0.02	0.02
\hat{p}_2 (0.25)	0.26	0.04	0.05	0.05	\hat{p}_2	0.25	0	0.02	0.02
$\hat{\alpha}$ (-)	1.44	-	0.15	1.05	$\hat{\alpha}$	1.40	-	0.07	0.99

Table 4.4: Monte Carlo study results for DP mixture vM model ($C = 3$)

$n = 100$	Est.	R.Bias	MCSE	SE	$n = 500$	Est.	R.Bias	MCSE	SE
$\hat{\mu}_1$	2.37	-	1.01	1.06	$\hat{\mu}_1$	2.14	-	1.10	0.87
$\hat{\mu}_2$ (3)	2.91	-0.03	0.39	0.18	$\hat{\mu}_2$	3.00	0	0.13	0.06
$\hat{\mu}_3$ (1)	1.00	0	0.09	0.07	$\hat{\mu}_3$	1.00	0	0.04	0.03
$\hat{\kappa}_1$	2.46	-	3.14	2.98	$\hat{\kappa}_1$	3.68	-	2.48	2.42
$\hat{\kappa}_2$ (8)	7.92	-0.01	3.39	2.99	$\hat{\kappa}_2$	8.15	0.01	1.92	1.54
$\hat{\kappa}_3$ (5)	5.21	0.04	0.98	0.90	$\hat{\kappa}_3$	4.75	-0.05	0.47	0.41
\hat{p}_1	0.08	-	0.09	0.07	\hat{p}_1	0.07	-	0.09	0.07
\hat{p}_2 (0.25)	0.21	-0.16	0.08	0.06	\hat{p}_2	0.22	0.12	0.06	0.04
\hat{p}_3 (0.75)	0.71	-0.05	0.09	0.07	\hat{p}_3	0.71	-0.05	0.08	0.07
$\hat{\alpha}$	5.81	-	0.85	2.34	$\hat{\alpha}$	5.77	-	0.84	2.35

In second example, we assume model parameters as $\boldsymbol{\mu} = (\mu_1, \mu_2) = (0.5, 3)$, $\boldsymbol{\rho} = (\rho_1, \rho_2) = (0.9, 0.7)$, $\boldsymbol{p} = (p_1, p_2) = (0.20, 0.80)$ from two mixture wrapped Cauchy distribution. Rose diagram of two mixture wC simulated data is displayed in Fig. 4.8.

We find these estimates for $C = 2$ as $\hat{\boldsymbol{\mu}} = (0.50, 3.00)$, $\hat{\boldsymbol{\rho}} = (0.88, 0.69)$, $\hat{\boldsymbol{p}} =$

Table 4.5: DP mixture vM model fits for Monte Carlo study

	C=2	C=3
n=100		
Deviance	223.48(16.24)	227.71(16.50)
BIC	246.51	264.55
n=500		
Deviance	1143.07(36.79)	1144.91(33.62)
BIC	1174.15	1194.62
No of parameters	5	8

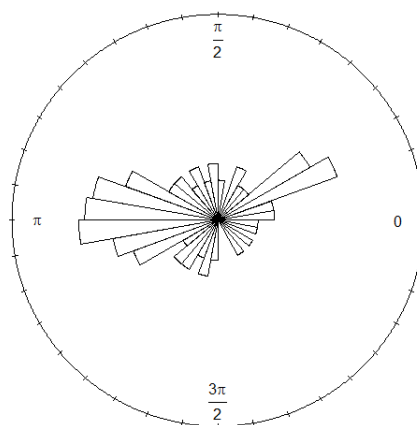


Figure 4.8: Rose diagram of two mixture wC data

$(0.20, 0.80)$ and the estimate concentration parameter of DP is $\hat{\alpha} = 5.23$ for $n = 100$. Additionally, for $n = 500$, we find same results but less MCSE and SE compared to $n = 100$ in Table 4.6. Finally, these results show that the estimators of all six parameters are approximately unbiased. The box-plots of estimated parameters for DP mixture wC model for $C = 2$ are displayed in Fig 4.9, Fig 4.10 and Fig. 4.11. For $C = 3$, we find similar estimates, but the estimates have more MCSE and SE compared to $C = 2$ in Table 4.7.

Similarly, to compare DP mixture wC models for $C = 2$ and $C = 3$, we compute Deviance and BIC criteria. These criteria give a slight to the true two group model ($C = 2$) and they show clearly favours true model in Table 4.8.

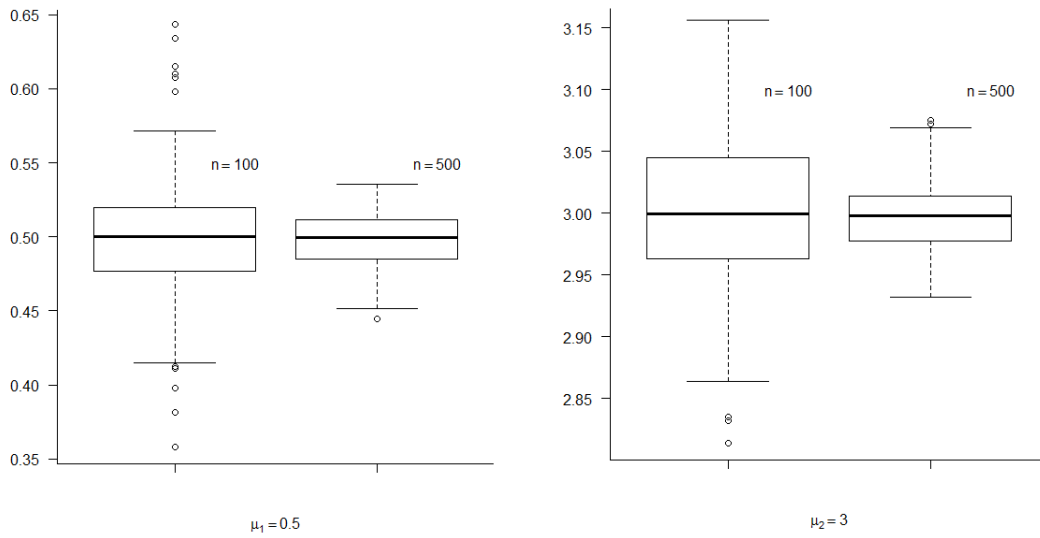


Figure 4.9: Boxplots of estimated circular mean directions for two mixture wC distributions

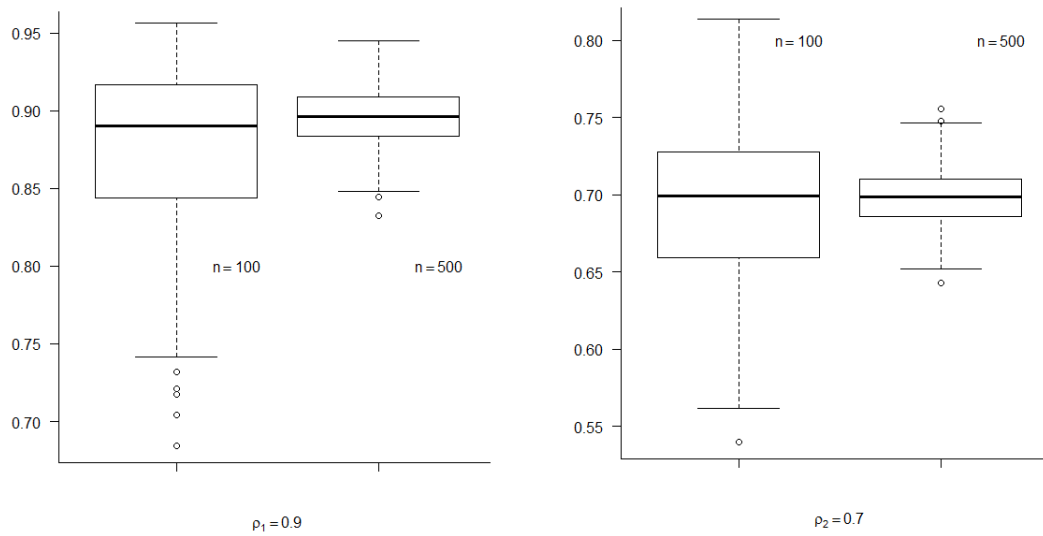


Figure 4.10: Boxplots of estimated concentration parameters for two mixture wC distributions

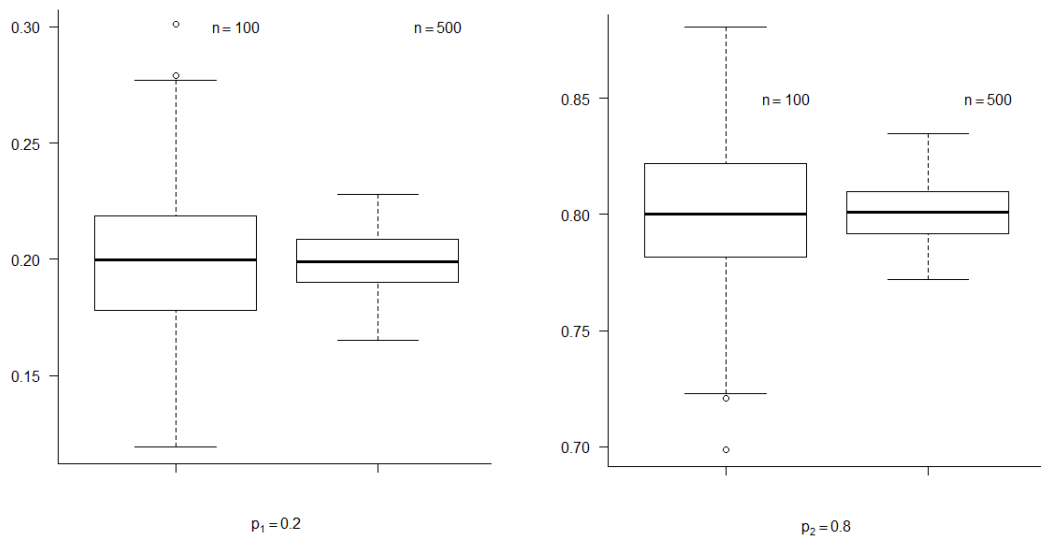


Figure 4.11: Boxplots of estimated weight parameters for two mixture wC distributions

4.4.3 Real data examples

In this section, we considered two real data sets to illustrate our circular DP modelling described in Section 4.3. Firstly, we used turtle data which is cited by [Stephens \(1969\)](#). Secondly, we analysed ant data that were randomly selected during an animal orientation experiment described in [Jander \(1957\)](#). Both data sets can be found in CircNNTSR package in R as online or in [Fisher's \(1991\)](#) book.

4.4.3.1 Turtle data

We re-consider turtle data by Gould's cited by [Stephens, \(1969\)](#). The data consists of orientations of 76 turtles after laying eggs. Previously, [Stephens \(1969\)](#) analysed this data set using two component mixture vM distribution under the assumption of same concentration and modes π radians. [Mardia \(1975\)](#) fitted two mixture vM distribution with having different concentration and modes parameters. [Wang and Gelfand \(2013\)](#) used the general projected normal model to fit this data and two mixture component vM distribution is not plausible for model specification. Main drawback of these approaches is that number of modes is fixed empirically prior to the estima-

Table 4.6: Monte Carlo study results for DP mixture wC model ($C = 2$)

$n = 100$	Est.	R.Bias	MCSE	SE	$n=500$	Est.	R.Bias	MCSE	SE
$\hat{\mu}_1$ (0.50)	0.50	0	0.04	0.05	$\hat{\mu}_1$	0.50	0	0.02	0.02
$\hat{\mu}_2$ (3.00)	3.00	0	0.06	0.06	$\hat{\mu}_2$	3.00	0	0.03	0.03
$\hat{\rho}_1$ (0.90)	0.88	-0.02	0.05	0.06	$\hat{\rho}_1$	0.90	0	0.02	0.02
$\hat{\rho}_2$ (0.70)	0.69	-0.01	0.05	0.05	$\hat{\rho}_2$	0.70	0	0.02	0.02
\hat{p}_1 (0.20)	0.20	0	0.03	0.05	\hat{p}_1	0.20	0	0.01	0.02
\hat{p}_2 (0.80)	0.80	0	0.03	0.05	\hat{p}_2	0.80	0	0.01	0.02
$\hat{\alpha}$	5.23	-	0.32	2.52	$\hat{\alpha}$	5.28	-	0.14	2.51

Table 4.7: Monte Carlo study results for DP mixture wC model ($C = 3$)

$n = 100$	Est.	R.Bias	MCSE	SE	$n=500$	Est.	R.Bias	MCSE	SE
$\hat{\mu}_1$ (0.50)	0.53	0.06	0.30	0.07	$\hat{\mu}_1$	0.50	0	0.02	0.02
$\hat{\mu}_2$ (3.00)	2.99	-0.003	0.35	0.46	$\hat{\mu}_2$	3.00	0	0.14	0.07
$\hat{\mu}_3$	2.89	-	0.51	0.43	$\hat{\mu}_3$	2.80	-	0.71	0.74
$\hat{\rho}_1$ (0.90)	0.87	-0.03	0.07	0.06	$\hat{\rho}_1$	0.90	0	0.02	0.02
$\hat{\rho}_2$ (0.70)	0.68	-0.03	0.13	0.16	$\hat{\rho}_2$	0.72	0.03	0.07	0.05
$\hat{\rho}_3$	0.68	-	0.15	0.12	$\hat{\rho}_3$	0.62	-	0.13	0.15
\hat{p}_1 (0.20)	0.19	-0.05	0.03	0.05	\hat{p}_1	0.19	-0.05	0.02	0.03
\hat{p}_2 (0.80)	0.20	-	0.18	0.17	\hat{p}_2	0.45	-	0.28	0.17
\hat{p}_3	0.61	-	0.17	0.18	\hat{p}_3	0.36	-	0.27	0.17
$\hat{\alpha}$	4.60	-	1.40	2.46	$\hat{\alpha}$	2.89	-	1.82	2.01

tions. On the other hand, our approach is flexible that number of modes is left unspecified and estimated along with other parameters. For prior distributions of DP mixture vM model, we consider concentrated (informative) priors for circular mean parameters as $\mu_k \sim \text{vM}(4, 7)$ and for other parameters, we consider weak priors as $\kappa_k \sim \text{Gamma}(6, 1)$ and $\alpha \sim \text{Gamma}(2, 2)$. For turtle data, if we consider weak priors for μ_k , there arise convergence problem for μ_k . In order to handle this problem, we suggest to use concentrated priors or informative priors for μ_k . Main reason of convergence problem is that model complexity is increasing or C is increasing. Hence, weak priors might not provide enough information in the data for such a complex hierarchical structure. Fig. 4.12 a shows posterior mean density estimate from our mixture DP vM approach with kernel density estimate. Table 4.10 is summarized

Table 4.8: DP mixture wC model fits for Monte Carlo study

	C=2	C=3
n=100		
Deviance	259.42(18.35)	261.57(19.80)
BIC	282.44	298.42
n=500		
Deviance	1320.01(42.46)	1321.75(40.92)
BIC	1343.04	1371.45
No of parameters	5	8

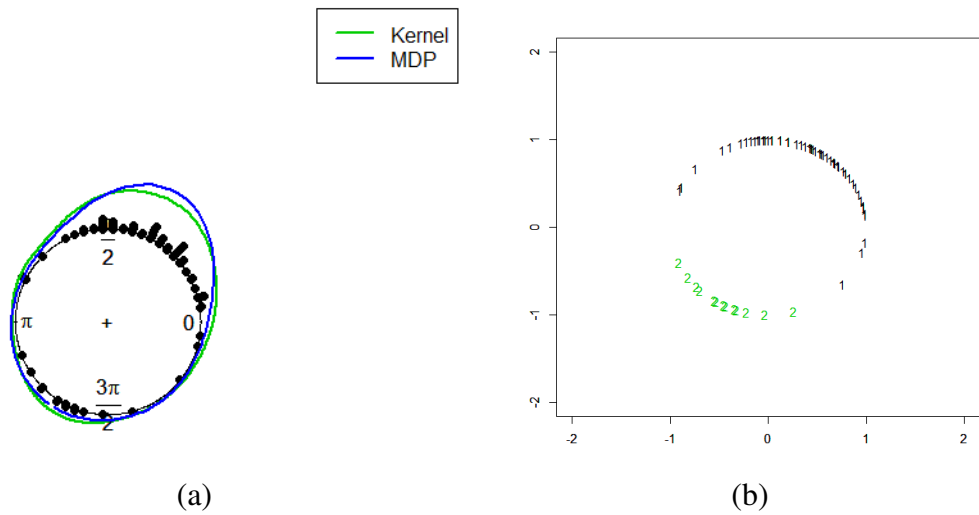


Figure 4.12: (a) Comparison of mixture DP vM model and kernel density estimation for turtle data. (b) Identified clusters for turtle data

