

ABSTRACT

RAVINDRAN, PALANIKUMAR. Bayesian Analysis of Circular Data Using Wrapped Distributions. (Under the direction of Associate Professor Sujit K. Ghosh).

Circular data arise in a number of different areas such as geological, meteorological, biological and industrial sciences. We cannot use standard statistical techniques to model circular data, due to the circular geometry of the sample space. One of the common methods used to analyze such data is the wrapping approach. Using the wrapping approach, we assume that, by wrapping a probability distribution from the real line onto the circle, we obtain the probability distribution for circular data. This approach creates a vast class of probability distributions that are flexible to account for different features of circular data. However, the likelihood-based inference for such distributions can be very complicated and computationally intensive. The EM algorithm used to compute the MLE is feasible, but is computationally unsatisfactory. Instead, we use Markov Chain Monte Carlo (MCMC) methods with a data augmentation step, to overcome such computational difficulties. Given a probability distribution on the circle, we assume that the original distribution was distributed on the real line, and then wrapped onto the circle. If we can *unwrap* the distribution off the circle and obtain a distribution on the real line, then the standard statistical techniques for data on the real line can be used. Our proposed methods are flexible and computationally efficient to fit a wide class of wrapped distributions. Furthermore, we can easily compute the usual summary statistics. We present extensive simulation studies to validate the performance of our method. We apply our method to several real data sets and compare our results to parameter estimates available in the literature. We find that

the Wrapped Double Exponential family produces robust parameter estimates with good frequentist coverage probability. We extend our method to the regression model. As an example, we analyze the association between ozone data and wind direction. A major contribution of this dissertation is to illustrate a technique to interpret the circular regression coefficients in terms of the linear regression model setup. Regression diagnostics can be developed after augmenting wrapping numbers to the circular data (refer Section 3.5). We extend our method to fit time-correlated data. We can compute other statistics such as circular autocorrelation functions and their standard errors very easily. We use the Wrapped Normal model to analyze the hourly wind directions, which is an example of the time series circular data.

Bayesian Analysis of Circular Data Using Wrapped Distributions

by

Palanikumar Ravindran

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial satisfaction of the
requirements for the Degree of
Doctor of Philosophy

Department of Statistics

Raleigh

2002

Approved By:

Dr. Peter Bloomfield

Dr. Sastry Pantula

Dr. Sujit K. Ghosh
Chair of Advisory Committee

Dr. John Monahan

To my parents

Biography

Palanikumar Ravindran was born in Virudhunagar, India, to parents Ravindran Palanichamy and Rajakumari Ravindran on Oct 25, 1976. He entered Indian Statistical Institute, Calcutta, India in 1994 and received a B.S. and M.S. in Statistics in 1997 and 1999, respectively. Since August 1999, he has studied for the doctoral degree in the Department of Statistics at North Carolina State University, under the supervision of Dr. Sujit Ghosh.

Acknowledgements

I would like to express my deepest gratitude and appreciation to my advisor Dr. Sujit Ghosh for his guidance, encouragement and support throughout my dissertation research. I would also like to thank the other committee members, Dr. John Monahan, Dr. Peter Bloomfield, Dr. Sastry Pantula and Dr. Gene Brothers for their careful reading of this manuscript and their helpful comments.

It has been a very pleasant experience to study in this department. I would like to convey my gratitude to the faculty and staff for their help and assistance. Thanks to my fellow students for their help and support. I would also like to thank Dr. Kaushik Ghosh, George Washington University, for the Splus code to plot the circular histograms.

Contents

List of Figures	vii
List of Tables	viii
1 Introduction to Circular Data	1
1.1 Some real life applications of circular data	1
1.2 Statistical Approaches to model circular data	6
1.2.1 The Embedding Approach	6
1.2.2 Intrinsic Approach	8
1.2.3 Wrapping Approach	9
2 Parameter Estimation for Wrapped Distributions	12
2.1 Previous Work	12
2.2 The Data Augmentation Approach	13
2.3 Model selection	21
2.4 Simulation studies	23
2.5 Validation approach for MCMC	32
2.6 Wrapped Extreme Value and Bimodal distributions	34
2.7 Application to real data sets	38
2.7.1 Jander's ant data	38
2.7.2 Ozone data set	42
2.8 Discussion	43
3 Circular Regression	45
3.1 Introduction	45
3.2 Previous work	46
3.3 Extension of the Data Augmentation Approach for Regression	48
3.4 Simulation studies	53
3.5 Regression Diagnostic plots	62
3.6 Application to real data sets	64
3.7 Discussion	70

4	Circular Time Series	71
4.1	Introduction	71
4.2	Previous work	72
4.3	Extension of the Data Augmentation Approach for Time Series	73
4.4	Simulation studies	78
4.5	Application to real data sets	79
4.6	Discussion	82
	Bibliography	84
A	Explicit full conditionals	90
A.1	Wrapped Normal Distribution	90
A.2	Wrapped Cauchy Distribution	92
A.3	Wrapped Double Exponential Distribution	93
A.4	Wrapped Extreme Value Distribution	95
A.5	Bimodal (Wrapped Beta) Distribution	96
B	Explicit full conditionals for regression	98
B.1	Wrapped Normal Distribution	98
B.2	Wrapped Cauchy Distribution	100
B.3	Wrapped Double Exponential Distribution	102
C	Explicit full conditionals for Time series	105
C.1	Wrapped Normal Distribution	105
D	Regression with categorical covariate	109

List of Figures

1.1	Circular histogram plot of the turtle data. Solid line indicates the circular mean and dashed line indicates the linear mean.	4
2.1	Probability plot of $\{H_k\}$ for μ parameter of WN model	33
2.2	Plot of ρ for the bimodal distribution	35
2.3	Plot of the bimodal density for $\rho = 0.087$	36
2.4	Plot of the bimodal density for $\rho = 0.087$ on the circle	36
2.5	Circular plot of the Jander's Ant data	39
2.6	Trace plots while fitting WC to Ant data	41
2.7	Posterior density of μ while fitting WC to Ant data	42
2.8	Plot of wind directions in the ozone data.	43
3.1	Joint data plot of the ozone data. The ozone concentrations are plotted as distances from the center.	46
3.2	Plot of the original values.	63
3.3	Regression Plot of the sample and fitted values.	63
3.4	Plot of wind directions in the ozone data.	65
3.5	Trace plots while fitting WN to the Ozone data	68
3.6	Regression Plot of the fitted values.	69
3.7	Plot of the residuals.	69
4.1	Plot of the median direction of face cleat collected at Wallsend Boreland Colliery, NSW in Australia.	72
4.2	Relation between circular autocorrelation function and autocorrelation function	77
4.3	Plot of hourly wind directions collected over three days.	80
4.4	Trace plots while fitting WARN(1) to Wind direction data using time series model	81
4.5	Plot of gamma, circular autocorrelation for WARN(1) model and sample circular autocorrelation at lags 1 to 7.	82

List of Tables

2.1	Fitting WN to WN distribution using different priors	24
2.2	Fitting WC to WC distribution using different priors	25
2.3	Fitting WDE to WDE distribution using different priors	25
2.4	Fitting WN, WC and WDE to WN distribution with $\rho \sim Beta(0.5, 0.5)$. .	27
2.5	Fitting WN, WC and WDE to WC distribution with $\rho \sim Beta(0.5, 0.5)$. .	28
2.6	Fitting WN, WC and WDE to WDE distribution with $\rho \sim Beta(0.5, 0.5)$. .	29
2.7	Fitting WN, WC and WDE to VM data with $\rho \sim Beta(0.5, 0.5)$	31
2.8	Fitting WEV to WEV distribution	37
2.9	Fitting WB to WB distribution	37
2.10	Fitting WN, WC and WDE to the Ant data set with $\rho \sim Beta(0.5, 0.5)$. .	40
2.11	Fitting WN, WC and WDE to the Ozone data set with $\rho \sim Beta(0.5, 0.5)$.	44
3.1	Regressing WN model to WN data	54
3.2	Regressing WC model to WC data	55
3.3	Regressing WDE model to WDE data	56
3.4	Regressing WN, WC and WDE models to WN data with $\beta_1 = 1.5$	58
3.5	Regressing WN, WC and WDE models to WC data with $\beta_1 = 1.5$	59
3.6	Regressing WN, WC and WDE models to WDE data with $\beta_1 = 1.5$	60
3.7	Regressing WN, WC and WDE models to VM data with $\beta_1 = 1.5$	61
3.8	Regressing WN, WC and WDE models to the Ozone data set with $\rho \sim$ $Beta(0.5, 0.5)$	67
4.1	Regressing WARN(1) model to WARN(1) data	79
4.2	Fitting WARN(1) to the wind direction data set with $\rho \sim Beta(0.5, 0.5)$. .	80
D.1	Regressing WN, WC and WDE models to WN data with categorical covariate	111
D.2	Regressing WN, WC and WDE models to WC data with categorical covariate	112
D.3	Regressing WN, WC and WDE models to WDE data with categorical covariate	113

Chapter 1

Introduction to Circular Data

Circular data arise from a number of sources in our daily lives, where we consider the circle to be the sample space. Some common examples are the migration paths of birds and animals, wind directions, ocean current directions and patients' arrival times in an emergency ward of a hospital. Many examples of circular data are found in various scientific fields such as earth sciences, meteorology, biology, physics, psychology and medicine. To motivate the use of circular data, we present a brief description of some examples from these fields.

1.1 Some real life applications of circular data

The study of earth sciences yields two good examples of circular data - The orientation of cross-bedding structures and the orientation of the long axis of unbroken sediment particles. Orientation of the cross-bedding structures gives us information about the properties of the rock structure for mineral and petroleum exploration. Orientation of pebbles

has proved to be useful in the study of glacial deposits and direction of ice movement. Pincus (1953) presents several such examples.

In meteorology, wind directions and ocean current directions give rise to circular data. Johnson and Wehrly (1977) did some analysis of wind directions. Seasonal weather changes such as the propensity for rainfall during the monsoon season is another example of circular data.