Table 4.9: Estimates of parameters for turtle data

Method	SU	EM	FCD	Spurr and Koutbeiy	MSBC	DP vM (C=2)
$\hat{\mu}_1$	67.67	63.47	63.50	63.20	63.15	63.31
$\hat{\mu}_2$	242.83	241.20	241.25	240.20	241.13	241.15
$\hat{\kappa}_1$	3.00	2.65	2.65	2.91	2.75	3.18
$\hat{\kappa}_2$	4.49	8.61	7.43	4.81	7.43	5.37
\hat{p}	0.82	0.84	0.84	0.82	0.83	0.82
U^2	0.032	0.019	0.019	0.018	0.018	0.020

posterior inferences of turtle data for $C = 10$. As far as the concentration parameter α is concerned, we consider three assumptions on it as α is set to 0.5, or uniform distribution with support (0.5,10) or Gamma distribution with shape and scale parameters that are equal to 2, respectively. Three inferences on α parameter have similar results, that is, the specific choice of the concentration parameter α is does not affect to model parameters' posterior results. For $\alpha \sim \text{Gamma}(2,2)$, we observe that there are two distinct clusters about % 91, associated with probabilities, 0.81, 0.10, respectively in Table 4.10. Here, we re-analyse this data for $C = 2$, but we use weak priors as $\mu_k \sim \text{vM}(0.1,0.1)$, and then, we get similar results with $C = 10$. Additionally, we make a comparison other methods which are self updating (SU), expectation and maximization (EM) algorithm, fuzzy-c directions (FCD) algorithm, Spurr and Koutbeiy algorithm and mean shift-based clustering (MSBC) in the literature. (see e.g. [Chang-Cien, et al. \(2012\)](#); [Hung, et al. \(2012\)](#)). Results of these methods for turtle data are taken from [Chang-Cien', et al. \(2012\)](#) paper . We find that the analysis results from SU, EM, FCD , Spurr and Koutbeiy, MSBC and DP vM model are very similar in Table 4.9 and most of turtles move around 63^0 and other turtles move around 241^0 . Fig. 4.12.b shows these two identified clusters. To compare results of six different fitting methods, we compute Watson- U^2 goodness of fit test of each method for two mixture vM distributions. Clearly, U^2 measures the discrepancy between the empirical distribution function denoted by F_n and the distribution function denoted by F . From this result, DP vM model shows superiority both number of modes and the estimates of parameter for turtle data.

4.4.3.2 Ant data

As second example, we analyse the ant data given by Appendix B.7 Fisher (1993). This data consists of 100 observations which are randomly collected by Jander (1957). Fisher (1993) shows that the vM distribution is not suitable model for this data with goodness of fit test statistics. Pewsey (2002) demonstrates that there is no evidence to show that the underlying distribution is asymmetric with test of circular reflective symmetry. Abe and Pewsey (2011) re-analyse this data set and the best fit based on model selection criteria is wC distribution. Here, we analyse this data in terms of number of modes using DP mixture wC model. We choose the following prior settings as $\mu_k \sim \text{vM}(1, 7)$, $\rho_k \sim \text{Beta}(3, 1)$, $\alpha \sim \text{Gamma}(2, 2)$ and $C = 10$. We have also examined robustness of the concentration parameter α of DP. We found no substantial differences on parameter inferences for ant data. Summary posterior inferences and posterior distribution of model parameters for ant data are indicated in Table 4.11. For $\alpha \sim \text{Gamma}(2, 2)$, we conclude that there is one cluster about % 95 proportions. This result is also consistent with the number of modes in the literature. Figure 4.13 shows the predictive density estimation for each group that are obtained using the mixture DP and Kernel density approaches for circular data. To sum up, our mixture DP wC approach is closer to that obtained by Kernel approach.

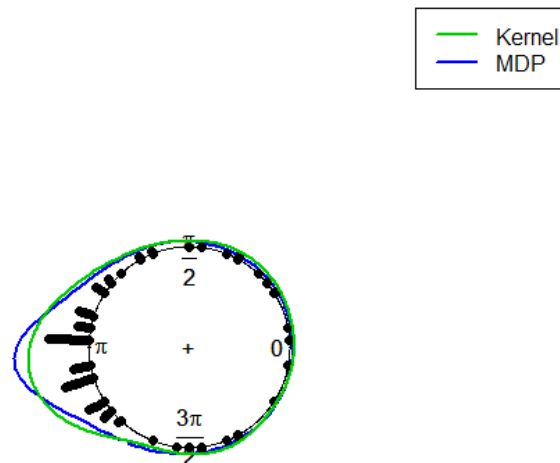


Figure 4.13: Comparison of mixture DP and kernel density estimation for ant data.

Table 4.10: Posterior means of the mixing probabilities and parameters of turtle data

c	$\alpha = 0.5$			$\alpha \sim U(0.5, 10)$			$\alpha \sim \text{Gamma}(2, 2)$		
	\hat{p}_c	$\hat{\mu}_c$	$\hat{\kappa}_c$	\hat{p}_c	$\hat{\mu}_c$	$\hat{\kappa}_c$	\hat{p}_c	$\hat{\mu}_c$	$\hat{\kappa}_c$
1	0.82 (0.05)	1.11(0.08)	3.52 (0.69)	0.80(0.05)	1.11 (0.08)	3.75 (0.67)	0.81 (0.05)	1.11 (0.08)	3.44 (0.69)
2	0.12 (0.06)	4.16 (0.27)	5.54(2.34)	0.08(0.06)	4.15(0.33)	5.65 (2.46)	0.11(0.06)	4.14(0.28)	5.60(2.35)
3	0.04 (0.05)	4.05(0.36)	5.63(2.48)	0.05(0.04)	4.12 (0.37)	5.51(2.43)	0.04 (0.05)	4.07(0.38)	5.65(2.51)
4	0.02 (0.03)	4.01(0.41)	5.66(2.43)	0.03 (0.03)	4.07 (0.39)	5.57 (2.47)	0.02 (0.03)	4.02 (0.40)	5.61 (2.5)
5	0.006(0.01)	4.00 (0.40)	5.66(2.52)	0.02(0.02)	4.05 (0.39)	5.64(2.51)	0.009(0.02)	4.01(0.39)	5.60(2.45)
6	0.002(0.005)	4.00(0.41)	5.65 (2.48)	0.009 (0.02)	4.03 (0.40)	5.69 (2.46)	0.005 (0.01)	4.01(0.40)	5.61(2.43)
7	0.0007(0.003)	3.99(0.39)	5.69 (2.41)	0.006(0.01)	4.01(0.40)	5.60 (2.47)	0.002(0.006)	4.00 (0.39)	5.59 (2.42)
8	0.0002(0.001)	4.00 (0.39)	5.72(2.52)	0.004 (0.008)	4.01(0.39)	5.64 (2.43)	0.001(0.004)	4.00 (0.40)	5.70 (2.48)
9	0.00008(0.0005)	4.00 (0.40)	5.62(2.42)	0.003(0.007)	4.00(0.40)	5.68(2.40)	0.0007 (0.003)	4.00(0.39)	5.70(2.48)
10	0.00004(0.0004)	4.00(0.40)	5.71(2.53)	0.006(0.01)	4.02(0.40)	5.64 (2.49)	0.001(0.004)	4.00 (0.39)	5.65(2.47)

Table 4.11: Posterior means of the mixing probabilities and parameters of ant data

c	\hat{p}_c	$\alpha = 0.5$		$\alpha \sim U(0.5, 10)$			$\alpha \sim \text{Gamma}(2, 2)$		
		$\hat{\mu}_c$	$\hat{\rho}_c$	\hat{p}_c	$\hat{\mu}_c$	$\hat{\rho}_c$	\hat{p}_c	$\hat{\mu}_c$	$\hat{\rho}_c$
1	0.96 (0.04)	3.22(0.07)	0.67 (0.04)	0.93(0.05)	3.23 (0.06)	0.68 (0.04)	0.95 (0.05)	3.23 (0.06)	0.67 (0.69)
2	0.03 (0.03)	1.00 (0.37)	0.74(0.20)	0.03 (0.03)	1.00 (0.37)	0.75(0.20)	0.03(0.03)	1.00(0.37)	0.74(0.20)
3	0.01 (0.02)	1.00(0.37)	0.79(0.20)	0.02 (0.02)	0.96(0.39)	0.76(0.20)	0.02 (0.02)	0.96(0.39)	0.76(0.20)
4	0.004 (0.009)	0.97(0.37)	0.79(0.19)	0.009 (0.01)	0.97 (0.39)	0.78 (0.20)	0.008 (0.03)	0.97 (0.39)	0.78 (0.20)
5	0.001(0.004)	0.98 (0.42)	0.80(0.19)	0.005(0.009)	0.98 (0.39)	0.80(0.20)	0.005(0.009)	0.98(0.39)	0.80(0.20)
6	0.0004(0.002)	0.99(0.39)	0.79 (0.19)	0.002 (0.006)	0.97 (0.40)	0.79 (0.19)	0.002 (0.006)	0.97(0.40)	0.79(0.19)
7	0.0001(0.0007)	1.00(0.39)	0.80 (0.19)	0.001(0.01)	0.97(0.40)	0.80 (0.19)	0.001(0.004)	0.97 (0.40)	0.80 (0.19)
8	0.00005(0.0004)	1.02 (0.39)	0.80(0.19)	0.0009 (0.003)	0.98(0.39)	0.79 (0.19)	0.0009(0.003)	0.99 (0.40)	0.80 (0.19)
9	0.00002(0.0002)	1.01 (0.39)	0.79(0.19)	0.0005(0.002)	0.99(0.38)	0.79(0.19)	0.0005 (0.002)	0.99(0.38)	0.79(0.19)
10	0.000007(0.00008)	1.02(0.41)	0.79(0.20)	0.001(0.003)	1.03(0.39)	0.79 (0.19)	0.001(0.004)	1.02 (0.39)	0.79(0.19)

4.5 Discussion

In many environmental and ecological studies, there is an uncertainty about the number of modes in circular data. For von Mises distribution, the main problem is convergence in the event of the multiple concentration parameter κ_k for each class k . [Bhattacharya and SenGupta \(2009\)](#) consider that G_0 is bivariate conjugate distribution under Polya-Urn representation, and then, they observe convergence problem in their MCMC application. To handle this problem, we consider that G_0 is bivariate non-conjugate distribution under stick breaking representation. Hence, our model overcomes the problem. Additionally, for wrapped Cauchy distribution, our proposed DP mixture model works quite well in the event of multiple μ_k and ρ_k parameters for each class k . Finally, it is seen that our models perform well in terms of estimates of parameters and number of modes for both simulated and real life data sets.

As future extension, we would like to emphasize that our DP mixture approach may be applied to mixture of skew circular distributions, namely, the sine-skewed von Mises and wrapped Cauchy distributions. In this chapter we propose two specific DP mixture models to determine the number of modes for time-independent circular data. In the next chapter, we propose Bayesian semi-parametric model for time-dependent circular data.

CHAPTER 5

BAYESIAN SEMI-PARAMETRIC MODEL FOR MULTI-MODAL CIRCULAR TIME SERIES DATA

The aim of this chapter is to define a new model for circular time series based on Dirichlet process (DP) mixture on a family of random probability measures indexed by the parameters of Möbius time series model. This new model is to define multi-modal circular time series as dependent mixtures of von Mises distributions. Our contribution is to provide a flexible circular time series model which overcomes both changing concentration parameter over time and the problem of multi-modality for time-dependent circular data. Real data examples are given from meteorology (wind directions) to illustrate our multi-modal circular time series approach.

5.1 Introduction

Research in this chapter is motivated by an hourly wind direction dataset we have received from a north-western wind farm details of which are given in section 5.4.2.2. As seen in Figure 5.1, hourly wind direction data at hand presented a multi-modal structure. Aim of this chapter is to develop a flexible methodology to analyse multi-modal circular time series data.

Circular observations with time structure such as the hourly or daily wind directions at fixed location have limited literature. [Fisher and Lee \(1994\)](#) proposed two main approaches used to model circular time series. For noisy series, they recommend to use circular model, while for non-noisy series, they propose transformation to a linear series with a link function. Another useful process called, a Wrapped Autoregressive

process WAR, was introduced by [Breckling \(1989\)](#). This method is based on to wrap a linear random variable around the circle. All of these methods are described in [Sec. 5.2](#).

In the context of Bayesian circular time series, [Coles \(1998\)](#) used MCMC methods to fit such class of models to circular data using wrapping process. [Ravindiran \(2002\)](#) developed a Bayesian methodology for the wrapping process based on data augmentation approach. Recently, [Lasinio, et al. \(2012\)](#) introduced Bayesian hierarchical model to overcome circular data based on adaptive truncation method. However, circular time series based on von Mises model have not received as much attention in Bayesian framework. In this study, we propose a new methodology to analyse multi-modal circular time series data based on Bayesian non-parametric approach.

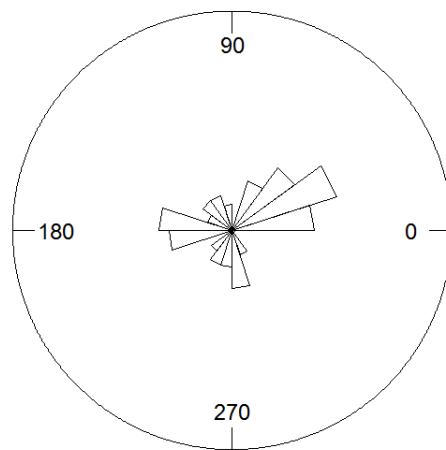


Figure 5.1: Rose diagram of a hourly wind direction data from Turkey

In many real data applications, the distribution of circular measurements is clearly multi-modal, in particular, wind direction in [Fig. 5.1](#). On this subject, [Holzmann et al.\(2006\)](#) introduced a new class of circular time series model based on hidden Markov models (HMM) for von Mises and wrapped distributions. They drew attention to Bayesian analysis of HMM, in particular, how to decide the number of states. In this chapter, we provide DP mixture circular Möbius time series model taking account of multi-modal structure.

Rest of this chapter is organized as follows. In the following section, we provide a brief summary of existing models for circular time series. In [Section 5.3](#), we intro-

duce Dirichlet process (DP) mixture Möbius time series model for circular data. Our Bayesian methodology is applied on simulated and real data sets for illustration in Section 5.4. We give some concluding remarks and future directions in Section 5.5

During this chapter, all circular valued random variables are assumed to take on values in the interval $[-\pi, \pi)$. To achieve symmetric support $[-\pi, \pi)$, we transform from χ to $\chi' = (\chi + \pi) \bmod 2\pi - \pi$.

5.2 Review of circular time series models

In this section, we present a brief review of time series models for circular observations defined in literature.

5.2.1 Linked process

A linked process is introduced by Fisher and Lee (1994). This method is based on a link function g . It is to associate a strictly monotonically increasing function which transforms values from real line $(-\infty, \infty)$ to circle $(-\pi, \pi)$. If $\{Y_t\}_{t=1,2,\dots}$ is a process on the line, g is link function, and $\mu \in [0, 2\pi)$, then the corresponding linked circular process $\{\Theta_t\}_{t=1,2,\dots}$ on the circle is defined by

$$\Theta_t = g(Y_t) + \mu.$$

For illustration, the useful link function form is $g(y) = 2\pi F(y)$, where $F(y)$ is a distribution function.

5.2.2 Circular autoregressive process

A circular AR(p) process, CAR(p) is also introduced by Fisher and Lee (1994). CAR(p) process, with link function g if Θ_t , given $\Theta = \theta_{t-1}, \theta_{t-2} = \theta_{t-2}, \dots, \theta_1 = \theta_1$ is von Mises $\text{vM}(\mu_t, \kappa)$ for $t > p$, where

$$\mu_t = \mu + g(\lambda_1 g^{-1}(\theta_{t-1} - \mu) + \dots + \lambda_p g^{-1}(\theta_{t-p} - \mu))$$

where μ_t is a mean direction and κ is a constant concentration parameter.

5.2.3 Wrapped process

The wrapping approach leads to following definition. Let $\{Y_t\}_{t=1,2,\dots}$ be a process on the real line and a corresponding process on the circle is $\{\Theta_t\}_{t=1,2,\dots}$ then, the wrapping process is obtained as follows

$$Y_t = \Theta_t + 2\pi k_t$$

where k_t is an unobserved integer. Thus, fitting such process leads to missing data problem. This problem can be handled with MCMC in Bayesian aspect. (see, e.g, [Coles \(1998\)](#); [Ravindran \(2002\)](#); [Lasinio et all \(2012\)](#)).

5.2.4 Projected Normal process

This process is defined as following construction. Let $\{X_t, Y_t\}_{t=1,2,\dots}$ a process on the plane. Thus, the radial projection to unit circle gives a corresponding process $\{\Theta_t\}_{t=1,2,\dots}$ on the circle is described as follows

$$X_t = R_t \cos(\Theta_t), \quad Y_t = R_t \sin(\Theta_t)$$

If $\{X_t, Y_t\}_{t=1,2,\dots}$ is a stationary Gaussian process then Θ_t is a projected normal distribution. Here, radial part $\{R_t\}_{t=1,2,\dots}$ of projected normal distribution is unobserved . Thus fitting such process leads to missing data problem. This problem can be handled with MCMC or EM algorithms.

5.2.5 Möbius time series model

[Downs and Mardia \(2002\)](#) introduced a circular-circular regression model. This model can be adopted for circular time series model (see, [Hughes \(2007\)](#)). Here, consider the mapping as follows

$$\tan \frac{1}{2}(\theta_t - \mu) = \lambda \tan \frac{1}{2}(\theta_{t-1} - \mu) \quad (5.1)$$

where μ is circular location parameter) on unit circle, λ is a slope parameter in closed interval $[-1, 1]$, and θ_t and θ_{t-1} are circular variables observed at time t and $t - 1$

respectively. Equation 5.1 has the unique solution as follows

$$\theta_t = \mu + 2\text{atan}\{\lambda \tan\frac{1}{2}(\theta_{t-1} - \mu)\}$$

Möbius time series model assumes that the conditional distribution of θ_t given θ_{t-1} has a von Mises distribution as shown below

$$\Theta_t | (\Theta_{t-1} = \theta_{t-1}) \sim \text{vM}(\mu + 2\text{atan}\{\lambda \tan\frac{1}{2}(\theta_{t-1} - \mu)\}, \kappa), t = 2, \dots, n$$

and the circular time series model becomes

$$\theta_t = \mu + 2\text{atan}\{\lambda \tan\frac{1}{2}(\theta_{t-1} - \mu)\} + \varepsilon_t$$

where $\varepsilon_t \sim \text{vM}(0, \kappa)$. Circular mean direction of conditional distribution of θ_t given θ_{t-1} is given by

$$\mu_t = \mu + 2\text{atan}\{\lambda \tan\frac{1}{2}(\theta_{t-1} - \mu)\}$$

Note that this time series model has the same form introduced by [Fisher and Lee \(1994\)](#), if $g(\cdot) = 2\text{atan}(\cdot)$. Main drawback of the model is that it gives poor fitting for multi-modal circular data sets. Another drawback is the potential problem of identifying μ when λ is close to -1 due to the behaviour of log likelihood function (see, [Hughes \(2007\)](#)). [Kato, \(2010\)](#) provides a new discrete Markov process by adapting Möbius circle transformation as regression curve. This new discrete Markov process has same regression curve with model 5.1 under the assumption of $0 < \lambda < 1$.

In the following section, we provide an extension of Möbius model depending on Bayesian non-parametric approach.

5.3 DP mixture model for circular time series

In this section, we present a general way for non-parametric circular autoregressive modelling using DP mixture. The idea is to provide a non-parametric extension of Möbius time series model. In the following section, we present the model that will be used in this chapter for time-dependent circular data.

5.3.1 DP mixture Möbius model

We adopt a flexible Möbius model on a family of random probability measures using DP defined by [Ferguson \(1973\)](#). Again we return the definition of DP which is almost surely discrete, that is, $G \sim DP(\alpha, G_0)$, where $\alpha > 0$ is a concentration parameter of DP and G_0 is a known baseline distribution. The representation of G as described in [Sethuraman \(1994\)](#) can be defined as

$$G(\cdot) = \sum_{k \geq 1} p_k I_{\varphi_k}(\cdot) \quad (5.2)$$

where $G(\cdot)$ is a random probability measure and I_{φ_k} is a indicator function (or point mass) at φ_k and the weights follow a stick breaking process, $p_k = \prod_{i < k} (1 - q_i) q_k$, with $q_k \sim \text{Beta}(1, \alpha)$ and $\varphi_k \sim G_0$.

We consider the Möbius circular autoregressive order-one $CAR(1)$, dependence case, that is, the conditional distribution $\Theta_t | \Theta_{t-1}, \dots, \Theta_1$ depends only on Θ_{t-1} for $t \geq 2$. This conditional distribution for Θ_t given Θ_{t-1} is a mixture von Mises distribution, but we assume that number of components is unknown. In this situation, the mixing measure G comes from the DP. Our proposed $CAR(1)$ -DP mixture model can be represented as follows

$$\begin{aligned} \Theta_t | \Theta_{t-1} = \theta_{t-1}, K_t = k, (\mu_k, \lambda_k, \kappa_k) &\sim \text{vM}(\mu_k + 2 \text{atan}\{\lambda_k \tan \frac{1}{2}(\theta_{t-1} - \mu_k)\}, \kappa_k) \\ \varphi_k = (\mu_k, \lambda_k, \kappa_k) &\sim G_0, \quad k = 1, 2, \dots, \infty \end{aligned} \quad (5.3)$$

where K_t denotes latent mixture component indicators with probability $P(K_t = k) = p_k$. The representation of [5.3](#) presents a hierarchical definition and the dependence structure is introduced in terms of the latent or state dependent parameters $\varphi_k = (\mu_k, \lambda_k, \kappa_k)$. This model also provides modelling of changing concentration parameter over time.

From the computational viewpoint, a simple format of model [5.3](#) can be achieved by truncating the infinite mixture applied by DP. This is based on selecting sufficiently large number of components, that is, maximum number of components say C . This simple format applies a stick break definition in terms of the mixture weights with $p_k = \prod_{i < k} (1 - q_i) q_k$ for $k = 1, 2, \dots, C$, where each p_k is distributed with a $\text{Beta}(1, \alpha)$ distribution for $1 \leq k \leq C$. ([Ishwaran and James \(2001\)](#)). The model of [5.3](#) can be

rewritten as shown below

$$\begin{aligned}
\Theta_t | \Theta_{t-1} = \theta_{t-1}, K_t, \boldsymbol{\varphi} &\sim \text{vM}(\mu_{K_t} + 2\text{atan}\{\lambda_{K_t} \tan \frac{1}{2}(\theta_{t-1} - \mu_{K_t})\}, \kappa_{K_t}), \quad t = 2, \dots, n \\
K_t | \boldsymbol{p} &\sim \text{Discrete}(p_1, \dots, p_C), \\
\boldsymbol{\varphi}_k = (\mu_k, \lambda_k, \kappa_k) &\sim G_0, \quad k = 1, \dots, C \\
\alpha &\sim \text{Gamma}(v_1, v_2)
\end{aligned} \tag{5.4}$$

$$\tag{5.5}$$

where G_0 can be chosen as $\text{vM}(\mu_0, \kappa_0) \otimes \text{Unif}(0, 1) \otimes \text{Gamma}(a_0, b_0)$ and \boldsymbol{p} is defined by stick breaking process, and $\boldsymbol{\varphi}$ can be decomposed as $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_C)$, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_C)$ and $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \dots, \kappa_C)$. We can implement a blocked Gibbs sampling approach described in [Ishwaran and James \(2002\)](#) under our model specifications for posterior density of $G(\cdot)$. Here, we can moderately draw samples from the following full conditional distributions

$$\begin{aligned}
&(\boldsymbol{\mu} | \boldsymbol{\lambda}, \boldsymbol{\kappa}, \boldsymbol{K}, \theta_n) \\
&(\boldsymbol{\lambda} | \boldsymbol{\mu}, \boldsymbol{\kappa}, \boldsymbol{K}, \theta_n) \\
&(\boldsymbol{\kappa} | \boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{K}, \theta_n) \\
&(\boldsymbol{K} | \boldsymbol{p}, \boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\kappa}, \theta_n) \\
&(\boldsymbol{p} | \boldsymbol{K}, \alpha) \\
&(\alpha | \boldsymbol{p})
\end{aligned} \tag{5.6}$$

This method produces values from posterior distribution $G(\cdot)$ and in each cycle of the sampler, we can track of $(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*, \boldsymbol{\kappa}^*, \boldsymbol{p}^*)$ which are sampled values for $(\boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\kappa}, \boldsymbol{p})$. These values present a random probability measure as follows

$$G^*(\cdot) = \sum_{k=1}^C p_k^* I_{(\mu_k^*, \lambda_k^*, \kappa_k^*)}(\cdot)$$

which is a draw from the posterior distribution $G(\cdot)$. Hence, G^* can be used to directly estimate posterior distribution $G | \theta_n$. To predict for a future observation $\boldsymbol{\varphi}_{n+1} = (\mu_{K_{n+1}}, \lambda_{K_{n+1}}, \kappa_{K_{n+1}})$, we can randomly draw from G^* and can write posterior predictive density $f(\theta_{n+1} | \theta_n)$ for the future observation as follows

$$f(\theta_{n+1} | \theta_n) = \int f_{\text{vM}}(\mu_{K_{n+1}} + 2\text{atan}\{\lambda_{K_{n+1}} \tan \frac{1}{2}(\theta_n - \mu_{K_{n+1}})\}, \kappa_{K_{n+1}}) dG(\boldsymbol{\varphi}_{n+1})$$

Finally, all full conditional distributions have non-standard forms. However all inferences for this model can be performed using slice updater and adaptive Metropolis

block updater in OpenBUGS. Both derivations of the full conditional distributions and OpenBUGS codes can be found in Appendix C for chapter 5.

We would like to emphasize that our model depends on latent state probabilities, but our model can be extended as based on a dependent DP model introduced by MacEachern (2000), that is, dependence is introduced at the level of responses, and not in terms of latent variables φ_k . The form of the dependent Dirichlet process as a collection of random probability measures is defined as $G_\theta = \sum_{k \geq 1} p_k(\theta) I_{\varphi_k(\theta)}$, $\theta \in \Theta$. However, from a computational point of view, DP mixture approach provides greatly simplified computation. Here, we consider DP mixture approach which is a special case of dependent DP approach with common weight.

5.4 Examples

In this section, we present a simulated data example and two real data examples and evaluate the performance of the proposed model. In all of these examples, for DP Möbius model parameters, we use a weakly informative prior as $\mu_0 = 0.1$ and $\kappa_0 = 0.1$ for μ_k and suggest a weakly informative prior for κ_k . For the concentration parameter of DP, we choose $a_o = 2$, $b_0 = 2$ (Ishwaran and James (2001)). For simulation of data based on von Mises distribution, we use the circular package in R. For all computation, we run using 40,000 iterations and dropped the first 20,000 as burn-in iterations with thinning 10, and we use R2OpenBUGS package in R. Additionally, standard diagnostic convergence criteria such as those available in the R2OpenBUGS package is applied to all parameters, indicating that convergence is achieved.

5.4.1 Simulated data example

To evaluate the performance of the proposed DP Möbius model, we considered the following simulation study. The proposed model is applied to estimate the model parameters and the resulting estimates are compared against the true parameters to assess the accuracy of the model.

In this section, we refer simulation studies doing by Artes and Toloï (2010). The best

convergences were found when the autoregressive parameter λ was closer to zero and time series size and concentration parameter were large. The bad convergences were found when the autoregressive parameter λ was larger. Generally, simulation studies have shown the difficulties involved in obtaining good estimates from larger autoregressive cases or from low concentration data or from small sample sizes. To avoid these difficulties in Monte Carlo experiment, we prefer simulated data example.

Here, we simulated two different Möbius time series model for each sample size of $N = 100$ and combined these series. we assumed model parameters as $\boldsymbol{\mu} = (\mu_1, \mu_2) = (-1, 0.5)$, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2) = (0.5, 0.8)$, $\boldsymbol{\kappa} = (\kappa_1, \kappa_2) = (2, 5)$, $\boldsymbol{p} = (p_1, p_2) = (0.5, 0.5)$. Corresponding to rose diagram of the simulated data are displayed in Fig. 5.2 . Clearly, there is one jump and can be seen two modes for simulated Möbius time series data

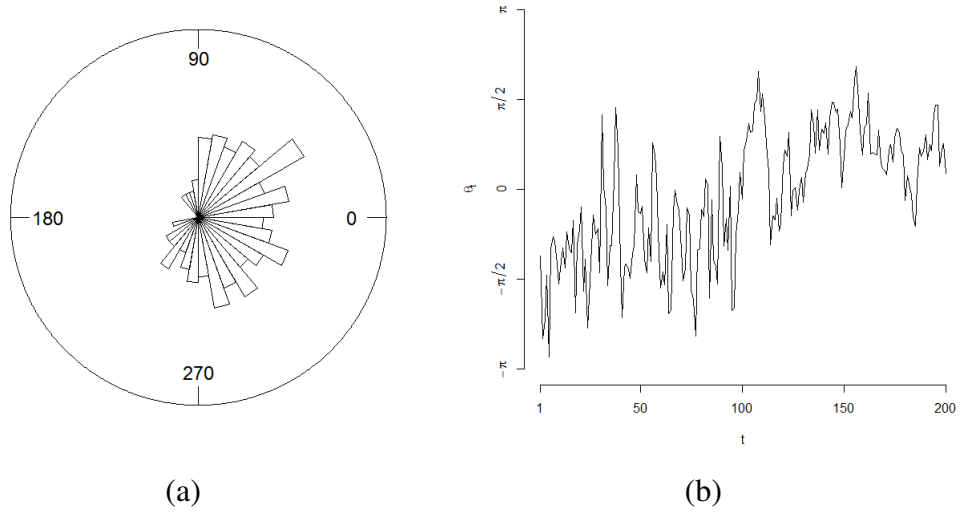


Figure 5.2: (a) Rose diagram of simulated Möbius time series data. (b) Plot of simulated Möbius time series data

In the proposed model, we take as $C = 2$ (the same as the number of components in simulated data set). For κ_k , we use weakly informative prior as $\text{Gamma}(0.01, 0.01)$. To overcome identifying problem on μ 's, we suggest to use a constraint as $\mu_2 = (\pi/2 - \mu_1)\delta_1$, where $\delta_1 \sim \text{Uniform}(0, 1)$. The resulting estimates are summarized as follows

- The posterior circular means $\hat{\boldsymbol{\mu}}$ are $(-1.008, 0.5882)$ and circular standard deviations are 0.32 and 0.36, respectively.

- The posterior means of concentration parameter $\hat{\mathbf{k}}$ are (1.80, 4.77) and standard deviations are (0.37 ,0.90).
- The posterior means of slope parameter $\hat{\boldsymbol{\lambda}}$ are (0.42, 0.82) and standard deviations are (0.16 ,0.09).
- The posterior means of mixing proportions $\hat{\mathbf{p}}$ are (0.33, 0.67) and standard deviations are (0.13, 0.13). The posterior means of concentration parameter $\hat{\alpha}$ of DP is 1.24 and standard deviation is (0.73)

Finally, the posterior estimators of all six parameters are very close to the true values and other mixing proportions are close to true values. Fig. 5.3 a shows the simulated circular time series and solid arrows represent the direction of each simulated value and Fig. 5.3 b shows the the predicted circular time series data and dashed arrows represent the direction of each predicted value. Both figures are very close. Fig. 5.4 shows both the posterior distribution of all parameters for three chains. For three chains, posterior densities are very similar, that is, the convergence of all parameters has achieved.