In the field of physics, before the discovery of isotopes, Von Mises (1918) proposed testing the hypothesis that atomic weights are integers subject to error. He converted the fractional parts of the atomic weights to angles. He regarded these angles as a random sample from a circular distribution with mean zero and tested for uniformity. He also introduced Von Mises distribution, a popular distribution on the circle. In another study, Rayleigh (1919) worked with a representation of sound waves. He was interested in the resultant of unit vectors and its distribution. He considered the unit vectors as points on the circle.

In psychology, circular data arises from experiments to study the behavior of the human mind. Consider the simulated tests of zero gravity. Scuba divers were required to turn somersault and reorient themselves to the vertical under various circumstances (for example, blindfolded or looking through a translucent faceplate, see Ross et al., 1969). The angles from the vertical were measured and analyzed. Circular data also occur in the studies of mental maps, which are used to represent surroundings. Individual subjects were led past a series of sites. They were asked at each site to point to the direction, and guess the direction of every other site from that site (Gordon, Jupp and Byrne, 1989).

In medicine, circular data arises as the time of onset of a particular disease at various times of the year (Lee, 1962). Another example is circadian rhythms, which is the time of adverse event occurrences throughout the day. Circadian rhythms are analyzed because it has been found that adverse events (for example, deaths, myocardial infarctions) do not occur randomly throughout the day but cluster at certain points in the day (Proschan and Follmann, 1997).

A very good source of circular data is the field of biology. Migration path of birds and animals has been the subject of many studies. The objective of these studies is to ascertain whether the direction of migration is uniform. An example of the migration of turtles is given in Figure 1.1. In the figure, we present the circular histogram plot of the data collected by Dr. E. Gould from John Hopkins University School of Hygiene and first cited by Stephens (1969). The data represents the directions taken by the sea turtles after laying their eggs. The predominant direction is 64° , which is the direction the turtles took to return to the sea (Fraser, 1979).

Standard statistical techniques cannot be used to analyze circular data. This is due to the circular geometry of the sample space. For example, the sample mean of a data set on the circle is not the usual sample mean. Let y_1, y_2, \dots, y_n be independent observations on the unit circle, such that $0 \leq y_j < 2\pi$, $j = 1, 2, \dots, n$. The mean direction \bar{y} is not given by the usual definition, $\frac{1}{n} \sum_{j=1}^n y_j$. This is illustrated in Figure 1.1, where the dashed arrow is the direction represented by $\frac{1}{n} \sum_{j=1}^n y_j$ and the solid arrow represents the mean obtained by vector addition. To find the circular mean, we use vector addition techniques.

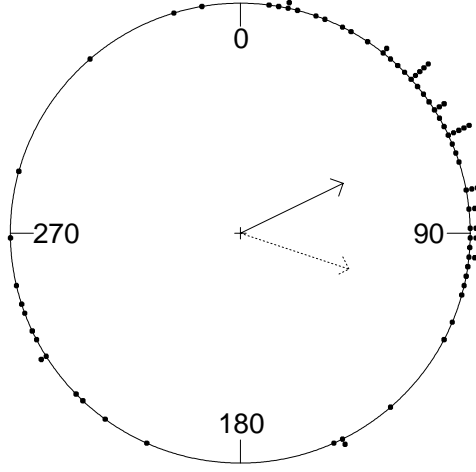


Figure 1.1: Circular histogram plot of the turtle data. Solid line indicates the circular mean and dashed line indicates the linear mean.

We consider $\bar{C} = \frac{1}{n} \sum_{j=1}^n \cos y_j$ and $\bar{S} = \frac{1}{n} \sum_{i=1}^n \sin y_j$ and define,

$$\bar{y} = \begin{cases} \arctan(\frac{\bar{S}}{\bar{C}}) & , \bar{C} \geq 0 \\ \arctan(\frac{\bar{S}}{\bar{C}}) + \pi & , \bar{C} < 0, \end{cases}$$

where \arctan takes values in $[-\frac{\pi}{2}, \frac{\pi}{2}]$. In general, the p^{th} theoretical moment of a circular distribution is defined as $E(e^{ipY}) = \alpha_p + i\beta_p$, for $p = 1, 2, \dots$. The mean direction is given by $\mu = \arctan(\beta_1/\alpha_1)$ and the mean resultant is defined as $\rho = \sqrt{\alpha_1^2 + \beta_1^2}$, so that $E(e^{iY}) = \rho e^{i\mu}$. In most applications it is of interest to estimate the location parameter μ and the scale parameter ρ . Nonparametric methods are suitable for this purpose. However, for the prediction problem, it would be of interest to develop parametric models. In addition, we exemplify that the class of parametric models that we develop is robust against erroneous models.

In Section 1.2, we discuss several statistical approaches to model circular data with

an emphasis on the wrapping approach. In this dissertation, we use the wrapping method to generate a flexible class of circular distributions. In chapter 2, we discuss classical and Bayesian methods to obtain estimates of the parameters of wrapped circular distributions. As the Bayesian methods have the advantage of obtaining a finite sample estimate of the variability (for example, s.e. of the estimates), we propose a data augmentation method for parameter estimation. In Section 2.2, we present the data augmentation method to obtain the posterior distribution of the parameters of several wrapped distributions. In Section 2.3, we describe the possible methods for model selection. In Section 2.4, we present extensive simulation studies to validate the frequentist performance of our method. In Section 2.6, we discuss the Wrapped Extreme Value and bimodal distributions. In Section 2.7.1, we apply our method to a real data set on the movement of ants. In chapter 3, we develop some models for regression where the response variable can be circular or linear. In Section 3.3, we present the extension of the data augmentation method proposed in Section 2.2 for regression to obtain the posterior distribution of the regression coefficients based on several wrapped distributions. In Section 3.4, we present extensive simulation studies to validate the performance of our method for regression. In Section 3.6, we fit a linear regression model to explore the relation between ozone concentration and wind direction. In Section 4.3, we illustrate how the data augmentation method used for regression in Section 3.3 can be easily extended for time series. Simulation studies with the Wrapped Normal density for the time series model is given in Section 4.4. In Section 4.5, we analyze hourly wind directions, which is an example of time series data. We use the Wrapped Normal model for this study.

1.2 Statistical Approaches to model circular data

Many methods and statistical techniques have been developed to analyze and understand circular data (see Mardia and Jupp, 1999). The popular approaches have been the *embedding* approach, *intrinsic* approach and *wrapping* approach. A brief description of each of the approaches is given below.

1.2.1 The Embedding Approach

In the embedding approach, the sample space (for example, the unit circle) is considered as a part of a larger space (for example, 2-dimensional plane). A common example is the representation of the points of the unit circle by unit complex numbers. There are many advantages of the embedding approach. Considering the points on the unit circle, as a vector $x = (\cos y, \sin y)^T$ in the plane enables the use of the traditional definition of expectations that is used for data in the Euclidean space. For instance, definition of the mean μ of a random variable y , defined on the unit circle is given by

$$\| E[(\cos y, \sin y)^T] \|^{-1} E[(\cos y, \sin y)^T] = (\cos \mu, \sin \mu)^T.$$

Several bivariate distributions on the Euclidean space can be embedded to produce distribution on the circle. For instance, the Projected Normal distribution is an example of the embedding approach. In the embedding approach, we start with the larger sample space (for example, 2-dimensional plane) and obtain the projection of this space into a smaller sample space. For example, if X has the Bivariate Normal distribution $N_2(\mu, \Sigma)$, then $\| X \|^{-1} X$ is said to have the Projected Normal distribution, $PN_2(\mu, \Sigma)$. This is used in meteorology, where the wind velocity is modeled by the Bivariate Normal distri-

bution and the resulting marginal distribution for wind direction is the Projected Normal distribution. The density of $PN_2(\mu, \Sigma)$ has been derived by Mardia (1972). The Projected Normal distribution can be extended to p -dimensions, where the distributions on \mathfrak{R}^p are projected onto S^{p-1} , the unit sphere in \mathfrak{R}^p . The density of the Projected Normal distribution, $PN_p(\mu, \Sigma)$ has been derived by Bingham (Watson, 1983, pp. 226-231) and a simpler form was derived by Pukkila & Rao (1988). However, in general, the densities obtained by embedding a generic distribution on \mathfrak{R}^p onto S^{p-1} , can turn out to be very complicated and hence obtaining the likelihood-based inference can be extremely challenging. Therefore, most of the literature is focused on developing statistical methods for the Projected Normal distributions only, which is a significant limitation of the embedding approach.

There are some techniques for parameter estimation in the embedding approach. Spherically projected multivariate linear (SPML) model using $PN_2(\mu, \Sigma)$ distribution was suggested by Presnell, Morrison and Littel (1998). They considered circular data as data sampled from a plane and projected onto the circle. They assumed that the distance from the center for each data point, $|y_i|$ was missing and used EM algorithm to estimate it. Embedding technique is also very useful to perform analysis of variance (ANOVA) for circular data. ANOVA for circular data was proposed by Harrison, Kanji and Gadsen (1986) and Harrison and Kanji (1988). We do not pursue any analysis based on the embedding approach in this research work, as the analysis becomes analytically and computationally intractable for most distributions other than $PN_2(\mu, \Sigma)$.

1.2.2 Intrinsic Approach

In the intrinsic approach for circular data, circle is used as the sample space. The directions (angles) are represented as points on the circle. In intrinsic approach, probability distributions are defined on the circle directly (for example, Von Mises and Cardioid distributions). Von Mises distribution is one of the most popular distributions that come out of this approach. The probability density function of the Von Mises distribution is given by $f_{VM}(y) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(y-\mu)}$, where I_0 denotes the modified Bessel function of the first kind and order 0. I_0 is defined by $I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} e^{\kappa \cos(y)} dy$. Von Mises distribution has been studied extensively. Mardia and Jupp (1999) give references for the genesis of the Von Mises distribution on the circle, which is analogous to the Normal distribution on the real line. Let $\mathbf{y} = \{y_1, y_2, y_3, \dots, y_n\}, 0 \leq y_i < 2\pi, i = 1 \dots n$ be a random sample from Von Mises distribution with location parameter μ and scale parameter κ . Define C, S and R as $C = \sum_{i=1}^n \cos y_i, S = \sum_{i=1}^n \sin y_i$ and $R^2 = C^2 + S^2$. $C = R \cos \bar{y}$ and $S = R \sin \bar{y}$, where \bar{y} is the mean as defined earlier (in Section 1.1). Mardia and Jupp (1999) provide the joint distribution of \bar{y} and R . They give the marginal densities of R, C and S using the results given by Greenwood and Durand (1955). Mardia (1972) showed that the conditional distribution of \bar{y} given R is Von Mises distribution with location parameter μ and scale parameter κR . Mardia and Jupp (1999) give results and references on several extensions of these results to multi-sample Von Mises populations. The asymptotic distributions of these statistics, as the sample size goes to infinity, are also available.