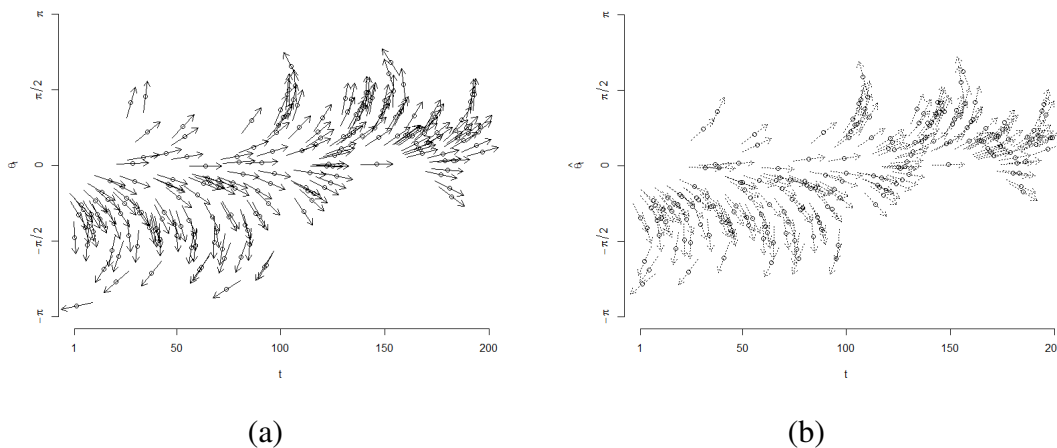


Figure 5.3: (a) Plot of direction of the simulated circular time series data (b) Plot of direction of the predicted circular time series data

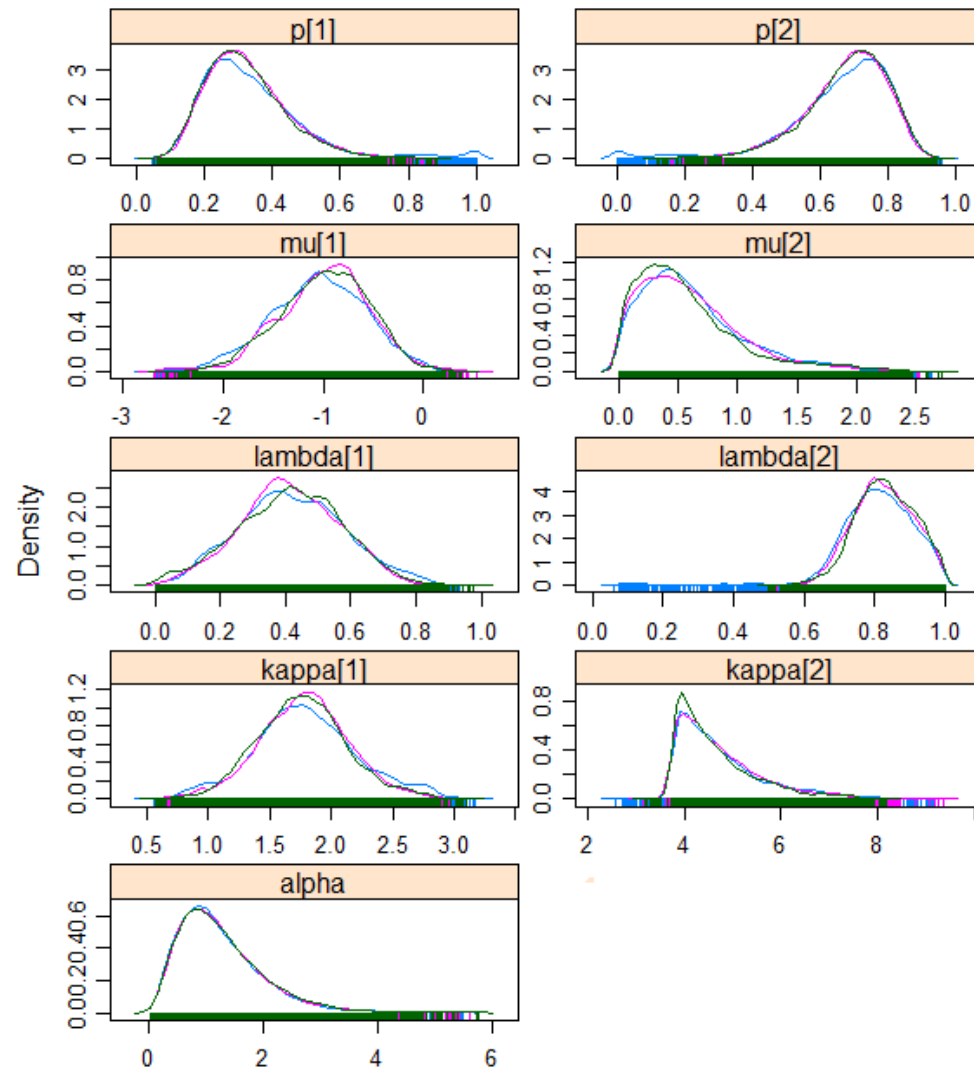


Figure 5.4: Posterior densities of all parameter of DP mixture Möbius model for simulated data

5.4.2 Real data examples

In this section, we consider two real data sets to illustrate our proposed model. At first, we analysed wind direction recorded at a site on Black Mountain, Australian Capital Territory, Australia (Cameron 1983). Secondly, we analysed the wind direction data recorded a hourly in a north-western wind farm in Turkey.

5.4.2.1 Wind directions in Australia

First circular time series example consists of 72 measurements of wind direction recorded at a site on Black Mountain, Australian Capital Territory, Australia (Cameron 1983). Fisher (1994) analysed this data using CAR(1) described in Sec. 5.2.2. Recently, the data is re-analysed for outlier detection in CAR(1) model by Abuzaid, et al. (2014). They concluded observations 5, 12 and 31 are Innovational outlier (IO) candidates, while observations 14,39, 50 are Additive outliers (AO) based on their three graphical procedures.

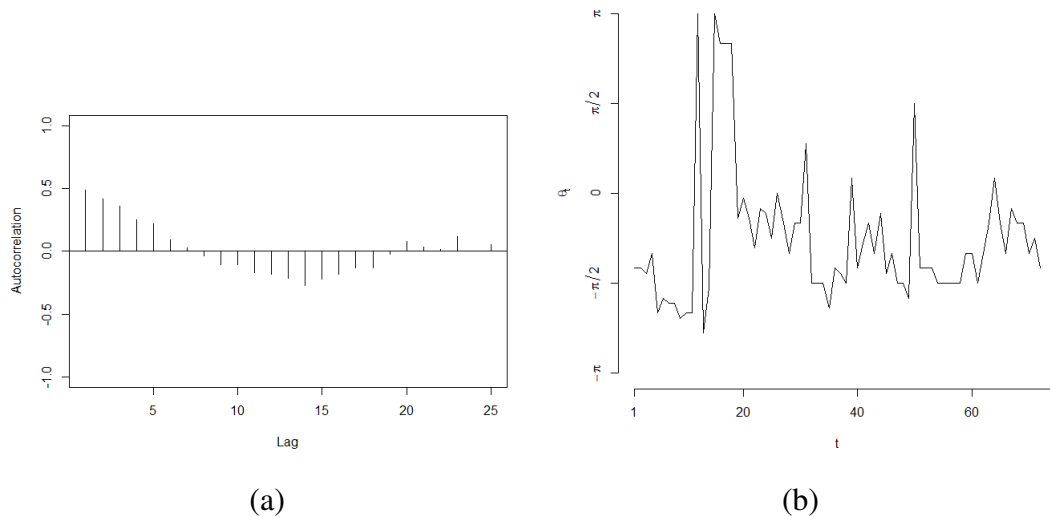


Figure 5.5: (a) Sample circular autocorrelations for the time series of wind directions in Australia (b) Plot of observed circular time series data in Australia

We also re-analyse this data using our approach. To overcome identifying problem, we use $\mu_2 = \mu_1 + \delta_1$ where $\delta_1 \sim \text{vM}(0, 0.001)I(\pi)$. We then set $\lambda = \lambda_1 = \lambda_2$. We take hyper-parameter as $a_0 = b_0 = 1$ for concentration parameter κ_k . The resulting

estimates of our proposed model are summarized as follows

- The posterior circular means are $\hat{\boldsymbol{\mu}} = (293.8^\circ, 22.6^\circ)$ and circular standard errors are 0.064 and 0.90 radians.
- The posterior means of concentration parameters are $\hat{\boldsymbol{\kappa}} = (4.67, 0.42)$ and the standard errors of them are 0.11 and 0.09
- The posterior means of mixing weights are $\hat{\boldsymbol{p}} = (0.80, 0.20)$ and the estimate of λ^* is 0.78 and standard error is 0.012. The estimate of concentration parameter of DP Möbius model is $\hat{\alpha} = 0.83$ and its standard error is 0.06. Posterior densities of all parameters of the proposed model are displayed in Fig. 5.8 for three chains.

We compare our model with models of [Fisher and Lee \(1994\)](#); [Abuzaid et al.\(2014\)](#) as follows

The estimates of [Fisher and Lee \(1994\)](#), with their standard errors given in parenthesis, are summarized as follows

$$\hat{\boldsymbol{\mu}} = 289.5^\circ(0.086), \hat{\boldsymbol{\kappa}} = 2.5(0.352), \hat{\lambda} = 0.68(0.138)$$

The estimates of [Abuzaid et al.\(2014\)](#) are below after adjustment of five outliers and refitting $CAR(1)$ model to reduced data

$$\hat{\boldsymbol{\mu}} = 210.19^\circ(0.12), \hat{\boldsymbol{\kappa}} = 2.27(0.10), \hat{\lambda} = 0.87(0.047)$$

According to these estimates, we obtain less standard error compare to other models. Fig .5.6 displays the predicted and actual rose diagrams of wind directions. These rose diagrams are similar.

Additionally, our second class observations are displayed as 12, 15, 19, 31, 32, 39, 40, 50 and 51 in Fig. 5.9. This finding is particularly remarkable in the sense that our method provided a formal way that was able to pin down the observations in this dataset having a distribution different than the bulk. Also, the method was able to identify the characteristics of the distribution to which the *outliers* belong. Accordingly, 80% of the wind direction data constitutes one cluster ($vM(293.8, 4.67)$) whereas 20% come from another distribution ($vM(22.6, 0.42)$).

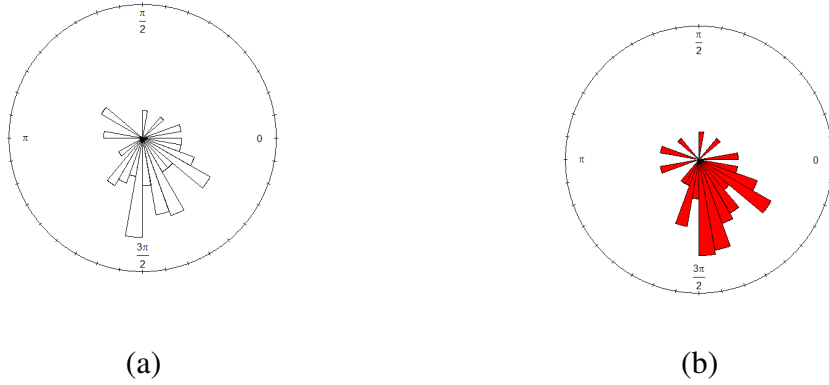


Figure 5.6: (a) Rose diagram of wind direction data. (b) Rose diagram of predicted wind direction data.

5.4.2.2 Wind directions in Turkey

As second example, we consider a time series of wind directions measured hourly in a north western wind farm in Turkey. Here, we analyse a time series of 120 wind directions measured hourly between 1.am. on May 1st and 11.p.m. on May 5th. Sample circular autocorrelation coefficients proposed by Fisher and Lee (1994) are displayed in Fig. 5.7.a , which shows that there is a evidence that a $CAR(1)$ model may be appropriate. We fit our proposed DP mixture Möbius model. To overcome

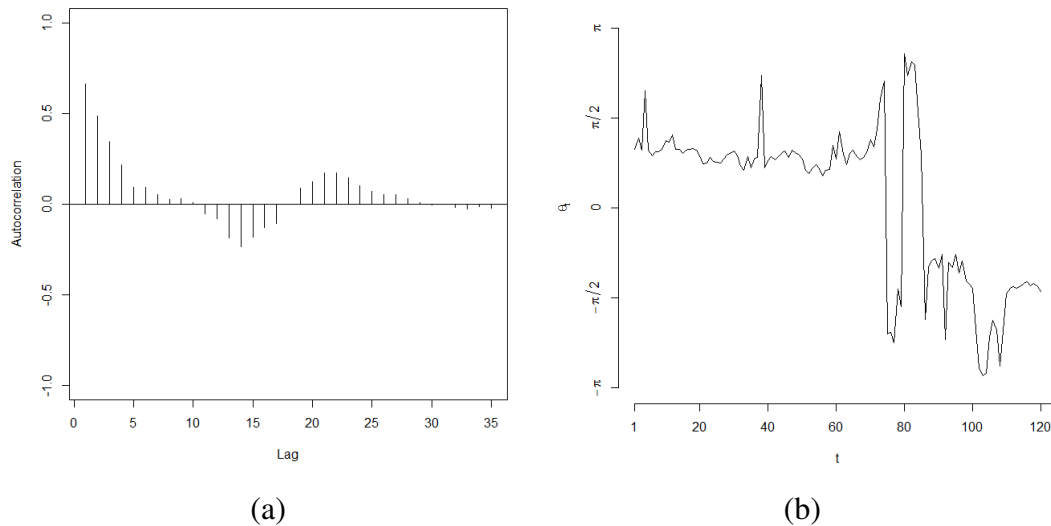


Figure 5.7: (a) Sample circular autocorrelations for the time series of wind directions in Turkey (b) Plot of observed circular time series data in Turkey

identifying problem for $C = 2$, we use $\mu_2 = (\pi/2 - \mu_1)\delta_1$, where $\delta_1 \sim \text{Unif}(0, 1)$. We take hyper-parameter as $a_0 = b_0 = 0.1$ for concentration parameters κ_k .

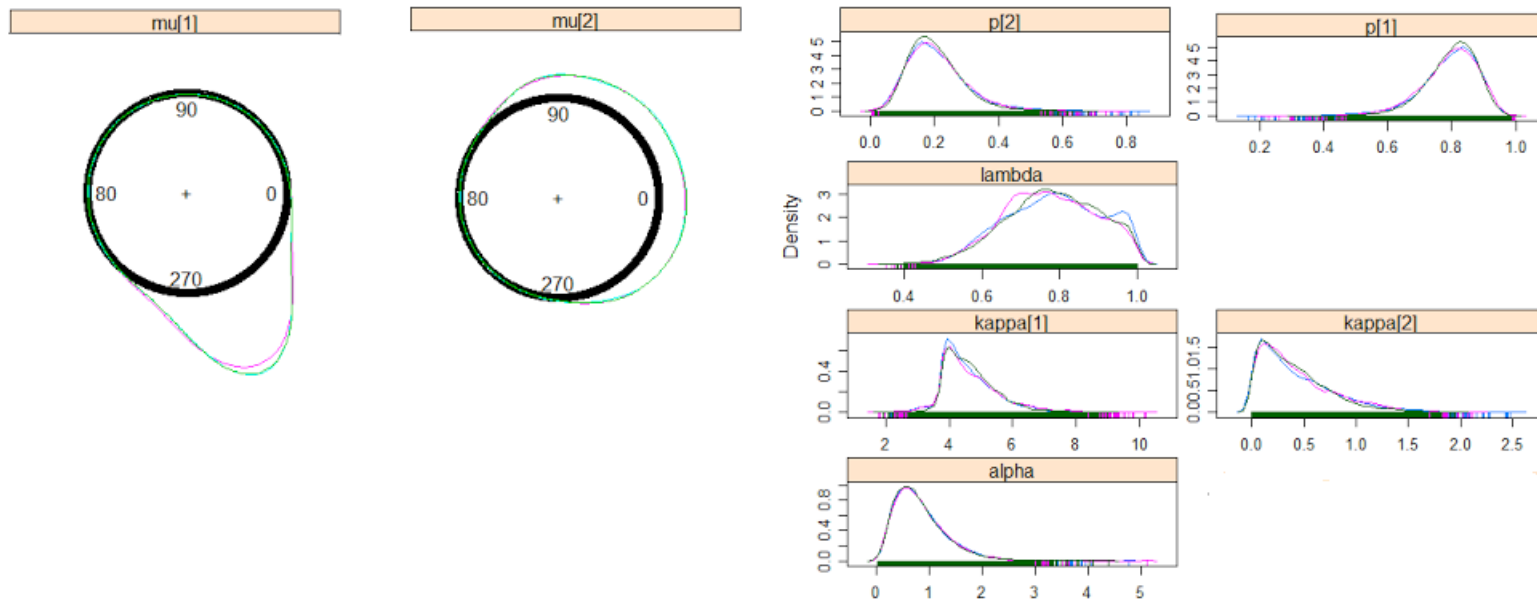


Figure 5.8: Posterior densities of parameters of DP Möbius model for wind direction data from Australia.