From a Bayesian perspective, the conjugate prior for the Von Mises distribution has been obtained by Guttorp and Lockhart (1988). Damien and Walker (1999) presents

full Bayesian analysis involving Von Mises distribution, where both the parameters are assumed to be unknown, and used its conjugate prior proposed by Guttorp and Lockhart. They proposed MCMC methods to simulate samples from the posterior distribution.

Mardia and Jupp (1999) describe the maximum likelihood estimates for the Von Mises distribution and give references and results for their large-sample asymptotic properties. They also provide a good overview of the various single-sample, two-sample and multi-sample hypothesis tests for the Von Mises distribution.

One of the main drawbacks of the intrinsic approach is that there are not many distributions available other than the Von Mises distribution and mixture of Von Mises distributions. Again, due to this limitation, we do not pursue any analysis using this approach. We use wrapping approach (see next section) and illustrate how a flexible class of models can be obtained.

1.2.3 Wrapping Approach

In the wrapping approach, given a known distribution on the real line, we wrap it around the circumference of the circle with unit radius. Technically this implies that if U is a random variable on the real line, then the corresponding random variable Y on the circle is given by $Y = U(\text{mod } 2\pi)$. Equivalently, the wrapped version of U is obtained by defining $Y = U - 2\pi \left[\frac{U}{2\pi} \right]$, where $[u] = \text{largest integer } \leq u$. Let the distribution function of U on the real line be denoted by F . The distribution function of Y denoted by F_w can be obtained as,

$$F_w(y) = \Pr(Y \leq y) = \sum_{k=-\infty}^{\infty} [F(y + 2\pi k) - F(2\pi k)].$$

This implies that if the density f of U exists, then the wrapped density f_w is given by

$$f_w(y) = \sum_{k=-\infty}^{\infty} f(y + 2\pi k), 0 \leq y < 2\pi.$$

An excellent overview of the properties of the wrapped distributions can be found in Mardia and Jupp (1999). One of the properties of the wrapped distributions is that the characteristic function (c.f.) of U is same as the c.f. of the Y . Thus, from the information about U , we can obtain the information about Y . More importantly, if the c.f. of U is integrable, then it can be shown that the density of Y , $f_w(y)$ can be represented as,

$$f_w(y) = \frac{1}{2\pi} \left[1 + 2 \sum_{p=1}^{\infty} (\alpha_p \cos py + \beta_p \sin py) \right],$$

where $E(e^{ipY}) = E(e^{ipU}) = \alpha_p + i\beta_p$. This result on the unit circle is analogous to the inversion theorem for continuous random variables on the real line. However, in most cases the above series cannot be written in closed form except in a few cases such as the Cauchy distribution. One of the most popular wrapped distributions is the Wrapped Cauchy distribution, introduced by Lévy (1939). This is because the density of the Wrapped Cauchy distribution has a closed form representation. The density of the Wrapped Cauchy distribution is given by $\sum_{k=-\infty}^{\infty} \frac{1}{\pi\sigma} \left[1 + \left(\frac{y+2\pi k-\mu}{\sigma} \right)^2 \right]^{-1}$, where μ is the location parameter and σ is the scale parameter. Using the inversion theorem, this can be represented as $\frac{1}{2\pi} \{1 + 2 \sum_1^{\infty} (\rho^p \cos p(y - \mu))\}$, where $\rho = e^{-\sigma}$. Considering this density as the real part of the geometric series $\sum_1^{\infty} \rho^p e^{-ip(y-\mu)}$, simplifies it to $\frac{1}{2\pi} \frac{1-\rho^2}{1+\rho^2-2\rho \cos(y-\mu)}$. Kent and Tyler (1988) and Mardia (1972) have shown the relation between the Wrapped Cauchy distribution and the Projected Normal distribution. Mardia and Jupp (1999) describe the wrapped stable family, with density given by $\frac{1}{2\pi} \{1 + 2 \sum_1^{\infty} (\rho^{p\alpha} \cos p(y - \mu))\}$, where $0 < \alpha \leq 2$.

This is a larger family of distributions containing the Wrapped Cauchy distribution ($\alpha = 1$) and the Wrapped Normal distribution ($\alpha = 2$).

It follows that a rich class of distributions on the circle can be obtained using the wrapping technique because we can wrap any known distribution on the real line onto the circle. Also, by wrapping a symmetric unimodal density, it is possible to obtain a bimodal density on the circle. An example is the Wrapped Beta distribution. We discuss this in Section 2.6. The main difficulty in working with the wrapping approach has been that, in most cases, the form of the densities and distribution functions are large sums, and cannot be simplified as closed forms. Due to this complexity, maximum likelihood techniques for point estimation and hypothesis testing cannot be easily implemented. The main contribution of this dissertation is to present a general approach to obtain parameter estimates of a wide class of wrapped distributions. The next paragraph outlines our approach.

For a given probability distribution on the circle, we make assumptions that the original circular distribution was distributed on a line and was wrapped onto the circle. Therefore if we can *unwrap* the distribution on the circle and obtain a distribution on the real line, we can use all the standard statistical techniques for data on the real line. We propose to perform this using the data augmentation approach. We use Bayesian methods, so that we can easily obtain parameter uncertainty estimates based on the finite sample. In many practical problems (as discussed in Section 1.1), the sample sizes are usually small because it is expensive to obtain such samples.

Chapter 2

Parameter Estimation for Wrapped Distributions

In this chapter, we consider parameter estimation for several wrapped distributions. We concentrate on symmetric unimodal and bimodal densities.

2.1 Previous Work

It was shown by Kent and Tyler (1988) that the maximum likelihood estimate for the Wrapped Cauchy distribution exists and is unique for samples of size greater than two. They also gave a simple iterative algorithm, which would always converge to the maximum likelihood estimate (MLE). Calculating the MLE for the Wrapped Cauchy distribution is possible because the density has a closed form representation. In general, wrapped distributions do not have closed form densities and consequently, computing the MLE is complicated.

A different approach for fitting wrapped distributions was given by Fisher and Lee (1994) who used the Expectation Maximization (EM) algorithm techniques to obtain parameter estimates from the Wrapped Normal distribution. However, the E-step involves ratio of large infinite sums, which needs to be approximated at each step. This makes the algorithm computationally inefficient. In addition, the standard errors of the MLEs have to be evaluated based on large-sample theory. We propose an alternative method that is more computationally efficient and flexible to entertain a large class of wrapped distributions. Also, as a by-product, we acquire finite sample interval estimates of the parameters of the wrapped distributions. This is done using the data augmentation approach described in the following section.

2.2 The Data Augmentation Approach

The data augmentation approach was originally proposed by Tanner and Wong (1987). Some references for this technique can also be found in the work of Damien, Wakefield and Walker (1999), Higdon (1998) and van Dyk and Meng (2001) and references therein. In the context of circular data, Damien and Walker (1999) used it to study Von Mises distribution. Coles (1998) used it to study the Wrapped Bivariate Normal distribution and wrapped autoregressive process. We present a generic approach that can be used for a broader class of wrapped distributions.

The main idea behind the data augmentation approach is to augment the original data with some "additional data" that would simplify the original likelihood to a form that is much easier to handle. In case of circular data, as $Y = U(\text{mod } 2\pi)$, the random variable

U on the real line, can be represented as $U = Y + 2\pi K$, where Y is the observed data on the circle, and K is the number of times U was wrapped to obtain Y . Therefore, in this case if we were able to "add" the information on K , and thus *unwrap* Y , then we could observe U . However, given Y , as the value of K is not unique, we obtain the conditional probability distribution of K given Y . To illustrate the unwrapping method, let us consider a location scale family $\frac{1}{\sigma}f(\frac{y-\mu}{\sigma})$ on the real line. Then the corresponding wrapped density is obtained as,

$$\begin{aligned} f_w(y) &= \sum_{k=-\infty}^{\infty} \frac{1}{\sigma} f\left(\frac{y+2\pi k-\mu}{\sigma}\right) \\ &= \sum_{k=-\infty}^{\infty} \frac{\frac{1}{\sigma} f\left(\frac{y+2\pi k-\mu}{\sigma}\right)}{F\left(\frac{2\pi(k+1)-\mu}{\sigma}\right) - F\left(\frac{2\pi k-\mu}{\sigma}\right)} \left[F\left(\frac{2\pi(k+1)-\mu}{\sigma}\right) - F\left(\frac{2\pi k-\mu}{\sigma}\right) \right]. \end{aligned} \quad (2.1)$$

In the above equations, we follow the convention that the location parameter $\mu = \mu_0 \pmod{2\pi}$, where μ_0 is the location parameter on the real line. In order to specify the full probability model, we consider several prior distributions for (μ, σ) . In most cases the mean resultant, ρ (defined in Section 1.1) can be expressed as a function of σ . In general, we will write $\rho = h(\sigma)$. For example, for the Wrapped Normal family, $\rho = e^{-\sigma^2/2}$. Notice that by definition, $0 < \rho \leq 1$. A class of non-informative prior for (μ, ρ) can be specified as,

$$[\mu, \rho] \propto I_{\mu}(0, 2\pi) \rho^{a_{\rho}-1} (1 - \rho)^{a_{\rho}-1}, \quad a_{\rho} > 0, \quad (2.2)$$

and hence the joint density of (μ, σ) is given by

$$[\mu, \sigma] \propto I_{\mu}(0, 2\pi) h(\sigma)^{a_{\rho}-1} (1 - h(\sigma))^{a_{\rho}-1} |h'(\sigma)|, \quad a_{\rho} > 0.$$

Viewing the wrapped number K to be a random variable, from (2.1), we see that

the conditional density of Y given $K = k$ and the parameters μ, ρ is given by

$$f_w(y|k, \mu, \sigma) = \frac{\frac{1}{\sigma} f\left(\frac{y+2\pi k-\mu}{\sigma}\right)}{F\left(\frac{2\pi(k+1)-\mu}{\sigma}\right) - F\left(\frac{2\pi k-\mu}{\sigma}\right)} I_y(0, 2\pi), \quad (2.3)$$

which is a truncated density of a location-scale family. It also follows from (2.1) that the marginal density of K given the parameters (μ, σ) can be obtained as

$$\Pr(K = k|\mu, \sigma) = F\left(\frac{2\pi(k+1)-\mu}{\sigma}\right) - F\left(\frac{2\pi k-\mu}{\sigma}\right), \quad k \in \mathcal{Z}.$$

Thus, we obtain a Bayesian hierarchical model by specifying the distribution of y given k, μ, σ , then the conditional distribution of k given μ, σ and finally the prior distribution for (μ, σ) .