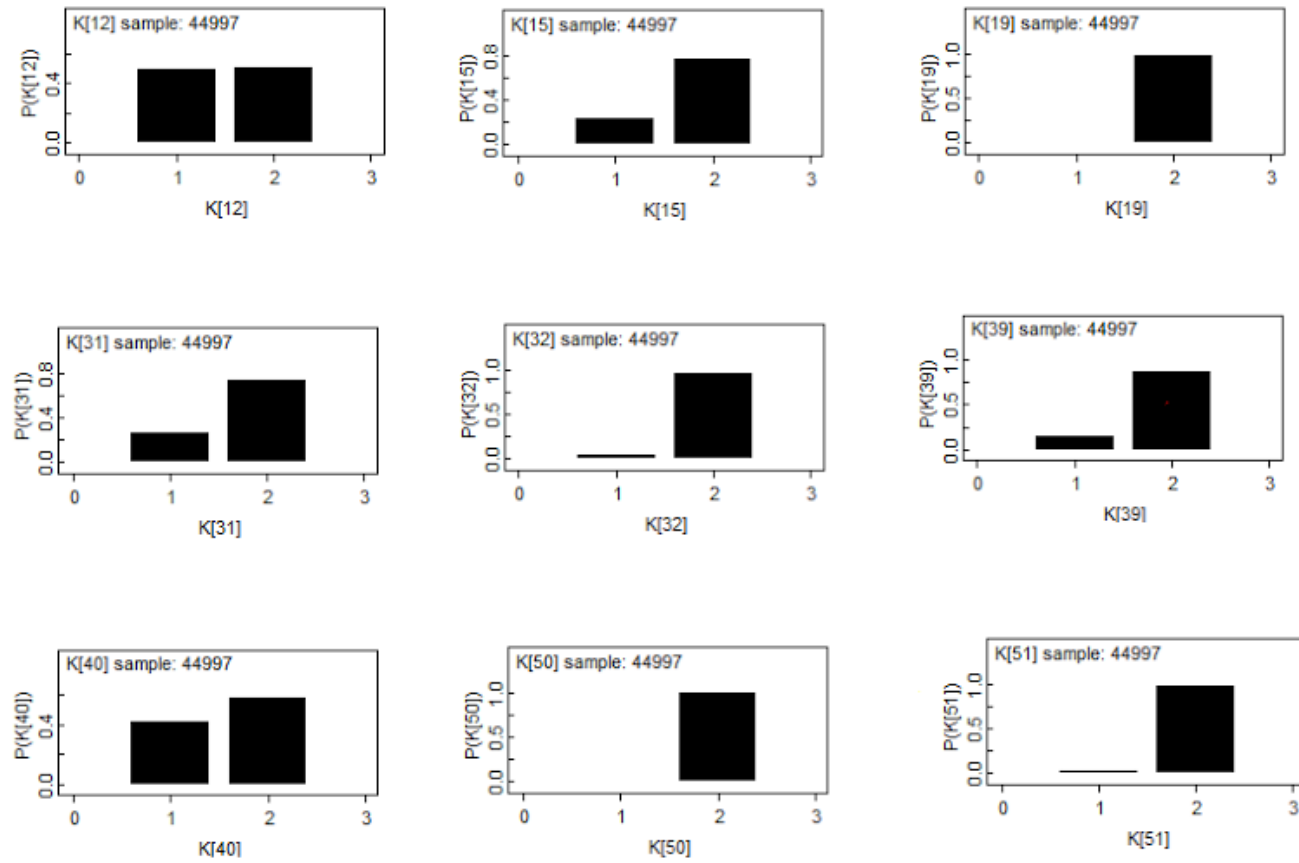


Figure 5.9: Posterior density of K latent variables which belong to second cluster

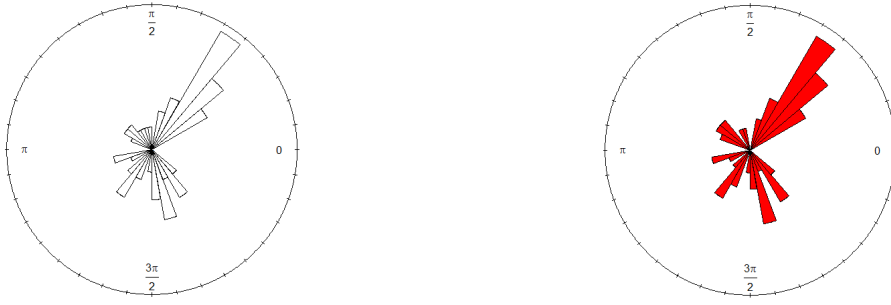
The resulting estimates of our Bayesian methodology for $C = 2$ are summarized as follows

- The posterior circular means are $\hat{\boldsymbol{\mu}} = (260.30^\circ, 79.06^\circ)$ and circular standard deviations are 0.94 and 0.96 radians.
- The posterior means of concentration parameters are $\hat{\boldsymbol{\kappa}} = (1.51, 50.18)$ and the standard deviations of them are 0.46 and 10.76
- The posterior means of mixing weights are $\hat{\boldsymbol{p}} = (0.27, 0.73)$ and the estimate of $\hat{\boldsymbol{\lambda}}$ is $(0.83, 0.97)$ and standard deviations are $(0.14, 0.02)$. The estimate of concentration parameter of DP Möbius model is $\hat{\alpha} = 1.16$ and its standard deviation is 0.73.

For $C = 3$, we use data-based identifying prior-constraint on $\boldsymbol{\mu}$'s as $\mu_1 \sim \text{vM}(0.1, 0.1)I(, 0)$, $\mu_2 \sim \text{vM}(0.1, 0.1)I(0, \pi/3)$ and $\mu_3 \sim \text{vM}(0.1, 0.1)I(\pi/3, \pi)$. The resulting estimates of our Bayesian methodology for $C = 3$ are summarized as follows

- The posterior circular means are $\hat{\boldsymbol{\mu}} = (240.37^\circ, 29.09^\circ, 110.45^\circ)$, and circular standard deviations are 0.88, 0.29 and 0.61 radians.
- The posterior means of concentration parameters are $\hat{\boldsymbol{\kappa}} = (1.47, 1.72, 48.97)$ and the standard deviations of them are 1.67, 4.80 and 11.16
- The posterior means of mixing weights are $\hat{\boldsymbol{p}} = (0.16, 0.12, 0.70)$ and standard deviations $(0.09, 0.08, 0.07)$ the estimate of $\hat{\boldsymbol{\lambda}}$ are $(0.44, 0.33, 0.96)$ and standard deviations are $(0.26, 0.29, 0.02)$ respectively. The estimate of concentration parameter of DP Möbius model is $\hat{\alpha} = 1.56$ and its standard deviation is 0.85.

Fig. 5.10 shows a rose diagram of the observed wind directions, white color, and compared with the posterior predicted wind directions, red color. These rose diagrams are very close. Additionally, Fig. 5.11 a shows the observed circular time series and solid arrows represent the direction of each observation value and Fig. 5.11 b shows the the predicted circular time series data and dashed arrows represent the direction



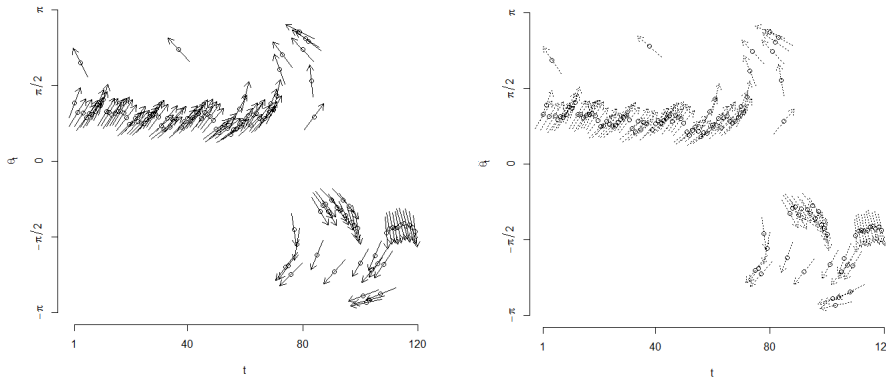
(a)

(b)

Figure 5.10: (a) Rose diagram of wind direction data in Turkey (b) Rose diagram of predicted wind direction data in Turkey

of each predicted value in Turkey. Both figures are very close. Additionally, these figures are constructed using teaching demos package in R.

To compare $C = 2$ and $C = 3$ our DP Möbius models, we use two metrics, namely, mean absolute cosine error (MACE) and mean cosine difference error (MCDE). We compute MACE and MCDE as $MACE = \frac{1}{n} \sum_{i=1}^n |\cos(\theta_i) - \cos(\hat{\theta}_i)|$ and $MCDE = 1 - \frac{1}{n} \sum_{i=1}^n \cos(\theta_i - \hat{\theta}_i)$ where $\hat{\theta}_i$ is posterior mean of direction.



(a)

(b)

Figure 5.11: (a) Plot of direction of the observed circular time series in Turkey (b) Plot of direction of the predicted circular time series in Turkey

Table 5.1 shows the computed MACE and MCDE values for two DP Möbius models. These values are close to zero. For $C = 3$, the computed MACE and MCDE are 0.20 and 0.09 which show a slight improvement over $C = 2$.

Table 5.1: Comparison with model selection criterion

Model	MACE	MCDE
DP Möbius(C=2)	0.21	0.10
DP Möbius(C=3)	0.20	0.09

5.5 Discussion

In many real data application, likelihood of a Möbius time series model has multiple local maxima and so convergence of the maximization algorithms provide no guarantee that the global maximum can be found when observing multi-modality. In this chapter, we have provided DP mixture Möbius model in analysing multi-modal circular data. Potential identifying problem on μ is handled with a prior constraint in Bayesian aspect. Identifying problem tends to elevate as C increases. Here, the order constraint on μ 's may not be proper since these parameters are circular in nature. The useful solution is to divide circular support as based on empirical rose diagrams when $C > 2$. The results of both simulated and real data examples indicate that our proposed model has been shown to perform well in terms of estimates of parameter and prediction error for multi-modal circular data. Finally, DP Möbius mixture models allow for great robustness when there are jumps in series or multi-modality in the time series of interest.

CHAPTER 6

CONCLUSION

In this dissertation, we investigate a number of problems which related with multi-modal circular data analysis in Bayesian panorama. We propose new models in the solution of these problems and contribute to the existing methods in the literature. Our contributions can be summarized as follows :

- For asymmetric and bi-modal circular distributions based on the extension of von Mises distribution, namely asymmetric generalised von mises (AGvM) and generalised von Mises (GvM), we proposed a general solution depending on SIR method for analysing asymmetric and bi-modal circular data. Main problem for analysing these distributions is the complex normalizing constants which are not available in closed forms. Our approach overcomes this problem. Additionally, in this dissertation, we define conjugate prior distributions for these problematic distributions and use the proposal distributions to obtain sample from the posterior distributions using SIR method. Finally, we propose the bivariate constrained joint prior distributions for the depended parameters of AGvM and GvM distributions in analysing asymmetric and bi-modal circular data.
- In many environmental and biological applications, circular data usually display multi modality and there is usually uncertainty about the number of modes as it is hard to determine from the sampled data. To overcome this uncertainty, we successfully adopted Dirichlet process (DP) mixture model to von Mises and wrapped Cauchy distributions. Our simulated and real data examples show the flexibility , utility and efficiency of the proposed approach in terms of the

parameter and unknown modal number estimation.

- DP mixture Möbius circular time series model is proposed for time-dependent circular data sets. Our aim is to model multi-modal circular time series observations with time structure. Our proposed Bayesian methodology depending on DP mixture model overcomes both varying concentration parameter κ over time and multi-modality problem in circular time series context. Additionally, Bayesian framework allows us to solve identifying problem on μ and computational issues.

REFERENCES

- [1] Abe, T., Pewsey, A. (2011). Sine-skewed circular distributions. *Stat Pap* 52, 683 - 707
- [2] Abuzaid, A.H., Mohamed, I.B., Hussin, A.G. (2014). Procedures for outlier detection in circular time series models, *Environmental Ecological Statistics*, 21, 793-809.
- [3] Abramowitz, M. and Stegun, A. (1972). Handbook of mathematical functions, graphs, and mathematical tables, New York: Dover Publications, ISBN 978-0-486-61272-0
- [4] Antonia, G.N. and Peña, E.G. (2005). A Bayesian analysis of directional data using von Mises Fisher distribution. *Communication in Statistics-Simulation and Computation*, 34, 989-999.
- [5] Antonia, G.N., Ausín and Wiper, M.P. (2014). Bayesian nonparametric models of circular variables based on Dirichlet process mixtures of normal distributions, *Journal of Agricultural, Biological and Environmental Statistics*, 20,1, 47-64.
- [6] Antoniak, C. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Annals of Statistics*, 2, 1152-1174.
- [7] Arnold, B.C., Ng, H.K.T. (2011). Flexible bivariate distributions, *Journal of Multivariate Analysis* 102, 1194-1202.
- [8] Arnold, B.C., Castillo, E., Sarabia, J.M. (1999). Conditional specification of statistical models, Springer-Verlag, New York.
- [9] Artes, R. and Tolo, M.C.C. (2010). An autoregressive model for time series of circular data, *Communications in Statistics-Theory and Methods*, 39, 186-194.
- [10] Bagchi, P, Kadane, J. B. (1991). Laplace approximations to posterior moments and marginal distributions on circle, spheres, and cylinders. *Canad. J. Statist.* 19, 67 - 77.
- [11] Blackwell, D. and MacQueen, J. B. (1973). Ferguson distributions via Polya urn schemes, *Annals of Statistics*, 1, 353–355.
- [12] Bhattacharya, S. SenGupta, A. (2009). Bayesian inference for circular distributions with unknown normalizing constant. *Journal of Statistical Planning and Inference*, 139,12, 4179-4192.

- [13] Bhattacharya, S. and SenGupta, A. (2009). Bayesian analysis of semiparametric linear-circular models. *Journal of Agricultural, Biological and Environmental Statistics*, 14, 33–65.
- [14] Breckling, J. (1989). Analysis of directional time series: Applications to wind speed and direction, *Lecture Notes in Statistics 61*, Springer-Verlag, Berlin.
- [15] Cameron, M.A (1983). The comparison of time series recorders. *Technometrics* 25, 9–22
- [16] Chang-Chien, S.J., Hung, W.L, and Yang, M.S. (2012). On mean shift based clustering for circular data, *Soft Comput.*, 16 , 6 , 1043-1060.
- [17] Coles, S. (1998). Inference for circular distributions and processes. *Statist. Comput.* 8 105-113.
- [18] Congdon, P. (2001). Bayesian statistical modelling. Wiley: Chichester.
- [19] Cox, D. R.(1975). Contribution to discussion of Mardia, *J. Roy. Statist. Soc. Ser. B*, 37, 380-381 (45,143, 273)
- [20] Damien P, Walker, S. G.(1999) A full Bayesian analysis of circular data using the von Mises distribution. *Canad. J. Statist.* 27, 291-298.
- [21] Durán, J.J.F (2004) Circular distributions based on nonnegative trigonometric sums. *Biometrics* 60, 499 - 503.
- [22] Durán, J.J.F. and Domínguez, M.M.G. (2014). Modeling angles in proteins and circular genomes using multivariate angular distributions based on multiple non-negative trigonometric sums, *Stat Appl Genet Mol Biol*, 13(1), 1-18.
- [23] Downs, T. D. and Mardia, K. V. (2002). Circular regression. *Biometrika* 89 683-697.
- [24] Ferguson, T. (1973) A Bayesian analysis of some non-parametric problems. *The Annals of Statistics*, 1, 209-230.
- [25] Fisher, N. I. (1993). Statistical analysis of circular data. Cambridge University Press.
- [26] Fisher, N. I. and Lee, A.J. (1994). Time series analysis of circular data. *J. R. Statist. Soc. B*, 56, 327-339.
- [27] Gatto, R., Jammalamadaka, S. (2007). The generalized von Mises distribution. *Stat. Methodology* 4, 341 - 353.
- [28] George, J.B. and Ghosh, K. (2006). A semiparametric Bayesian model for circular-linear regression, *Communications in Statistics - Simulation and Computation*, 35,4, 911-923.