Given a random sample on the circle, $\mathbf{y} = \{y_1, y_2, y_3, \dots, y_n\}$, $0 \leq y_j < 2\pi$, $j = 1 \dots n$, we unwrap the data by obtaining samples from the conditional distribution of \mathbf{k} given μ, σ and the observed data \mathbf{y} , where $\mathbf{k} = \{k_1, k_2, k_3, \dots, k_n\}$. This is referred to as the data augmentation step. Then, conditional on the augmented data \mathbf{y}, \mathbf{k} , we obtain samples from the joint posterior distribution of (μ, σ) , to complete the Gibbs cycle of the MCMC method. From these samples, we obtain the marginal posterior distribution of μ and σ given the observed data \mathbf{y} , using the Ergodic Theorem of Markov Chain.

We provide a generic method to implement the MCMC method for a general class of location scale family, whose density function is invertible. A function $y = f(x)$ is invertible if f can be analytically or numerically inverted, or if f can be factorized into functions, which can be analytically or numerically inverted. That is, we assume that $f(x) = \prod_{t=1}^T f_t(x)$ and $\{f_t^{-1}(y) : j = 1 \dots T\}$ are explicitly known functions or functions that can be computed using numerical methods. For example, $f(x) = e^{-x}e^{-e^{-x}}$ does not

have an explicitly known inverse, but e^{-x} and $e^{-e^{-x}}$ have explicitly known inverses. It is also possible to compute the inverse of $e^{-x}e^{-e^{-x}}$ using numerical methods such as bisection method. It is assumed that $\rho = h(\sigma)$ is a monotone decreasing function of σ and can be inverted. This is a reasonable assumption because in general, as $\sigma \downarrow 0$, $\rho \uparrow 1$ and as $\sigma \uparrow \infty$, $\rho \downarrow 0$. These assumptions are satisfied for most wrapped distributions including Wrapped Normal (WN), Wrapped Cauchy (WC) and Wrapped Double Exponential (WDE) distributions. The general method will work even if $\rho = h(\sigma)$ is not a monotone decreasing function of σ . These wrapped distributions are all symmetric and unimodal. However, our method is not restricted only to these distributions.

In order to specify the required conditional distributions, we use the notation $[\theta_1, \theta_2, \dots, \theta_n]$ to represent the joint density of $\theta_1, \theta_2, \dots, \theta_n$ and $[\theta_1 | \theta_2, \dots, \theta_n]$ to represent the conditional density of θ_1 given $\theta_2, \dots, \theta_n$. By the term *full conditional density of θ_1* , we mean the conditional density of θ_1 , given the rest of the parameters. We use the notation $a \vee b$ to represent $\max(a, b)$ and $a \wedge b$ to represent $\min(a, b)$. The notation $DU[p, q]$ stands for the Discrete Uniform distribution on $[p, q]$, $p \in \mathcal{Z}$, $q \in \mathcal{Z}$, $p < q$. That is, if $U \sim DU[p, q]$, then the p.d.f. of U is given by

$$f_U(u) = \begin{cases} \frac{1}{q-p+1} & u = p, p+1, \dots, q \\ 0 & \text{otherwise.} \end{cases}$$

The general method is as follows. As mentioned above, the joint density of (μ, σ) is given by

$$[\mu, \sigma] \propto I_\mu(0, 2\pi) h(\sigma)^{a_\rho-1} (1 - h(\sigma))^{a_\rho-1} |h'(\sigma)|, \quad a_\rho > 0, \quad \rho = h(\sigma).$$

For each observed direction y_j , we augment a random wrapping number k_j . The joint

density of $\mathbf{y}, \mathbf{k}, \mu$ and σ is given by

$$\begin{aligned}
[\mathbf{y}, \mathbf{k}, \mu, \sigma] &\propto [\mathbf{y}|\mathbf{k}, \mu, \sigma^2] [\mathbf{k}|\mu, \sigma^2] [\mu, \sigma^2] \\
&\propto \prod_{j=1}^n \left(\frac{1}{\sigma} f\left(\frac{y_j - \mu + 2\pi k_j}{\sigma}\right) \right) h(\sigma)^{a_\rho - 1} (1 - h(\sigma))^{a_\rho - 1} |h'(\sigma)| I_\mu(0, 2\pi) \\
&\propto \frac{1}{\sigma^{n+n_0}} \prod_{j=1}^n f\left(\frac{y_j - \mu + 2\pi k_j}{\sigma}\right) h(\sigma)^{a_\rho - 1 + n_1} (1 - h(\sigma))^{a_\rho - 1} h_1(\sigma) I_\mu(0, 2\pi),
\end{aligned}$$

where $|h'(\sigma)|$ can be factorized as $\frac{1}{\sigma^{n_0}} h(\sigma)^{n_1} h_1(\sigma)$. It is assumed that $h_1(\sigma)$ is invertible. In our examples, $|h'(\sigma)|$ factorizes as $\frac{1}{\sigma^{n_0}} h(\sigma)^{n_1}$ with no extra $h_1(\sigma)$ term. In general, this is not true. This general technique assumes that there is an $h_1(\sigma)$ term. Also, in general, $h(\sigma)^{a_\rho - 1}$ and $(1 - h(\sigma))^{a_\rho - 1}$ cannot be factorized. However, we see that for the Wrapped Double Exponential distribution, $(1 - h(\sigma))^{a_\rho - 1}$ can be factorized as a function of $h(\sigma)$ and σ .

The full conditional densities of \mathbf{k}, μ and σ are nonstandard densities. Therefore, we introduce auxiliary variables \mathbf{z} and \mathbf{v} . Let $\mathbf{z} = \{z_0, z_1, z_2, z_3\}$ and $\mathbf{v} = \{v_1, v_2, v_3, \dots, v_n\}$, such that

$$[\mathbf{y}, \mathbf{k}, \mu, \sigma] \propto \int [\mathbf{y}, \mathbf{k}, \mu, \mathbf{z}, \mathbf{v}, \sigma] d\mathbf{z} d\mathbf{v}.$$

The joint density of $\mathbf{y}, \mathbf{k}, \mu, \mathbf{z}, \mathbf{v}$ and σ is given by

$$\begin{aligned}
[\mathbf{y}, \mathbf{k}, \mu, \mathbf{z}, \mathbf{v}, \sigma] &\propto I_{z_0}\left(0, \frac{1}{\sigma^{n+n_0}}\right) \prod_{j=1}^n I_{v_j}\left(0, f\left(\frac{y_j - \mu + 2\pi k_j}{\sigma}\right)\right) I_{z_1}\left(0, h(\sigma)^{a_\rho - 1 + n_1}\right) I_\mu(0, 2\pi) \\
&\quad \{I_{z_2}\left(0, (1 - h(\sigma))^{a_\rho - 1}\right) I(a_\rho \neq 1) + I(a_\rho = 1)\} I_{z_3}(0, h_1(\sigma)).
\end{aligned}$$

This data augmentation method, where we introduce auxiliary variables, is called slice sampling. The advantage of slice sampling method over the Metropolis-Hastings method is that we don't have to choose a proposal density. Furthermore, it has been shown that slice sampling method increases mixing.

We now show that all full conditional densities are standard distributions, which can be easily sampled using subroutines available in SAS, Splus and other standard software applications. In general, let f be a multimodal density, which is invertible. Then, given v_j , we get $\{(gm_l(v_j), gM_l(v_j)), l = 1 \dots q, q \in \mathcal{Z}\}$ such that,

$$\left\{0 < v_j < f\left(\frac{y_j - \mu + 2\pi k_j}{\sigma}\right)\right\} = \bigcup_{l=1}^q \left\{gm_l(v_j) < \left(\frac{y_j - \mu + 2\pi k_j}{\sigma}\right) < gM_l(v_j)\right\}$$

is a union of disjoint sets because $gm_1(v_j) < gM_1(v_j) < gm_2(v_j) < gM_2(v_j) < \dots < gm_q(v_j) < gM_q(v_j)$.

If f is not easily invertible, then it can be factorized into functions, which are analytically or numerically invertible. That is, $f(x) = \prod_{t=1}^T f_t(x)$ and $\{f_t^{-1}(y) : j = 1 \dots T\}$ can be computed. In this case, instead of $\prod_{j=1}^n I_{v_j}\left(0, f\left(\frac{y_j - \mu + 2\pi k_j}{\sigma}\right)\right)$, use $\prod_{t=1}^T \prod_{j=1}^n I_{v_{tj}}\left(0, f_t\left(\frac{y_j - \mu + 2\pi k_j}{\sigma}\right)\right)$ in the expression for joint density of $\mathbf{y}, \mathbf{k}, \mu, \mathbf{z}, \mathbf{v}$ and σ . Use a similar procedure as above for all $\{v_{tj} : t = 1 \dots T, j = 1 \dots n\}$.