- [29] Ghosh, K., Jammalamadaka S.R, Tiwari, R. (2003). Semiparametric Bayesian techniques for problems in circular data. *Journal of Applied Statistics*, 30, 145–161.
- [30] Guttorp, P, Lockhart, R. A. (1988) Finding the location of a signal: a Bayesian analysis. *JASA*. 83, 322-329.
- [31] Hall, P., Watson, G. S., and Cabrera, J. (1987). Kernel density estimation with spherical data. *Biometrika*, 74, 751-762.
- [32] Hughes, G. (2007). Multivariate and time series models for circular data with applications to protein conformational angles. Ph.D. thesis, Univ. Leeds, Leeds, England, UK.
- [33] Hung, W.L., Chang-Chien, S.J and Yang, M.S. (2009). Self updating clustering algorithm for estimating parameters in mixtures of von Mises distributions, *Journal of Applied Statistics*, 39 , 10, 2259-2274.
- [34] Holzmann, H., Munk, A., Suster, M., Zucchini, W. (2006). Hidden Markov models for circular and linear-circular time series, *Environmental and Ecological Statistics*, 13, 3, 325-347.
- [35] Ishwaran, H. and Zarepour, M. (2000). Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika*, 87, 371-339.
- [36] Ishwaran, H. and James, L. (2001). Gibbs sampling methods for stick-breaking priors, *JASA*, 96, 161-173.
- [37] Ishwaran H, James L.F. (2002). Approximate Dirichlet process computing infinite normal mixtures: smoothing and prior information. *Journal of Computational and Graphical Statistics* , 11, 508 –532.
- [38] Jammalamadaka. S.R and Sengupta. A. (2001). Topics in circular statistics, World Scientific Press.
- [39] Jander, R. (1957). Die optische richtangsorientierung der roten waldameise. *Z.vergl. Physiologie*, 40 162-238.
- [40] Jeffreys.H (1961), Theory of probability, Oxford, UK Oxford University Press.
- [41] Kato, S. (2010). A Markov process circular data, *J. R. Statist. Soc. B*,72, 665-672.
- [42] Kent, J. T. Tyler, D. E. (1988). Maximum likelihood estimation for the wrapped Cauchy distribution. *Journal of Applied Statistics* 15, 247-254.
- [43] Kim, S., SenGupta, A.(2013). A three parameter generalized von Mises distribution. *Stat Pap* 54, 685-693.

- [44] Lasino, G. J. , Gelfand, A. and Lasinio, M.J.(2012). Spatial analysis of wave direction data using wrapped gaussian process, *The Annals of Applied Statistics*, Vol 6, No 4, 1478-1498.
- [45] Lund, U. (1999). Least circular distance regression for directional data. *Journal of Applied Statistics* 26, 723-733.
- [46] Lunn, D., Jackson, C. , Best, N. , Thomas, A. and Spiegelhalter, D. (2012). The BUGS Book: A practical introduction to Bayesian analysis, Chapman & Hall/CRC Texts in Statistical Science Paperback
- [47] MacEachern, S. N. (2000). Dependent dirichlet processes, Department of Statistics, The Ohio State University.
- [48] Maksimov, V. (1967) Necessary and sufficient statistics for the family of shifts of probability distributions on continuous bicomact groups. *Theoria Verоятna* , 307 - 321.
- [49] Mardia, K.V. (2010). Bayesian analysis for bivariate von Mises distribution, *Journal of Applied Statistics*, 37.3, 515-528.
- [50] Mardia, K.V. (1975). Statistics of directional data, *Journal of the Royal Statistical Society Series B Methodological*. 37, 349-393.
- [51] Mardia, K. V. (1972). Statistics of directional data, Academic Press, London.
- [52] Mardia, K.V. and Jupp, P.E. (1999). Directional statistics. John Wiley, Chichester.
- [53] Mardia, K.V. and Sutton, T.W.(1975). On the modes of a mixture of two von Mises distributions, *Biometrika*, 62, 699-701.
- [54] Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models, *Journal of Computational and Graphical Statistics*, 9, 249–265.
- [55] Oliveria, M., Crujeiras, R. M., and Rodríguez-Casal, A. (2012). A plug-in rule for bandwidth selection in circular density estimation. *Computational Statistics and Data analysis*, 56, 3898-3908.
- [56] Olkin, I. and Liu, R.,(2003). A bivariate beta distribution, *Statistics and Probability Letters* 62 , 407-412.
- [57] Pewsey, A. (2002). Testing circular symmetry. *Canadian Journal of Statistics*, 30, 591-600.
- [58] Pewsey, A., Neuhäuser, M. and Ruxton, G.D. (2013). Circular statistics in R, Oxford Press.
- [59] Raftery, A.E. (1995). Bayesian model selection in social research (with Discussion) *Sociological Methodology* 25, 111-196.

- [60] Ravindran, P (2002). Bayesian analysis of circular data using wrapped distributions, Phd.thesis.
- [61] R Development Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>.
- [62] Rubin D. B. (1988) Using the SIR algorithm to simulate posterior distributions with discussion, *Bayesian Statistics 3*, *Oxford: University Press*, 395-402.
- [63] Schafer JL, Graham JW.(2002). Missing data: our view of the state of the art. *Psychological Methods*, 7, 147–177
- [64] Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, 4, 639-650.
- [65] SenGupta A., Laha, A.K. (2008) A Bayesian analysis of change point problem for directional data. *Journal of Applied Statistics*, 35.6, 693-700.
- [66] SenGupta A., Kim, S. and Arnold, B.C. (2013). Inverse circular-circular regression, *Journal of Multivariate Analysis*.119, 200-208.
- [67] Smith, A.F.M and Gelfand, A.E. (1992) Bayesian statistics without tears: a sampling-resampling perspective. *The American Statistician*, 46, 84-88.
- [68] Spurr, B. D. (1981). On estimating the parameters in mixtures of circular normal distributions, *Mathematical Geology*, 13, 163-173.
- [69] Stephens, M.A.(1969). Techniques for directional data, Technical Report, Stanford University.
- [70] Taylor, C.C. (2008). Automatic bandwidth selection for circular density estimation. *Computational Statistics and Data Analysis*, 52, 3493-500.
- [71] Wang, F. and Gelfand, A. E.(2013.) Directional data analysis under the general projected normal distribution. *Statistical Methodology*, 10, 113-127.
- [72] Wilks, S.S.(1963). *Mathematical Statistics*, 2nd edition, John Wiley and Sons, New York.
- [73] Yfantis, E.A. and Borgman, L.E.(1982). An extension of the von Mises distribution, *Communications in Statistics, Theory and Methods* 11, 1695–1706.

APPENDIX A

APPENDIX FOR CHAPTER 3

In this section, we show posterior distribution-conjugacy for each sub-model and give technical and computational details for these prior joint distributions of dependent parameters for each sub-model described in Section 3.2

A.1 Posterior distribution-Conjugacy for Generalised von Mises distribution

The posterior density of GvM is proportionally to $L(\theta|\mu_1, \mu_2, \kappa_1, \kappa_2) \times p(\mu_1, \mu_2, \kappa_1, \kappa_2)$ which is given by

$$\begin{aligned} & \{c(\kappa_1, \kappa_2)\}^{-n} \exp\left(\sum_i \kappa_1 \cos(\theta_i - \mu_1) + \sum_i \kappa_2 \cos 2(\theta_i - \mu_2)\right) \times \\ & \{c(\kappa_1, \kappa_2)\}^{-r} \exp(\kappa_1 R_{01} \cos(\mu_1 - \mu_{01}) + \kappa_2 R_{02} \cos 2(\mu_2 - \mu_{02})) \\ = & \{c(\kappa_1, \kappa_2)\}^{-(n+r)} \exp\left\{\sum_i \kappa_1 \cos(\theta_i - \mu_1) + \kappa_1 R_{01} \cos(\mu_1 - \mu_{01})\right\} \\ & \times \exp\left\{\sum_i \kappa_2 \cos 2(\theta_i - \mu_2) + \kappa_2 R_{02} \cos 2(\mu_2 - \mu_{02})\right\} \end{aligned}$$

Here, the first exponential form can be separately expanded as

$$\begin{aligned} \sum_i \kappa_1 \cos(\theta_i - \mu_1) + \kappa_1 R_{01} \cos(\mu_1 - \mu_{01}) &= \kappa_1 \cos \mu_1 \sum_i \cos \theta_i + \kappa_1 \sin \mu_1 \sum_i \sin \theta_i \\ & \quad + \kappa_1 R_{01} \cos \mu_1 \cos \mu_{01} + \kappa_1 R_{01} \sin \mu_1 \sin \mu_{01} \\ &= \kappa_1 \cos \mu_1 \{R_{01} \cos \mu_{01} + \sum_i \cos \theta_i\} + \kappa_1 \sin \mu_1 \{R_{01} \sin \mu_{01} + \sum_i \sin \theta_i\} \\ &= \kappa_1 \cos \mu_1 \{R_{n1} \cos \mu_{n1}\} + \kappa_1 \sin \mu_1 \{R_{n1} \sin \mu_{n1}\} = \kappa_1 R_{n1} \cos(\mu_1 - \mu_{n1}) \end{aligned}$$

The second exponential form is given by

$$\begin{aligned}
\sum_i \kappa_2 \cos 2(\theta_i - \mu_2) + \kappa_2 R_{02} \cos 2(\mu_2 - \mu_{02}) &= \kappa_2 \cos 2\mu_2 \sum_i \cos 2\theta_i + \kappa_2 \sin 2\mu_2 \sum_i \sin 2\theta_i \\
&\quad + \kappa_2 R_{02} \cos 2\mu_2 \cos 2\mu_{02} + \kappa_2 R_{02} \sin 2\mu_2 \sin 2\mu_{02} \\
&= \kappa_2 \cos 2\mu_2 \{R_{02} \cos 2\mu_{02} + \sum_i \cos 2(\theta_i)\} + \kappa_2 \sin 2\mu_2 \{R_{02} \sin 2\mu_{02} + \sum_i \sin 2(\theta_i)\} \\
&= \kappa_2 \cos 2\mu_2 \{R_{n2} \cos 2\mu_{n2}\} + \kappa_2 \sin 2\mu_2 \{R_{n2} \sin 2\mu_{n2}\} = \kappa_2 R_{n2} \cos 2(\mu_2 - \mu_{n2})
\end{aligned}$$

A.1.1 Posterior distribution

Here, the posterior distribution of GvM is given by

$$\{c(\delta, \kappa_1, \kappa_2)\}^{-m} \exp(\kappa_1 R_{n1} \cos(\mu_1 - \mu_{n1}) + \kappa_2 R_{n2} \cos 2(\mu_2 - \mu_{n2}))$$

where $m = r + n$ and R_{n1}, R_{n2} and μ_{n1}, μ_{n2} are obtained from the following equations

$$\begin{aligned}
R_{n1} \cos \mu_{n1} &= R_{01} \cos \mu_{01} + \sum_i \cos \theta_i, \quad R_{n1} \sin \mu_{n1} = R_{01} \sin \mu_{01} + \sum_i \sin \theta_i \\
R_{n2} \cos 2\mu_{n2} &= R_{02} \cos 2\mu_{02} + \sum_i \cos 2\theta_i, \quad R_{n2} \sin 2\mu_{n2} = R_{02} \sin 2\mu_{02} + \sum_i \sin 2\theta_i
\end{aligned} \tag{A.1}$$

A.2 Posterior distribution-Conjugacy for Asymmetric Generalised von Mises distribution

The posterior density of AGvM is proportionally to $L(\theta|\mu, \kappa_1, \kappa_2) \times p(\mu, \kappa_1, \kappa_2)$.

$$\begin{aligned}
&\{c(\kappa_1, \kappa_2)\}^{-n} \exp\left(\sum_i \kappa_1 \cos(\theta_i - \mu) + \sum_i \kappa_2 \sin 2(\theta_i - \mu)\right) \\
&\times \{c(\kappa_1, \kappa_2)\}^{-r} \exp(\kappa_1 R_{01} \cos(\mu - \mu_0) + \kappa_2 R_{02} \sin 2(\mu - \mu_0)) \\
&= \{c(\kappa_1, \kappa_2)\}^{-(n+r)} \exp\left\{\sum_i \kappa_1 \cos(\theta_i - \mu) + \kappa_1 R_{01} \cos(\mu - \mu_0)\right\} \\
&\quad \times \exp\left\{\sum_i \kappa_2 \sin 2(\theta_i - \mu) + \kappa_2 R_{02} \sin 2(\mu - \mu_0)\right\}
\end{aligned} \tag{A.2}$$

Here, we can be expanded separately the summation of forms as follows

$$\begin{aligned}
\sum_i \kappa_1 \cos(\theta_i - \mu) + \kappa_1 R_{01} \cos(\mu - \mu_0) &= \kappa_1 \cos \mu \sum_i \cos \theta_i + \kappa_1 \sin \mu \sum_i \sin \theta_i \\
&\quad + \kappa_1 R_{01} \cos \mu \cos \mu_0 + \kappa_1 R_{01} \sin \mu \sin \mu_0 \\
&= \kappa_1 \cos \mu \{R_{01} \cos \mu_0 + \sum_i \cos(\theta_i)\} + \kappa_1 \sin \mu \{R_{01} \sin \mu_0 + \sum_i \sin(\theta_i)\} \\
&= \kappa_1 \cos \mu \{R_{n1} \cos \mu_n\} + \kappa_1 \sin \mu \{R_{n1} \sin \mu_n\} = \kappa_1 R_{n1} \cos(\mu - \mu_n)
\end{aligned}$$

The second summation is expanded as:

$$\begin{aligned}
\sum_i \kappa_2 \sin 2(\theta_i - \mu) + \kappa_2 R_{02} \sin 2(\mu - \mu_0) &= \kappa_2 \cos 2\mu \sum_i \sin 2\theta_i - \kappa_2 \sin 2\mu \sum_i \cos 2\theta_i \\
&\quad + \kappa_2 R_{02} \sin 2\mu \cos 2\mu_0 - \kappa_2 R_{02} \cos 2\mu \sin 2\mu_0 \\
&= -\kappa_2 \cos 2\mu \{R_{02} \sin 2\mu_0 - \sum_i \sin 2\theta_i\} + \kappa_2 \sin 2\mu \{R_{02} \cos 2\mu_0 - \sum_i \cos 2\theta_i\} \\
&= -\kappa_2 \cos 2\mu \{R_{n2} \sin 2\mu_n\} + \kappa_2 \sin 2\mu \{R_{n2} \cos 2\mu_n\} = \kappa_2 R_{n2} \sin 2(\mu - \mu_n).
\end{aligned}$$

A.2.1 Posterior distribution

Here, the posterior distribution of AGvM is given by

$$\{c(\delta, \kappa_1, \kappa_2)\}^{-m} \exp(\kappa_1 R_{n1} \cos(\mu - \mu_n) + \kappa_2 R_{n2} \sin 2(\mu - \mu_n))$$

where $m = r + n$ and R_{n1}, R_{n2}, μ_n obtained from the following equation

$$\begin{aligned}
R_{n1} \cos \mu_n &= R_{01} \cos \mu_0 + \sum_i \cos \theta_i, & R_{n1} \sin \mu_n &= R_{01} \sin \mu_0 + \sum_i \sin \theta_i \\
R_{n2} \cos 2\mu_n &= R_{02} \cos 2\mu_0 - \sum_i \cos 2\theta_i, & R_{n2} \sin 2\mu_n &= R_{02} \sin 2\mu_0 - \sum_i \sin 2\theta_i
\end{aligned} \tag{A.3}$$