The full conditional densities of $\mathbf{k}, \mathbf{v}, \mathbf{z}$ and μ are given by

$$\begin{aligned} v_j | \mathbf{y}, \mathbf{k}, \mu, \mathbf{z}, \mathbf{v}_{-j}, \sigma &\sim U\left[0, f\left(\frac{y_j - \mu + 2\pi k_j}{\sigma}\right)\right] \\ k_j | \mathbf{y}, \mathbf{k}_{-j}, \mu, \mathbf{z}, \mathbf{v}, \sigma &\sim DU\left[\bigcup_{l=1}^q \left\{\left\lceil \frac{1}{2\pi}(\mu - y_j + \sigma gm_l(v_j)) \right\rceil, \left\lfloor \frac{1}{2\pi}(\mu - y_j + \sigma gM_l(v_j)) \right\rfloor\right\}\right], \end{aligned}$$

where DU stands for Discrete Uniform.

Without loss of generality, let $l_o \in \{1, 2, \dots, q\}$ be such that,

$$\left\lceil \frac{1}{2\pi}(\mu - y_j + \sigma gm_{l_o}(v_j)) \right\rceil \leq k_j \leq \left\lfloor \frac{1}{2\pi}(\mu - y_j + \sigma gM_{l_o}(v_j)) \right\rfloor$$

$$z_0 | \mathbf{y}, \mathbf{k}, \mu, \mathbf{z}_{-0}, \mathbf{v}, \sigma \sim U\left[0, \frac{1}{\sigma^{n+n_0}}\right]$$

$$z_1 | \mathbf{y}, \mathbf{k}, \mu, \mathbf{z}_{-1}, \mathbf{v}, \sigma \sim U\left[0, h(\sigma)^{a_\rho - 1 + n_1}\right]$$

$$\begin{aligned}
z_2|\mathbf{y}, \mathbf{k}, \mu, \mathbf{z}_{-2}, \mathbf{v}, \sigma &\sim \begin{cases} U [0, (1 - h(\sigma))^{a_\rho - 1}], & a_\rho \neq 1 \\ z_2 \text{ is not needed, } & a_\rho = 1 \end{cases} \\
z_3|\mathbf{y}, \mathbf{k}, \mu, \mathbf{z}_{-3}, \mathbf{v}, \sigma &\sim \begin{cases} U [0, h_1(\sigma)], & h_1(\sigma) \neq 1 \\ z_3 \text{ is not needed, } & h_1(\sigma) \equiv 1 \end{cases} \\
\mu|\mathbf{y}, \mathbf{k}, \mathbf{z}, \mathbf{v}, \sigma &\sim U [m_\mu, M_\mu],
\end{aligned}$$

$$\text{where } m_\mu = \max_{j=1}^n [y_j + 2\pi k_j - \sigma g M_{l_0}(v_j)] \bigvee 0$$

$$M_\mu = \min_{j=1}^n [y_j + 2\pi k_j - \sigma g m_{l_0}(v_j)] \bigwedge (2\pi).$$

Since $h(\sigma)$ is a monotone decreasing function of σ and is invertible, h^{-1} is also a monotone function. This enables us invert the relationship between z_1 , z_2 and σ and obtain some parts of the distribution for σ .

Also, since $h_1(\sigma)$ is invertible, given z_3 , we get $\{(gm_l^*(z_3), gM_l^*(z_3)), l = 1 \dots q, q \in \mathcal{Z}\}$ such that, $\{0 < z_3 < h_1(\sigma)\} = \bigcup_{l=1}^q \{gm_l^*(z_3) < \sigma < gM_l^*(z_3)\}$ is a union of disjoint sets because $gm_1^*(z_3) < gM_1^*(z_3) < gm_2^*(z_3) < gM_2^*(z_3) < \dots < gm_q^*(z_3) < gM_q^*(z_3)$. If $h_1(\sigma)$ is not invertible, but can be factorized into invertible functions, then this method can also be easily extended in a manner similar to the extension given for f .

$$\begin{aligned}
\sigma|\mathbf{y}, \mathbf{k}, \mu, \mathbf{z}, \mathbf{v} &\sim U \left[(m_\sigma, M_\sigma) \bigcap \bigcup_{l=1}^q \{gm_l^*(z_3), gM_l^*(z_3)\} \right], \\
\text{where } m_\sigma &= \begin{cases} \max_{j=1}^n h_m \left(\frac{y_j - \mu + 2\pi k_j}{gm_{l_0}(v_j)}, \frac{y_j - \mu + 2\pi k_j}{gM_{l_0}(v_j)} \right) \bigvee h^{-1} \left(1 - z_2^{\frac{1}{a_\rho - 1}} \right), & a_\rho > 1 \\ \max_{j=1}^n h_m \left(\frac{y_j - \mu + 2\pi k_j}{gm_{l_0}(v_j)}, \frac{y_j - \mu + 2\pi k_j}{gM_{l_0}(v_j)} \right), & a_\rho \leq 1 \end{cases}
\end{aligned}$$

$$M_\sigma = \begin{cases} \min_{j=1}^n h_M \left(\frac{y_j - \mu + 2\pi k_j}{gm_{l_0}(v_j)}, \frac{y_j - \mu + 2\pi k_j}{gM_{l_0}(v_j)} \right) \wedge \frac{1}{z_0^{\frac{1}{n+n_0}}} \wedge h^{-1} \left(z_1^{\frac{1}{a_\rho - 1 + n_1}} \right), \\ \text{where } (a_\rho \geq 1) \cup ((a_\rho < 1) \cap (z_2 \leq 1)) \\ \left[\min_{j=1}^n h_M \left(\frac{y_j - \mu + 2\pi k_j}{gm_{l_0}(v_j)}, \frac{y_j - \mu + 2\pi k_j}{gM_{l_0}(v_j)} \right) \wedge \frac{1}{z_0^{\frac{1}{n+n_0}}} \wedge h^{-1} \left(z_1^{\frac{1}{a_\rho - 1 + n_1}} \right) \right. \\ \left. \wedge h^{-1} \left(1 - \frac{1}{z_2^{\frac{1}{1-a_\rho}}} \right) \right], \text{ where } ((a_\rho < 1) \cap (z_2 > 1)), \end{cases}$$

where $h_m \left(\frac{y_j - \mu + 2\pi k_j}{gm_{l_0}(v_j)}, \frac{y_j - \mu + 2\pi k_j}{gM_{l_0}(v_j)} \right)$ and $h_M \left(\frac{y_j - \mu + 2\pi k_j}{gm_{l_0}(v_j)}, \frac{y_j - \mu + 2\pi k_j}{gM_{l_0}(v_j)} \right)$ are defined as follows.

$$h_m \left(\frac{y_j - \mu + 2\pi k_j}{gm_{l_0}(v_j)}, \frac{y_j - \mu + 2\pi k_j}{gM_{l_0}(v_j)} \right) = \begin{cases} \frac{y_j - \mu + 2\pi k_j}{gM_{l_0}(v_j)}, & gm_{l_0}(v_j) \geq 0 \\ \frac{y_j - \mu + 2\pi k_j}{gm_{l_0}(v_j)}, & gM_{l_0}(v_j) \leq 0 \\ \frac{y_j - \mu + 2\pi k_j}{gm_{l_0}(v_j)} \vee \frac{y_j - \mu + 2\pi k_j}{gM_{l_0}(v_j)}, & gm_{l_0}(v_j) < 0 < gM_{l_0}(v_j) \end{cases}$$

$$h_M \left(\frac{y_j - \mu + 2\pi k_j}{gm_{l_0}(v_j)}, \frac{y_j - \mu + 2\pi k_j}{gM_{l_0}(v_j)} \right) = \begin{cases} \frac{y_j - \mu + 2\pi k_j}{gm_{l_0}(v_j)}, & gm_{l_0}(v_j) \geq 0 \\ \frac{y_j - \mu + 2\pi k_j}{gM_{l_0}(v_j)}, & gM_{l_0}(v_j) \leq 0 \\ h_M \text{ is not needed, } & gm_{l_0}(v_j) < 0 < gM_{l_0}(v_j). \end{cases}$$

For symmetric unimodal densities,

$$h_m \left(\frac{y_j - \mu + 2\pi k_j}{gm_{l_0}(v_j)}, \frac{y_j - \mu + 2\pi k_j}{gM_{l_0}(v_j)} \right) = \frac{|y_j - \mu + 2\pi k_j|}{gM_{l_0}(v_j)}$$

and h_M is not needed.

We illustrate the above technique for several popular distributions such as Wrapped Normal, Wrapped Cauchy and Wrapped Double Exponential distributions. In Appendix A, we provide more details on the exact form of the above full conditional distributions for these three families of wrapped distributions. We study the performance of the proposed method by several simulation experiments.

2.3 Model selection

We now have a large class of parametric models that can be fitted to the circular data. In order to select the best fitting model, good model selection techniques are required. A few methods for model selection are Deviance, Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC) and Gelfand and Ghosh (1998) Criteria (GGC).

Deviance (McCullagh and Nelder, 1989) is defined as twice the negative of the log-likelihood. For example, for any wrapped location-scale density, $f_w()$ the deviance (Dev) is,

$$\begin{aligned} \text{Dev} &= -2\log\left(\prod_{i=1}^n f_w(y_i)\right) \\ &= -2\sum_{i=1}^n \log\left(\sum_{k=-\infty}^{\infty} \frac{1}{\sigma} f\left(\frac{y_i+2\pi k-\mu}{\sigma}\right)\right) \\ &\approx -2\sum_{i=1}^n \log\left(\sum_{k=-L}^L \frac{1}{\sigma} f\left(\frac{y_i+2\pi k-\mu}{\sigma}\right)\right), \end{aligned}$$

where L is a very large positive number. For Wrapped Normal and Wrapped Double Exponential densities, $L = 5\sigma$ and $L = 10\sigma$ respectively, works well. We could not use this approximation technique for the Wrapped Cauchy density as Cauchy is a flat density. For the Wrapped Cauchy density, we used $L = 100\sigma$ and approximated the tail with the integral. We could have also used the closed form for the Wrapped Cauchy density.