A.3 Constrained joint prior distribution for dependent parameters of GvM

A.3.1 Bivariate exponential conditionals distribution

The density function of bivariate exponential conditionals (BEC) distribution is given by

$$f(\kappa_1, \kappa_2 | \alpha, \beta, \gamma) = c \exp(-(\alpha \kappa_1 + \beta \kappa_2 + \gamma \kappa_1 \kappa_2)) \tag{A.4}$$

for $\alpha > 0, \beta > 0, \gamma > 0$ and $\kappa_1, \kappa_2 > 0$, where c denotes unknown normalizing constant. (see e.g, [Arnold et al. \(1999\)](#))

For our case, κ_1, κ_2 are constrained as $0 < \kappa_1 < 4\kappa_2$, $0 < \kappa_2 < \infty$, and hence, we have a truncated BEC distribution. The conditional distributions of κ_1 and κ_2 are truncated exponential and exponential distribution $f(\kappa_1|\kappa_2) = \text{TExp}(\alpha + \gamma\kappa_2, 4\kappa_2)$ and $f(\kappa_2|\kappa_1) = \text{Exp}(\beta + \gamma\kappa_1)$, respectively. Here, in order to obtain a sample from truncated BEC distribution, we use two stage Gibbs sampler algorithm as follows

Algorithm 1: Simulation of bivariate exponential conditionals distribution

Give a starting point $\phi^{(0)} = (\kappa_1^{(0)}, \kappa_2^{(0)})$,

$$\kappa_1^{(s)} \sim f(\kappa_1|\kappa_2^{(s-1)})$$

$$\kappa_2^{(s)} \sim f(\kappa_2|\kappa_1^{(s)})$$

$$\phi^{(s)} = (\kappa_1^{(s)}, \kappa_2^{(s)})$$

A.4 Constrained joint prior distributions of dependent parameters for AGvM

A.4.1 Bivariate beta distribution

Suppose that independent random variables Y_1, Y_2, Y_3 have standard gamma distribution with respective shape parameters, a, b, c and

$$\kappa'_1 = \frac{Y_1}{Y_1+Y_3}, \quad \kappa'_2 = \frac{Y_2}{Y_2+Y_3}$$

The marginal distribution of κ'_1 and κ'_2 are beta distributions with parameters (a, c) and (b, c) respectively. The joint probability density of bivariate beta distribution is defined as:

$$f(\kappa'_1, \kappa'_2) = \frac{(\kappa'_1)^{a-1} (\kappa'_2)^{b-1} (1 - \kappa'_1)^{b+c-1} (1 - \kappa'_2)^{a+c-1}}{B(a, b, c) (1 - \kappa'_1 \kappa'_2)^{a+b+c}}, \quad 0 < \kappa'_1, \kappa'_2 < 1 \quad (\text{A.5})$$

where $B(\cdot)$ is beta function. ([Olkin and Liu \(2003\)](#)).

A.4.2 Bivariate Dirichlet distribution

Let Y_1, Y_2, Y_3 be independent random variables which have standard gamma distribution with respective shape parameters, a, b, c and

$$\kappa'_1 = \frac{Y_1}{Y_1+Y_2+Y_3}, \quad \kappa'_2 = \frac{Y_2}{Y_1+Y_2+Y_3}$$

The joint density of κ'_1, κ'_2 is defined by

$$f(\kappa'_1, \kappa'_2) = \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)} (\kappa'_1)^{a-1} (\kappa'_2)^{b-1} (1-\kappa'_1-\kappa'_2)^{c-1}, \quad \kappa'_1 + \kappa'_2 < 1. \quad (\text{A.6})$$

(see [Wilks \(1963\)](#)) This distribution may be considered as a special case when the κ'_1 and κ'_2 parameters are too small.

A.4.3 Bivariate beta conditionals distribution

This distribution is a special case of the following theorem.

Theorem : Suppose that $f_1(\kappa'_1; \theta)$ and $f_2(\kappa'_2; \eta)$ denote l_1 and l_2 parameter exponential families respectively. Let $f(\kappa'_1, \kappa'_2)$ be a bivariate density whose conditional densities satisfy $f(\kappa'_1 | \kappa'_2) = f_1(\kappa'_1; \underline{\theta}(\kappa'_2))$ and $f(\kappa'_2 | \kappa'_1) = f_2(\kappa'_2; \underline{\eta}(\kappa'_1))$ for some functions $\underline{\theta}(\kappa'_2)$, and $\underline{\eta}(\kappa'_1)$. Then the joint density $h(\kappa'_1, \kappa'_2)$ is defined by

$$h(\kappa'_1, \kappa'_2) = r_1(\kappa'_1) r_2(\kappa'_2) \exp(\underline{q}^{(1)}(\kappa'_1) M_{(l_1+1) \times (l_2+1)} \underline{q}^{(2)}(\kappa'_2)) \quad (\text{A.7})$$

where $\underline{q}^{(1)}(\kappa'_1) = (q_{10}(\kappa'_1), q_{11}(\kappa'_1), \dots, q_{1l_1}(\kappa'_1))$ and $\underline{q}^{(2)}(\kappa'_2) = (q_{20}(\kappa'_2), q_{21}(\kappa'_2), \dots, q_{2l_2}(\kappa'_2))$ with $q_{10}(\kappa'_1) = q_{20}(\kappa'_2) \equiv 1$ and $M_{(l_1+1) \times (l_2+1)}$ is matrix of constant parameters (see [Arnold et al. \(1999\)](#)).

The joint density of bivariate beta conditionals distribution is defined below, following by above Theorem.

$$\frac{1}{\kappa'_1(1-\kappa'_1)} \frac{1}{\kappa'_2(1-\kappa'_2)} \exp\left(\begin{pmatrix} 1 & \log \kappa'_1 & \log(1-\kappa'_1) \end{pmatrix} M_{3 \times 3} \begin{pmatrix} 1 \\ \log \kappa'_2 \\ \log(1-\kappa'_2) \end{pmatrix}\right) \quad (\text{A.8})$$

$$0 < \kappa'_1, \kappa'_2 < 1$$

where $M_{3 \times 3} = \begin{pmatrix} m_{00} & m_{01} & m_{02} \\ m_{10} & m_{11} & 0 \\ m_{20} & 0 & 0 \end{pmatrix}$ and the conditional distributions of κ'_1, κ'_2 are specified by

$$f(\kappa'_2 | \kappa'_1) = \text{Beta}(m_{11} \log(\kappa'_1) + m_{01}, m_{02}),$$

$$f(\kappa'_1 | \kappa'_2) = \text{Beta}(m_{11} \log(\kappa'_2) + m_{10}, m_{20})$$

respectively. The marginal distribution is not a well known form and it can be showed as shown below

$$f(\kappa'_1) = c x^{m_{10}-1} (1-x)^{m_{20}-1} \frac{\Gamma(m_{11} \log x + m_{21} \log(1-x) + m_{01}) \Gamma(m_{12} \log x + m_{21} \log(1-x) + m_{02})}{\Gamma((m_{11} + m_{12}) \log x + (m_{21} + m_{22}) \log(1-x) + m_{01} + m_{02})} \quad (\text{A.9})$$

where c is the constant of the integral and corresponds to $\exp(m_{00})$.

In order to obtain a sample from these bivariate distributions, we use the following algorithms. The first and second algorithms are related to direct simulation, and the last algorithm is related to Gibbs sampler to obtain a bivariate conditional beta distribution by using conditional distributions.

Algorithm 2: Simulation of bivariate beta distribution

Generate independently gamma random variables, Y_1, Y_2, Y_3 with shape parameter a, b, c , respectively.

$$\kappa'_1 = \frac{Y_1}{Y_1+Y_3}, \kappa'_2 = \frac{Y_2}{Y_2+Y_3}$$
$$\kappa_2 = 2\kappa'_2 - 1 \text{ and } \kappa_1 = 2\kappa'_1|\kappa_2|$$

Accept if $\kappa_1 < 2|\kappa_2|$ else return first step

Algorithm 3: Simulation of bivariate Dirichlet distribution

Generate independently gamma random variables, Y_1, Y_2, Y_3 with shape parameter a, b, c , respectively.

$$\kappa'_1 = \frac{Y_1}{Y_1+Y_2+Y_3}, \kappa'_2 = \frac{Y_2}{Y_1+Y_2+Y_3}$$
$$\kappa_2 = 2\kappa'_2 - 1 \text{ and } \kappa_1 = 2\kappa'_1|\kappa_2|$$

Accept if $\kappa_1 < 2|\kappa_2|$ else return first step.

Algorithm 4: Simulation of bivariate beta conditionals distribution

Let $\phi^{(0)} = (\kappa_1^{(0)}, \kappa_2^{(0)})$, be a starting point

$$\kappa_1^{(s)} \sim f(\kappa_1 | \kappa_2^{(s-1)})$$

$$\kappa_2^{(s)} \sim f(\kappa_2 | \kappa_1^{(s)})$$

$$\kappa_2^{(s)} = 2\kappa_2^{(s)} - 1 \text{ and } \kappa_1^{(s)} = 2\kappa_1^{(s)}|\kappa_2^{(s)}|$$

$$\phi^{(s)} = (\kappa_1^{(s)}, \kappa_2^{(s)})$$

APPENDIX B

APPENDIX FOR CHAPTER 4

B.1 Posterior computation for DP mixture von Mises model

The full conditional distributions for the parameters of DP mixture von Mises model are as follows:

Let K_1^*, \dots, K_m^* be the current m unique values of \mathbf{K} . The conditional distribution $\boldsymbol{\varphi}|\mathbf{K}, \boldsymbol{\theta}$ can be decomposed as $\boldsymbol{\mu}|\boldsymbol{\kappa}, \mathbf{K}, \boldsymbol{\theta}$ and $\boldsymbol{\kappa}|\boldsymbol{\mu}, \mathbf{K}, \boldsymbol{\theta}$. In each iteration of Gibbs sampler, we simulate as

Conditional distribution for $\boldsymbol{\mu}$: For each $j \in K_1^*, \dots, K_m^*$, draw

$$\mu_j|\boldsymbol{\kappa}, \mathbf{K}, \boldsymbol{\theta} \propto \exp(\kappa_0 \cos(\mu_j - \mu_0) + \sum_{i:K_i=j} \kappa_j \cos(\theta_i - \mu_j)).$$

In OpenBUGS, block-hybrid sampling algorithm is performed to obtain random samples from $\mu_j|\boldsymbol{\kappa}, \mathbf{K}, \boldsymbol{\theta}$. Also for each $j \in \mathbf{K} - K_1^*, \dots, K_m^*$, independently simulate $\mu_j \sim \text{vM}(\mu_0, \kappa_0)$.

Conditional distribution for $\boldsymbol{\kappa}$: For each $j \in K_1^*, \dots, K_m^*$, draw

$$\kappa_j|\boldsymbol{\mu}, \mathbf{K}, \boldsymbol{\theta} \propto \frac{\kappa_j^{b_0-1}}{I_0(\kappa_j)^{n_j}} \exp\left(\sum_{i:K_i=j} \kappa_j \cos(\theta_i - \mu_j) - a_0 \kappa_j\right)$$

where $n_j = \#\{i : K_i = j\}$. Slice sampling algorithm is used to obtain random samples from the full conditional distribution of κ_j . Also for each $j \in \mathbf{K} - K_1^*, \dots, K_m^*$, independently simulate $\kappa_j \sim \text{Gamma}(a_0, b_0)$.

Conditional distribution for \mathbf{K} :

$$(K_i|\mathbf{p}, \boldsymbol{\mu}, \boldsymbol{\kappa}, \boldsymbol{\theta}) \sim \sum_{k=1}^C p_{k,i} I_k(\cdot), \quad i = 1, \dots, n$$

where

$$(p_{1,i}, \dots, p_{C,i}) \propto \frac{p_1}{I_0(\kappa_1)} \exp(\kappa_1(\cos(\theta_i - \mu_1))), \dots, \frac{p_C}{I_0(\kappa_C)} \exp(\kappa_C(\cos(\theta_i - \mu_C))).$$

Discrete slice sampling algorithm is used to obtain random samples from the full conditional distribution of \mathbf{K}

Conditional distribution for \mathbf{p} :

$$p_1 = q_1^* \text{ and } p_k = (1 - q_1^*)(1 - q_2^*) \dots (1 - q_{k-1}^*) q_k^*, \quad k = 2, \dots, C - 1$$

where

$$q_k^* \sim \text{Beta}(1 + n_k, \alpha + \sum_{l=k+1}^C n_l), \quad k = 1, \dots, C - 1$$

where $n_k = \#\{i : K_i = k\}$, that is, n_k saves the number of K_i values which set to k

Conditional distribution for α :

$$\alpha | \mathbf{p} \sim \text{Gamma}(C + v_1 - 1, v_2 - \sum_{k=1}^{C-1} \log(1 - q_k^*))$$

where q_k^* are same values in the simulation of \mathbf{p} .

B.2 Posterior computation for DP mixture wrapped Cauchy model

The full conditional distributions for the parameters of DP mixture wrapped Cauchy model are given by:

Let K_1^*, \dots, K_m^* be the current m unique values of \mathbf{K} . The conditional distribution $\boldsymbol{\varphi} | \mathbf{K}, \boldsymbol{\theta}$ can be decomposed as $\boldsymbol{\mu} | \boldsymbol{\rho}, \mathbf{K}, \boldsymbol{\theta}$ and $\boldsymbol{\rho} | \boldsymbol{\mu}, \mathbf{K}, \boldsymbol{\theta}$. In each iteration of Gibbs sampler, we simulate as

Conditional for $\boldsymbol{\mu}$: For each $j \in K_1^*, \dots, K_m^*$, draw

$$\mu_j | \boldsymbol{\rho}, \mathbf{K}, \boldsymbol{\theta} \propto \exp(\kappa_0 \cos(\mu_j - \mu_0)) + \sum_{i:K_i=j} \log\left(\frac{1}{1 + \rho_j^2 - 2\rho_j \cos(\theta_i - \mu_j)}\right)$$

it can be used Taylor expansion of $\log(1/(1+x))$ with ignored high order terms, we repeatedly write the full conditional distributions for $\boldsymbol{\mu}$ as follows

$$\propto \exp(\kappa_0 \cos(\mu_j - \mu_0)) + \sum_{i:K_i=j} 2\rho_j \cos(\theta_i - \mu_j) - \rho_j^2$$

and then, the block-hybrid sampling algorithm is performed to obtain random samples from the full conditional distribution of μ_j . Also for each $j \in \mathbf{K} - K_1^*, \dots, K_m^*$, independently simulate $\mu_j \sim \text{vM}(\mu_0, \kappa_0)$.

Conditional distribution for ρ : For each $j \in K_1^*, \dots, K_m^*$, draw

$$\rho_j | \boldsymbol{\mu}, \mathbf{K}, \boldsymbol{\theta} \propto \rho_j^{a_0-1} (1 - \rho_j)^{b_0-1} \prod_{i:K_i=j} \frac{1 - \rho_j^2}{1 + \rho_j^2 - 2\rho_j \cos(\theta_i - \mu_j)}.$$

Slice sampling algorithm is performed to obtain random samples from $\rho_j | \boldsymbol{\mu}, \mathbf{K}, \boldsymbol{\theta}$.

Also for each $j \in \mathbf{K} - K_1^*, \dots, K_m^*$, independently simulate $\rho_j \sim \text{Beta}(a_0, b_0)$.

Conditional distribution for \mathbf{K} :

$$(K_i | \boldsymbol{\rho}, \boldsymbol{\mu}, \boldsymbol{\kappa}, \boldsymbol{\theta}) \sim \sum_{k=1}^C p_{k,i} I_k(\cdot), \quad i = 1, \dots, n$$

where

$$(p_{1,i}, \dots, p_{C,i}) \propto p_1 \frac{1 - \rho_1^2}{1 + \rho_1^2 - 2\rho_1 \cos(\theta_i - \mu_1)}, \dots, p_C \frac{1 - \rho_C^2}{1 + \rho_C^2 - 2\rho_C \cos(\theta_i - \mu_C)}.$$

The discrete slice sampling algorithm is used to obtain random samples from the full conditional distribution of \mathbf{K} . The others full conditional distributions are same in given by Appendix B.1.

B.3 OpenBUGS codes

In this part, we share our OpenBUGS codes. We define von Mises distribution and wrapped Cauchy distribution via of new specifying distributions using ones trick. For new prior distribution, the likelihood for θ when this is combined with a flat prior for θ the correct prior results. Our codes with respect to wrapped Cauchy distribution are shown below

```
#Wrapped Cauchy distribution
model{
  const<-10000
  Pi <- 3.14159265359
  for (i in 1:N) {
```

```

z[i]<- 1
z[i] ~ dbern(wc[i])
L[i] <- (1/(2*Pi))*(1-rho[ K[i] ]*rho[ K[i] ])/(1+rho[ K[i] ]*
rho[ K[i] ] -2*rho[K[i]] *cos(theta[i]-mu[ K[i] ] ))
wc[i]<-L[i]/const
K[i] ~ dcat(p[])
}
#Constructive DPP
#stick-breaking prior
p[1]<- q[1]; q[C]<-1
for (j in 2:C)
{p[j]<-q[j]*(1-q[j-1])*p[j-1]/q[j-1] }
for (k in 1:C-1) {q[k]~dbeta(1,alpha)}

# Baseline distribution

for (k in 1:C){
rho[k]~dbeta(0.5,0.5)
}
Ikappa0<-exp(kappa0)/sqrt(2*Pi*kappa0)
#hyperparameters
kappa0<-7
mu0<-0
for (k in 1:C){
mu[k]~dflat()
z1[k]<-1
z1[k]~dbern(phi[k])
L2[k]<- 1/(Ikappa0) *exp(kappa0*cos(mu[k]-mu0))
phi[k]<-L2[k]/const
}

#DPP parameter prior
alpha~dgamma(1,1)

```

```

#Programing for calculating summary statistics#
for ( i in 1:N) { for (j in 1:C) {
SC[i,j]<-equals(j, K[i])
} }
#total clusters #
for (j in 1:C) {cl[j]<-step(sum(SC[,j])-1)}
Cluster<-sum(cl[])
}
}

```

The likelihood of von Mises distribution in OpenBUGS is defined as follows

```

#Likelihood

const<-10000
pi<-3.14159265359
for (i in 1:N) {
z[i]<-1
z[i]~dbern(phi[i])
t[i]<-kappa[K[i]]/3.75
Ikappa00[i]<-1+3.5156229*pow(t[i],2)+3.0899424*pow(t[i],4)+
1.2067492*pow(t[i],6)+
0.2659732*pow(t[i],8)+0.0360768*pow(t[i],10)
+0.0045813*pow(t[i],12)
Ikappa01[i]<-exp(kappa[K[i]])/sqrt(2*pi*kappa[K[i]])
L1[i]<- 1/(Ikappa00[i]) * exp(kappa[K[i]]*cos(theta[i]-mu[K[i]]))
L2[i]<-1/(Ikappa01[i]) * exp(kappa[K[i]]*cos(theta[i]-mu[K[i]]))
L[i]<-L1[i]*step(3.75-kappa[K[i]])+L2[i]*step(kappa[K[i]]-3.75)
K[i] ~ dcat(p[])
phi[i]<-L[i]/const
}
}

```


APPENDIX C

APPENDIX FOR CHAPTER 5

C.1 Circular-Circular association

Let (Θ_1, Ψ_1) and (Θ_2, Ψ_2) be two independent random vectors of (Θ, Ψ) , and the circular correlation coefficient introduced by Fisher and Lee (1983) is defined as follows

$$\rho_T = \frac{E\{\sin(\Theta_1 - \Theta_2)\sin(\Psi_1 - \Psi_2)\}}{\{E[\sin^2(\Theta_1 - \Theta_2)]E[\sin^2(\Psi_1 - \Psi_2)]\}^{1/2}}$$

This circular correlation coefficient takes values between -1 and 1 , and is 0 if Θ and Ψ are independent otherwise dependent.

Given a random sample of n observations of (Θ, Ψ) , $(\theta_1, \psi_1), \dots, (\theta_n, \psi_n)$, the estimate of ρ_T is given by

$$\hat{\rho}_T = \frac{4(AB - CD)}{\{(n^2 - E^2 - F^2)(n^2 - G^2 - H^2)\}^{1/2}}$$

where

$$\begin{aligned} A &= \sum_{j=1}^n \cos\theta_j \cos\psi_j, & B &= \sum_{j=1}^n \sin\theta_j \sin\psi_j, \\ C &= \sum_{j=1}^n \cos\theta_j \sin\psi_j, & D &= \sum_{j=1}^n \sin\theta_j \cos\psi_j, \\ E &= \sum_{j=1}^n \cos 2\theta_j, & F &= \sum_{j=1}^n \sin 2\theta_j, & G &= \sum_{j=1}^n \cos 2\psi_j, & H &= \sum_{j=1}^n \sin 2\psi_j. \end{aligned}$$

In order to examine the correlation patterns of Θ_t , we use this circular correlation coefficient. Here, given a random sample of n observations of Θ_t , which can be written as $(\theta_1, \theta_{k+1}), \dots, (\theta_{n-k}, \theta_n)$, $k \geq 0$, then we compute $\hat{\rho}_{T,k}$ which is defined as k -lag sample circular autocorrelation.

C.2 Posterior computation for DP mixture Möbius model

The full conditional distributions for the parameters of DP mixture Möbius time series model are as follows:

Let K_1^*, \dots, K_m^* be the current m unique values of \mathbf{K} . In each iteration of the Gibbs sampler, we simulate

Conditional for $\boldsymbol{\mu}$: For each $j \in K_1^*, \dots, K_m^*$, draw

$$\mu_j | \boldsymbol{\lambda}, \boldsymbol{\kappa}, \mathbf{K}, \boldsymbol{\theta} \propto \exp(\kappa_0 \cos(\mu_j - \mu_0) + \sum_{t:K_t=j} \kappa_j \cos(\theta_t - (\mu_j + 2\text{atan}\{\lambda_j \tan \frac{1}{2}(\theta_{t-1} - \mu_j)\}))).$$

Adaptive Metropolis Hastings algorithm is performed to obtain random samples from $\mu_j | \boldsymbol{\lambda}, \boldsymbol{\kappa}, \mathbf{K}, \boldsymbol{\theta}$. Also for each $j \in \mathbf{K} - K_1^*, \dots, K_m^*$, independently simulate $\mu_j \sim \text{vM}(\mu_0, \kappa_0)$.

Conditional for $\boldsymbol{\lambda}$: For each $j \in K_1^*, \dots, K_m^*$, draw

$$\lambda_j | \boldsymbol{\mu}, \boldsymbol{\kappa}, \mathbf{K}, \boldsymbol{\theta} \propto \exp(\sum_{t:K_t=j} \kappa_j \cos(\theta_t - (\mu_j + 2\text{atan}\{\lambda_j \tan \frac{1}{2}(\theta_{t-1} - \mu_j)\}))).$$

Slice sampling algorithm is performed to obtain random samples from $\lambda_j | \boldsymbol{\mu}, \boldsymbol{\kappa}, \mathbf{K}, \boldsymbol{\theta}$. Also for each $j \in \mathbf{K} - K_1^*, \dots, K_m^*$, independently simulate $\lambda_j \sim \text{Unif}(a_0, b_0)$.

Conditional distribution for $\boldsymbol{\kappa}$: For each $j \in K_1^*, \dots, K_m^*$, draw

$$\kappa_j | \boldsymbol{\mu}, \mathbf{K}, \boldsymbol{\theta} \propto \frac{\kappa_j^{b_0-1}}{I_0(\kappa_j)^{n_j}} \exp(\sum_{t:K_t=j} \kappa_j \cos(\theta_t - (\mu_j + 2\text{atan}\{\lambda_j \tan \frac{1}{2}(\theta_{t-1} - \mu_j)\}))) - a_0 \kappa_j$$

where $n_j = \#\{t : K_t = j\}$. The slice sampling algorithm is used to obtain random samples from the full conditional distribution of κ_j . Also for each $j \in \mathbf{K} - K_1^*, \dots, K_m^*$, independently simulate $\kappa_j \sim \text{Gamma}(a_0, b_0)$.

Conditional distribution for \mathbf{K}

$$(K_t | \mathbf{p}, \boldsymbol{\mu}, \boldsymbol{\kappa}, \boldsymbol{\theta}) \sim \sum_{k=1}^C p_{k,t} I_k(\cdot), \quad t = 2, \dots, n$$

where

$$(p_{1,t}, \dots, p_{C,t}) \propto \frac{p_1}{I_0(\kappa_1)} \exp(\kappa_1 (\cos(\theta_t - (\mu_1 + 2\text{atan}\{\lambda_1 \tan \frac{1}{2}(\theta_{t-1} - \mu_1)\}))), \dots, \frac{p_C}{I_0(\kappa_C)} \exp(\kappa_C (\cos(\theta_t - (\mu_C + 2\text{atan}\{\lambda_C \tan \frac{1}{2}(\theta_{t-1} - \mu_C)\}))).$$

Conditional distribution for \mathbf{p}

$$p_1 = q_1^* \text{ and } p_k = (1 - q_1^*)(1 - q_2^*) \dots (1 - q_{k-1}^*)q_k^*, \quad k = 2, \dots, C - 1$$

where

$$q_k^* \sim \text{Beta}(1 + n_k, \alpha + \sum_{l=k+1}^C n_l), \quad k = 1, \dots, C - 1$$

where $n_k = \#\{t : K_t = k\}$, that is, n_k saves the number of K_t values which set to k

Conditional distribution for α :

$$\alpha | \mathbf{p} \sim \text{Gamma}(C + v_1 - 1, v_2 - \sum_{k=1}^{C-1} \log(1 - q_k^*))$$

where q_k^* are same values in the simulation of \mathbf{p} .

C.3 OpenBUGS codes

In the following codes, we give R2OpenBUGS codes for our Möbius time series model. Stick breaking implementation is similar with previous chapter.

```

circmodel <- function() {
const<-10000
pi<-3.14159265359
for (i in 2:N) {
z[i]<-1
z[i]~dbern(phi[i])
t[i]<-kappa[T[i]]/3.75
Ikappa00[i]<-1+3.5156229*pow(t[i],2)+3.0899424*pow(t[i],4)+
1.2067492*pow(t[i],6)+
0.2659732*pow(t[i],8)+0.0360768*pow(t[i],10)
+0.0045813*pow(t[i],12)
Ikappa01[i]<-exp(kappa[T[i]])/sqrt(2*pi*kappa[T[i]])
L1[i]<- 1/(Ikappa00[i]) * exp(kappa[T[i]]*cos(theta[i]-mut[i]))
L2[i]<-1/(Ikappa01[i]) * exp(kappa[T[i]]*cos(theta[i]-mut[i]))
L[i]<-L1[i]*step(3.75-kappa[T[i]])+L2[i]*step(kappa[T[i]]-3.75)
}
}

```

```

T[i] ~ dcat(p[])
phi[i]<-L[i]/const
}
for (i in 2:N){
w[i]<-theta[i-1]-mu[T[i]]
mut[i] <-mu[T[i]]+2*arctan(lambda[T[i]]*tan(w[i]/2))
}
mut[1]<-theta[1]
# Constructive DPP

#stick-breaking prior
p[1]<- r[1]; r[C]<-1
for (j in 2:C) {p[j]<-r[j]*(1-r[j-1])*p[j-1]/r[j-1] }
for (k in 1:C-1) {r[k]~dbeta(1,alpha)}

# Baseline distribution
kappa0<-0.1
t1<-kappa0/3.75
mu0<-0.1
Ikappa0p<-1+3.5156229*pow(t1,2)+3.0899424*pow(t1,4)+
1.2067492*pow(t1,6)+
0.2659732*pow(t1,8)+0.0360768*pow(t1,10)+
0.0045813*pow(t1,12)
#Ikappa0p<-exp(kappa0)/sqrt(2*pi*kappa0)
#for (k in 1:C){
mu[1]~dunif(-3.14159265359,3.14159265359)
z1[1]<-1
z1[1]~dbern(hiph[1])
L3[1]<-1/(Ikappa0p)*exp(kappa0*cos(mu[1]-mu0))
hiph[1]<-L3[1]/const
#}

```

```

for ( k in 2:C){
mu[k]<-(pi/2-mu[k-1])*delta[k-1]
}
for(k in 1:C-1)
{
delta[k]~dunif(0,1)
}
for (k in 1:C){
kappa[k]~dgamma(0.01,0.01)
}
#DPP parameter prior
#alpha~dunif(0.5,10)
alpha~dgamma(2,2)
for (k in 1:C){
lambda[k]~dunif(0,1)
}
}

circcdata <- list("theta","N","C")
circinits <- function() {
  list(alpha=1,
mu=c(-1.5,NA),
kappa=c(2,0.2),lambda=c(0.5,0.5),r=c(0.5,NA))}
## Uses default settings for n.burnin = n.iter/2; n.thin=10;
circout <- bugs(data = circcdata, inits = circinits,
parameters.to.save = c("alpha",
"mu", "kappa","p","lambda"), model.file = circmodel,
n.chains = 3,n.burnin=40001, n.iter = 80001,n.thin=10,
debug=TRUE)

```


CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Kılıç, Muhammet Burak

Nationality: Turkish (TC)

Date and Place of Birth: 02.11.1986, Elazığ

Marital Status: Married

Phone: 0 530 242 28 55

EDUCATION

Degree	Institution	Year of Graduation
M.S.	Firat University, Department of Statistics	M.S. 2011
B.S.	Firat University, Department of Statistics	B.S. 2009
B.S.	Firat University, Department of Mathematics	B.S. 2009

PROFESSIONAL EXPERIENCE

Year	Place	Enrollment
2010-2011	Mehmet Akif Ersoy University, Burdur	Research Assistant
2011-2015	Middle East Technical University, Ankara	Research Assistant
2014	University of California, Riverside	Visiting Graduate Student
2015	Indian Statistical Institute, Kolkata	Visiting Researcher

PUBLICATIONS

Papers

Guzel, E, Yasar, M, Kılıç, M.B. and Canyılmaz, M. (2013). A new approach on investigation of the relationship of vlf signals by using longitudinal analysis model, *International Journal of Antennas and Propagation*, Volume 2013, Article ID 861682, 7 pages.

Unpublished Papers

A Bayesian analysis of asymmetric-bimodal circular data

Bayesian semi parametric models for circular data

Bayesian semi parametric time series model for circular data

International Conference Publications

Kılıç, M.B, SenGupta, A (2014). A Bayesian analysis of asymmetric generalized von Mises distribution, International Indian Statistical Association Conference 2014, 11-13 July, Riverside, USA.

Kılıç, M.B, Kalaylioglu, Z. (2013). A Bayesian model of wind direction, 9th PhD Seminar on Wind Energy in Europe, p.44, Uppsala University Campus Gotland, September 18-20, Sweden.

Aglaz, M., Kılıç, M.B., and Purutcuoglu, V. (2013). Deterministic modelling of gene network via parametric and non-parametric approaches , Proceeding of the 29th Meeting of Statisticians, Budapest, Hungary.

Yasar, M., Kılıç, M.B. Canyılmaz, M. and Guzel, E. (2012). Statistical approach on very low frequency Electromagnetic Wave, 29. International Physic Congress, p.646, Nurol cultural center, Bodrum.

Aydođdu S., Grcan, M., Kılıç, M.B. (2012). Generalized beta distribution and its properties, 13th International Conference on Econometrics, Operations Research and Statistics, Famagusta, 24-26 May, Turkish Republic of Northern Cyprus.

Grcan, M. and Kılıç M.B. (2012). Exchangeable parameters binomial approximation, International Conference on Applied Mathematics and Approximation Theory, pp.76 May 17-19, Ankara, Turkey.

Kılıç , M.B., Grcan, M. and Kaya, M.O. (2011). Statistical estimation method aid of Szasz Mirakjiyan operator, International 7th Statistical Congress, 28 April-1 May, Antalya, Turkey

Kılıç M.B, Grcan, M. and Aydođdu, S. (2010). Bernstein polinomlarının bazı özellikleri ve istatistiksel çıkarsamaları, İstatistik Gnleri Sempozyumu, pp. 10-11, 28-30 Haziran, Ankara, Turkey.