AIC (Akaike, 1973) and BIC (Schwartz, 1978) can be calculated by adding an appropriate penalty term to the posterior mean of the deviance. This penalty term is a function of the dimension of the parameters and sample size, which are same for all the wrapped densities that we have considered. That is,

$$AIC = Dev + 2m,$$

$$BIC = Dev + m \log(n),$$

where $m = 2$ is the number of parameters (μ and σ (or ρ)) and n is the sample size. Therefore comparing AIC or BIC is the same as comparing the deviances. Therefore, we just use the Dev to select the best fitting model.

Gelfand and Ghosh Criteria (GGC) is based on posterior predictive distribution. Define $y^{obs} = (y_1, \dots, y_n)$ as the observed data and $y^{pred} = (y_1^{pred}, \dots, y_n^{pred})$ as the predictive data obtained from the following posterior predictive distribution,

$$p(y^{pred}|y^{obs}) = \int \int p(y^{pred}|\mu, \rho)p(\mu, \rho|y^{obs})d\mu d\rho,$$

where $p(y^{pred}|\mu, \rho)$ denotes the sampling distribution of the data, which is the wrapped density evaluated y^{pred} given μ and ρ . $p(\mu, \rho|y^{obs})$ denotes the posterior distribution of the parameters (μ, ρ) given the observed data y^{obs} . Let the loss function be the Square Predicted Errors (SPE) function defined as,

$$SPE = \sum_{i=1}^n (y_i^{pred} - y_i)^2.$$

GGC is defined as,

$$\begin{aligned} GGC &= E[SPE|y^{obs}] \\ &= G + P, \end{aligned}$$

where $G = \sum_{i=1}^n (y_i - E(y_i^{pred}|y^{obs}))^2$ and $P = \sum_{i=1}^n Var(y_i^{pred}|y^{obs})$. G is the goodness-of-fit term and P is the penalty term. The expectation is taken with respect to the posterior predictive distribution defined above. As models become more complex, the G term usually decreases and the P term increases to account for model complexity. Thus, this criteria

selects a model that compromises between G and P . We use the criteria to choose a model that minimizes GGC .

2.4 Simulation studies

For our simulation studies, we generate samples of size $n = 50$ from Wrapped Normal (WN), Wrapped Cauchy (WC) and Wrapped Double Exponential (WDE) distributions with parameters set at $\mu = \pi/2 = 1.5708$ and $\rho = 0.5$. In order to study the sensitivity of priors, we fit each of the three distributions with priors given in equation (2.2) with $a_\rho = 0.5, 1$ and 2 . We compute the posterior mean, standard deviation, 2.5 percentile, median and 97.5 percentile for μ and ρ for each simulation. The percentiles are computed with 0 radians as the reference point on the circle. Usually, the reference point on the circle is chosen diametrically opposite the sample mean, or where the samples are sparsely distributed. Note that, changing the reference point does not affect the circular mean. Bayesian highest posterior density (HPD) can be used instead of percentiles. HPD is unaffected by the change in the reference point. This is described in more detail in Section 2.7.1. The simulation standard errors for each of these summary values are also computed. We also compute the nominal coverage probability for the 95% posterior interval given by the 2.5 and 97.5 percentile of the posterior distribution. In each simulation, we choose the burn-in period to be 2000 samples (i.e. throw away first 2000 samples from the MCMC chain) and then keep 5000 samples after burn-in, to obtain posterior summary values. The sample size was decided after some preliminary studies using Geweke diagnostics, Gelman-Rubin diagnostics, Raftery-Lewis method, autocorrelations, cross-correlations and

Table 2.1: Fitting WN to WN distribution using different priors

$a_\rho = 0.5$	mean	std dev	2.5 %	50 %	97.5 %	cov prob
μ	1.58	0.20	1.18	1.58	1.98	0.96
s.e.	0.20	0.09	0.22	0.19	0.29	0.01
ρ	0.48	0.08	0.31	0.48	0.61	0.94
s.e.	0.08	0.01	0.10	0.08	0.06	0.01
$a_\rho = 1$	mean	std dev	2.5 %	50 %	97.5 %	cov prob
μ	1.59	0.20	1.20	1.59	1.98	0.96
s.e.	0.19	0.04	0.21	0.19	0.21	0.01
ρ	0.48	0.07	0.32	0.49	0.61	0.92
s.e.	0.08	0.01	0.09	0.08	0.06	0.01
$a_\rho = 2$	mean	std dev	2.5 %	50 %	97.5 %	cov prob
μ	1.57	0.19	1.18	1.57	1.95	0.94
s.e.	0.20	0.04	0.22	0.20	0.21	0.01
ρ	0.49	0.07	0.33	0.49	0.61	0.94
s.e.	0.07	0.01	0.08	0.07	0.06	0.01

trace plots. This was done using the CODA program. Tables 2.1 through 2.7 contain all summary values based on these final 5000 samples. We repeat the entire procedure 500 times to see the frequentist performance of the proposed Bayes method. In SAS, on a sparc 20 machine, on an average it took about 50 minutes to perform the entire simulation for a given wrapped distribution.

Bayesian highest posterior density (HPD) can also be used instead of percentiles. For example, in Figure 2.7, we plot the posterior density of μ while fitting WC to Ant data. Since the posterior density has most its mass close to 3.24 radians (185°), we computed the percentiles with 0 radians as the reference point. The percentiles will change if we change the reference point, but the change will be minimal for the Jander's ant data set, if the reference point is away from 3.24 radians. However, HPD is not affected by the change in the reference point.

Table 2.2: Fitting WC to WC distribution using different priors

$a_\rho = 0.5$	mean	std dev	2.5 %	50 %	97.5 %	cov prob
μ	1.57	0.17	1.23	1.57	1.92	0.95
s.e.	0.18	0.14	0.20	0.17	0.34	0.01
ρ	0.49	0.07	0.34	0.49	0.61	0.90
s.e.	0.09	0.02	0.10	0.09	0.08	0.01
$a_\rho = 1$	mean	std dev	2.5 %	50 %	97.5 %	cov prob
μ	1.56	0.17	1.23	1.56	1.90	0.93
s.e.	0.17	0.05	0.20	0.17	0.19	0.01
ρ	0.48	0.07	0.33	0.49	0.61	0.94
s.e.	0.07	0.01	0.09	0.07	0.06	0.01
$a_\rho = 2$	mean	std dev	2.5 %	50 %	97.5 %	cov prob
μ	1.57	0.17	1.25	1.57	1.91	0.95
s.e.	0.17	0.13	0.20	0.16	0.34	0.01
ρ	0.49	0.07	0.35	0.49	0.62	0.92
s.e.	0.08	0.01	0.09	0.08	0.07	0.01

Table 2.3: Fitting WDE to WDE distribution using different priors

$a_\rho = 0.5$	mean	std dev	2.5 %	50 %	97.5 %	cov prob
μ	1.56	0.18	1.21	1.56	1.92	0.96
s.e.	0.17	0.07	0.20	0.17	0.23	0.01
ρ	0.48	0.08	0.31	0.48	0.63	0.91
s.e.	0.09	0.01	0.10	0.09	0.08	0.01
$a_\rho = 1$	mean	std dev	2.5 %	50 %	97.5 %	cov prob
μ	1.58	0.17	1.24	1.58	1.92	0.97
s.e.	0.17	0.06	0.20	0.17	0.24	0.01
ρ	0.49	0.08	0.33	0.49	0.63	0.90
s.e.	0.09	0.01	0.10	0.09	0.08	0.01
$a_\rho = 2$	mean	std dev	2.5 %	50 %	97.5 %	cov prob
μ	1.58	0.18	1.23	1.58	1.92	0.93
s.e.	0.17	0.05	0.19	0.17	0.21	0.01
ρ	0.49	0.08	0.33	0.49	0.63	0.92
s.e.	0.08	0.01	0.09	0.08	0.07	0.01

From Table 2.1, Table 2.2 and Table 2.3, we see that the proposed method performs very well in terms of maintaining the nominal coverage probability when the underlying distribution is true. In addition, the posterior mean and median can serve as good point estimates of the parameters. Although we do not see that the posterior distribution is sensitive to the choice of the hyper parameter a_ρ , in general, we would recommend $a_\rho = 0.5$ for all the wrapped distributions. Therefore, we use this value for a_ρ for our application and other simulations.

In order to study the sensitivity of the sampling distribution, we generated data from Normal, Cauchy and Double Exponential distributions on the real line and wrapped them onto the circle $(0, 2\pi)$. For our simulations, we fixed $\mu = \pi/2$ and $\rho = 0.5$. We generated $n = 50$ observations from the $WN(\pi/2, 0.5)$, $WC(\pi/2, 0.5)$ and $WDE(\pi/2, 0.5)$. We then fitted Wrapped Normal, Wrapped Cauchy and Wrapped Double Exponential distributions to each of the three datasets. We repeated the method 500 times to see the frequentist performance of the Bayes method for erroneous models. For model fitting we used $a_\rho = 0.5$ for all priors. In our study (as shown previously in Table 2.1, Table 2.2 and Table 2.3) we did not find the posterior summary to be very sensitive to the choice of a_ρ . Therefore, we did not report the posterior summary values for other choices of a_ρ . As before, we report the posterior mean, standard deviation, and two equal tail percentiles along with the Monte Carlo standard error. Table 2.4, Table 2.5 and Table 2.6 contain the summary statistics.

Comparing the results in Table 2.4, Table 2.5 and Table 2.6, we see that the model selection criteria GGC and deviance work well and select the right distribution. A brief description of GGC and Dev is given in Section 2.3. While fitting the WN, WC and

Table 2.4: Fitting WN, WC and WDE to WN distribution with $\rho \sim Beta(0.5, 0.5)$

<i>WN</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
μ	1.56	0.20	1.17	1.56	1.95	0.95
s.e.	0.19	0.05	0.21	0.19	0.22	0.01
ρ	0.48	0.08	0.32	0.49	0.61	0.94
s.e.	0.08	0.01	0.10	0.08	0.06	0.01
GGC	228.88	47.90	146.15	225.20	332.62	
s.e.	48.48	6.41	36.13	48.99	58.81	
Dev	156.45	2.14	154.36	155.79	162.28	
s.e.	8.77	0.32	8.71	8.75	9.04	
<i>WC</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
μ	1.56	0.23	1.12	1.56	2.03	0.90
s.e.	0.24	0.13	0.27	0.24	0.47	0.01
ρ	0.42	0.07	0.27	0.42	0.56	0.81
s.e.	0.08	0.02	0.09	0.08	0.07	0.02
GGC	241.26	47.78	155.66	238.58	342.14	
s.e.	41.53	5.43	31.98	41.71	50.77	
Dev	160.30	2.07	158.18	159.69	165.75	
s.e.	8.08	0.52	7.93	8.10	8.31	
<i>WDE</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
μ	1.56	0.21	1.16	1.56	1.97	0.91
s.e.	0.23	0.07	0.26	0.23	0.29	0.01
ρ	0.47	0.08	0.30	0.47	0.62	0.94
s.e.	0.08	0.01	0.09	0.09	0.07	0.01
GGC	230.78	47.36	147.43	227.66	331.92	
s.e.	45.39	5.78	35.03	45.76	54.49	
Dev	158.93	2.05	156.82	158.32	164.37	
s.e.	8.75	0.43	8.68	8.77	8.82	

Table 2.5: Fitting WN, WC and WDE to WC distribution with $\rho \sim Beta(0.5, 0.5)$

<i>WN</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
μ	1.57	0.23	1.11	1.56	2.04	0.94
s.e.	0.22	0.10	0.25	0.22	0.36	0.01
ρ	0.45	0.08	0.29	0.46	0.59	0.87
s.e.	0.09	0.01	0.10	0.09	0.07	0.02
GGC	232.80	48.21	148.73	229.32	336.40	
s.e.	47.06	5.97	35.22	47.64	56.38	
Dev	158.93	2.10	156.84	158.29	164.62	
s.e.	9.51	0.32	9.46	9.51	9.65	
<i>WC</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
μ	1.57	0.17	1.23	1.57	1.92	0.95
s.e.	0.17	0.11	0.21	0.16	0.34	0.01
ρ	0.48	0.07	0.33	0.49	0.62	0.89
s.e.	0.09	0.02	0.10	0.09	0.07	0.01
GGC	222.66	43.76	144.79	220.00	315.54	
s.e.	40.01	5.06	30.97	40.15	48.86	
Dev	154.76	2.07	152.66	154.14	160.27	
s.e.	10.71	0.44	10.60	10.72	10.91	
<i>WDE</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
μ	1.57	0.18	1.23	1.57	1.92	0.95
s.e.	0.17	0.11	0.20	0.16	0.34	0.01
ρ	0.49	0.08	0.33	0.50	0.64	0.92
s.e.	0.09	0.01	0.10	0.09	0.08	0.01
GGC	221.19	44.46	142.90	218.25	316.11	
s.e.	42.38	4.95	32.96	42.76	50.35	
Dev	154.95	2.05	152.87	154.34	160.46	
s.e.	10.46	0.42	10.40	10.45	10.57	

Table 2.6: Fitting WN, WC and WDE to WDE distribution with $\rho \sim Beta(0.5, 0.5)$

<i>WN</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
μ	1.58	0.22	1.14	1.58	2.02	0.96
s.e.	0.21	0.09	0.24	0.21	0.31	0.01
ρ	0.46	0.08	0.29	0.46	0.59	0.89
s.e.	0.09	0.01	0.10	0.09	0.07	0.01
GGC	229.32	47.88	146.09	225.76	332.56	
s.e.	45.60	6.24	33.76	46.08	55.58	
Dev	158.08	2.13	155.99	157.43	163.88	
s.e.	9.45	0.34	9.40	9.43	9.70	
<i>WC</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
μ	1.56	0.18	1.22	1.56	1.92	0.96
s.e.	0.17	0.05	0.21	0.17	0.19	0.01
ρ	0.47	0.07	0.32	0.47	0.60	0.92
s.e.	0.08	0.01	0.09	0.08	0.07	0.01
GGC	225.45	44.73	145.81	222.74	320.27	
s.e.	37.60	5.07	28.91	37.68	46.19	
Dev	156.09	2.09	153.97	155.45	161.63	
s.e.	9.32	0.44	9.25	9.30	9.55	
<i>WDE</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
μ	1.57	0.17	1.22	1.57	1.91	0.96
s.e.	0.16	0.05	0.20	0.17	0.19	0.01
ρ	0.49	0.08	0.32	0.49	0.64	0.92
s.e.	0.08	0.01	0.10	0.08	0.07	0.01
GGC	221.53	44.91	142.45	218.56	317.35	
s.e.	40.95	5.16	31.56	41.31	49.06	
Dev	155.61	2.09	153.50	154.98	161.21	
s.e.	9.62	0.47	9.54	9.60	9.79	

WDE distributions, the model generated from the true distribution has the lowest GGC and Dev value. Also, both μ and ρ have been well estimated, as we have fitted the correct distribution.

From Table 2.4, Table 2.5 and Table 2.6, we see that in general, the location parameter is estimated well, even when the models are erroneous. For WN model, coverage probability of ρ is less than the nominal level. This indicates that WN is not robust in estimating the mean resultant length when the distribution is not specified correctly. In the case of fitting WC model to WN data, we see that the mean resultant is not well estimated. However, WC model performs considerably better while fitting WDE data. This is expected as WC is closer to WDE than WN. Finally, we see that for WDE model, the parameters μ and ρ have been well estimated, even when the distribution is incorrect. Therefore, WDE model is robust in estimating μ and ρ .

We generate data from Von Mises (VM) distribution on the circle $(0, 2\pi)$. To generate the data from Von Mises distribution, we use the algorithm given by Best and Fisher (1978). As before we set the $\mu = \pi/2$ and $\rho = 0.5$. We fit Wrapped Normal, Wrapped Cauchy and Wrapped Double Exponential distributions to Von Mises data and repeat the procedure 500 times to study frequentist performance.

In Table 2.7, we see that WN is robust in estimating μ and ρ when they come from a Von Mises distribution. This is expected as WN closely approximates the Von Mises distribution. WC model does not perform well in estimating the mean resultant length. The lower than nominal coverage probability of ρ indicates that WC is not robust in estimating the mean resultant length of the distribution. However, the location parameter is estimated

Table 2.7: Fitting WN, WC and WDE to VM data with $\rho \sim \text{Beta}(0.5, 0.5)$

<i>WN</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
μ	1.57	0.22	1.15	1.57	2.00	0.95
s.e.	0.20	0.10	0.23	0.20	0.35	0.01
ρ	0.47	0.08	0.31	0.47	0.60	0.91
s.e.	0.08	0.01	0.10	0.08	0.07	0.01
GGC	228.56	47.84	145.67	224.93	331.84	
s.e.	47.59	6.21	35.56	48.14	57.36	
Dev	157.36	2.10	155.29	156.72	163.06	
s.e.	9.32	0.33	9.25	9.29	9.54	
<i>WC</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
μ	1.57	0.20	1.18	1.57	1.97	0.94
s.e.	0.20	0.06	0.23	0.20	0.24	0.01
ρ	0.44	0.07	0.29	0.45	0.58	0.87
s.e.	0.07	0.01	0.09	0.07	0.07	0.02
GGC	232.33	46.21	149.83	229.60	330.11	
s.e.	38.81	5.13	30.09	38.91	47.82	
Dev	158.30	2.07	156.21	157.67	163.79	
s.e.	8.69	0.42	8.63	8.68	8.88	
<i>WDE</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
μ	1.58	0.20	1.19	1.57	1.97	0.92
s.e.	0.21	0.11	0.22	0.20	0.34	0.01
ρ	0.48	0.08	0.31	0.48	0.63	0.93
s.e.	0.08	0.01	0.10	0.08	0.07	0.01
GGC	220.16	40.54	146.23	218.27	304.63	
s.e.	26.06	2.07	22.78	26.29	28.54	
Dev	157.43	2.09	155.30	156.81	163.03	
s.e.	9.23	0.46	9.17	9.23	9.30	

very well. For the WDE model, the parameters μ and ρ have been well estimated. Therefore, we see again that the WDE model is robust in estimating μ and ρ when the distribution is erroneous. Using GGC, we find that the WDE model fits the best to VM. However, Dev indicates that WDE model and WN model fit well to VM.

2.5 Validation approach for MCMC

In this section, we implement the method suggested by Monahan and Boos (1992) for validation based on coverage probabilities. Consider any nonnegative integrable likelihood, $L(y|t)$. Let the underlying distribution for parameter θ be any absolutely continuous prior $p(t)$ and let $f(y|t)$ be the correct model for the data. R_α is defined as the posterior coverage set function of level α for a prior p and likelihood L if, and only if, for every y , $P[\theta \in R_\alpha(y)] = \alpha$ under conditional measure on θ .

Define H to be the random integral,

$$H = \int_{-\infty}^{\theta} p_*(t|Y) dt,$$

where $p_*(t|y) \propto p(t)L(y|t)$ is the posterior. If the posterior $p_*(t|y)$ is valid by coverage, then H is $U(0,1)$ under $p(t)f(y|t)$ on (Y, θ) . Posterior $p_*(t|y)$ is defined to be valid by coverage for a model $f(y|t)$ if, and only if, for every posterior coverage set function $R_\alpha(y)$ for that prior p and likelihood L , $P[\theta \in R_\alpha(y)] = \alpha$ under the joint measure $p(t)f(y|t)$ on (Y, θ) .

Using this result, the procedure is as follows. We generate a sample $\{\theta_k, k = 1, \dots, m$ independently from the prior, $p(t)$. Then, for each θ_k , generate data $Y^{(k)}$ from

$f(y|\theta_k)$, and compute H_k for sample $Y^{(k)}$ given by

$$H_k = \int_{-\infty}^{\theta_k} p_*(t|Y^{(k)}) dt \approx \frac{1}{B-b} \sum_{l=b+1}^B I(\theta_k^{(l)} \leq \theta_k),$$

where $\{\theta_k^{(l)}, l = 1, \dots, B\}$ is the MCMC chain generated for θ_k , b is the burn-in (initial samples rejected) and B is the total length of the chain. This sample $\{H_k, k = 1, \dots, m\}$ can be tested for goodness of fit to the $U(0, 1)$.

If the distribution of H_k fails to follow the Uniform distribution for any prior, then the likelihood L cannot be a coverage proper Bayesian likelihood. A likelihood L is defined to be a coverage proper Bayesian likelihood if, and only if, for every absolutely continuous prior $p(t)$, the posterior $p_*(t|y) \propto p(t)L(y|t)$ is valid by coverage. If L cannot be a coverage proper Bayesian likelihood, then L cannot be a proper Bayesian likelihood and will not give valid posterior inference.

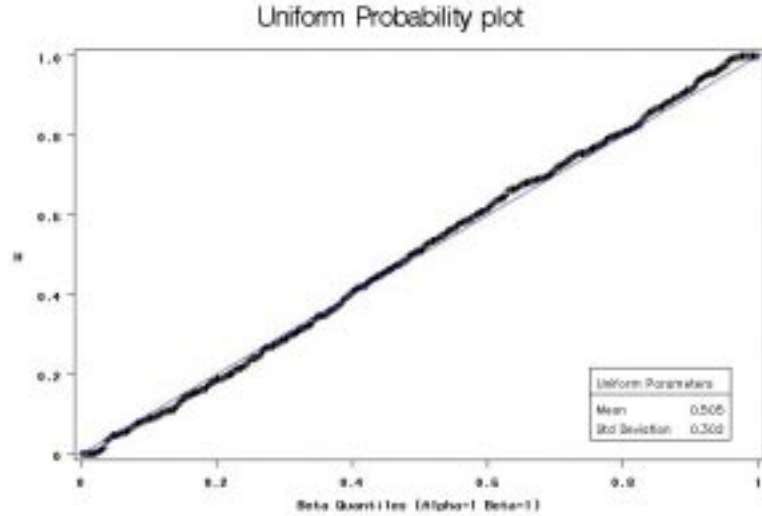


Figure 2.1: Probability plot of $\{H_k\}$ for μ parameter of WN model

We use the above method to validate WN, WC and WDE models. For each model,

using Kolmogorov-Smirnoff test, we test if the samples $\{H_k, k = 1, \dots, m\}$ are from $U(0, 1)$ distribution. Though this was done for both parameters μ and ρ in WN, WC and WDE models, we present the results only for μ parameter of WN model in Figure 2.1. The Kolmogorov-Smirnov distance is 0.0354 and the p-value was greater than 0.25. Therefore, we conclude that the sample $\{H_k, k = 1, \dots, 500\}$ is from $U(0, 1)$ distribution, which is expected. We obtained similar results for all the other goodness of fit tests.

2.6 Wrapped Extreme Value and Bimodal distributions

In this section, we introduce two new wrapped distributions - Wrapped Extreme Value distribution (WEV) and bimodal or Wrapped Beta distribution (WB). When we wrap a Beta distribution from the real line onto the circle, we obtain a bimodal distribution for low values of ρ . The general theory is the same as that given in Section 2.2. The only difference is that for simplicity, we use an Inverse-Gamma(0.5,0.5) prior for σ instead of Beta(0.5, 0.5) prior for ρ .

The Extreme Value density is given by

$$\frac{1}{\sigma} e^{-\left(\frac{u-\mu}{\sigma}\right)} e^{-e^{-\left(\frac{u-\mu}{\sigma}\right)}}, \quad -\infty < u < \infty.$$

The Wrapped Extreme Value density is obtained by wrapping the Extreme value density onto the circle. The mean resultant, $\rho = \sqrt{\frac{\pi\sigma}{\sinh(\pi\sigma)}}$ is a strictly decreasing function of σ .

The general Beta density is defined as

$$\frac{1}{\text{Beta}(a,b)} \frac{(u-l)^{a-1}(r-u)^{b-1}}{(r-l)^{a+b-1}}, \quad l < u < r, \quad a > 0, \quad b > 0.$$

Consider the case where $a = 2$, $b = 2$, $l = \mu - \sigma$ and $r = \mu + \sigma$. Then, the density on the

real line is,

$$\frac{3}{4\sigma} \left[1 - \left(\frac{u-\mu}{\sigma} \right)^2 \right], \quad \mu - \sigma < u < \mu + \sigma. \quad (2.4)$$

The bimodal density is obtained by wrapping the density given in (2.4) onto the circle for most of the values of $\sigma > \pi$. The mean resultant, $\rho = \left| \frac{-3}{\sigma^3} (-\sin(\sigma) + \cos(\sigma)\sigma) \right|$ is not a strictly decreasing function of σ . A plot of ρ as a function of σ is given in Figure 2.2. For our simulation, we use the value of $\sigma = 4$, that is $\rho = 0.087$. A plot of the bimodal density for $\rho = 0.087$ is given in Figure 2.3 and Figure 2.4. In Figure 2.3, we plot the bimodal density between $-\pi$ to π . This clearly shows the bimodal structure of the density. However, note that the density is very flat. It varies between 0.13 and 0.19. Therefore, when we plot this density on the circle in Figure 2.4, it looks similar to a Circular Uniform density.

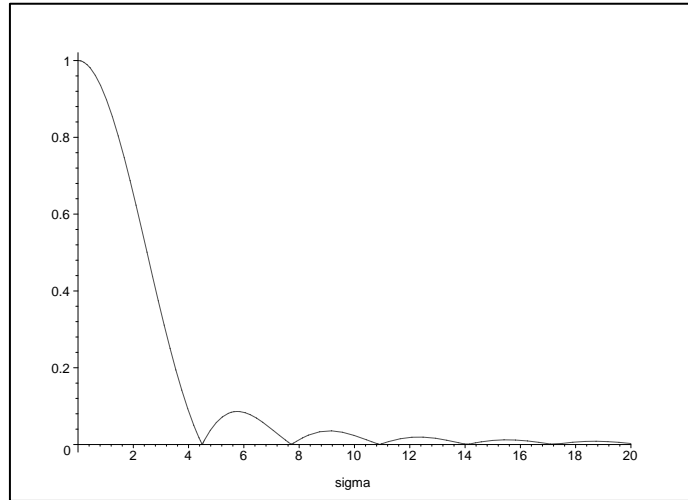


Figure 2.2: Plot of ρ for the bimodal distribution

For our simulation studies, we generate samples of size $n = 50$ from the Wrapped Extreme Value (WEV) distribution with parameters set at $\mu = \pi/2 = 1.5708$ and $\rho = 0.5$. We also generate samples of size $n = 50$ from the Wrapped Beta (WB) distribution

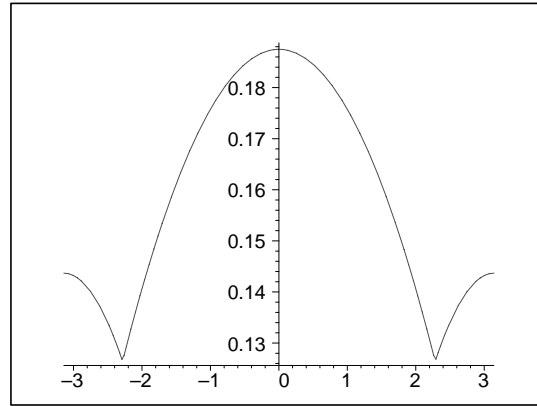


Figure 2.3: Plot of the bimodal density for $\rho = 0.087$

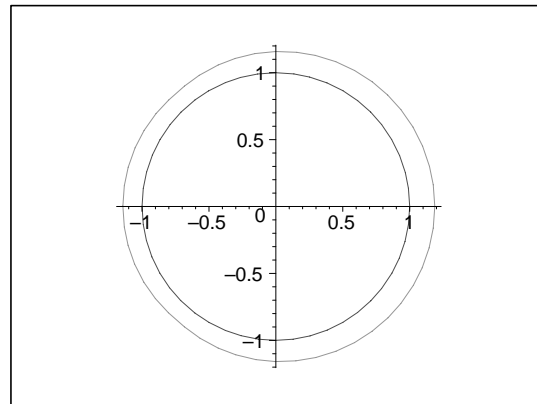


Figure 2.4: Plot of the bimodal density for $\rho = 0.087$ on the circle

Table 2.8: Fitting WEV to WEV distribution

	mean	std dev	2.5 %	50 %	97.5 %	cov prob
μ	1.58	0.18	1.25	1.58	1.95	0.91
s.e.	0.20	0.05	0.19	0.19	0.27	0.01
ρ	0.50	0.07	0.35	0.50	0.62	0.91
s.e.	0.07	0.01	0.09	0.07	0.06	0.01

Table 2.9: Fitting WB to WB distribution

	mean	std dev	2.5 %	50 %	97.5 %	cov prob
μ	3.10	1.41	0.64	3.10	5.57	0.96
s.e.	0.78	0.46	0.82	0.98	0.92	0.01
ρ	0.09	0.06	0.01	0.08	0.22	0.88
s.e.	0.06	0.02	0.02	0.06	0.08	0.01

with parameters set at $\mu = \pi = 3.1416$ and $\rho = 0.087$. These values are chosen for WB distribution parameters to obtain a bimodal distribution. We fit the true distributions and compute the usual statistics as described in Section 2.4. In each simulation, we choose the burn-in period to be 2000 samples, and then keep 5000 samples after burn-in, to obtain posterior summary values. Table 2.8 and Table 2.9 contain summary values based on these final 5000 samples. We repeat the entire procedure 500 times to see the frequentist performance of the proposed Bayes method. In SAS, on a sparc 20 machine, on an average it took about 50 minutes to perform the entire simulation for a given wrapped distribution.

From Table 2.8, we see that for the WEV distribution, as the underlying distribution is true, both μ and ρ have been well estimated. However, for the bimodal (WB) distribution in Table 2.9, we see that μ is well estimated but the coverage probability for ρ is lower than expected. One reason for this could be that the original value of ρ is 0.087,

which is quite low (that is, the value of σ is high).

2.7 Application to real data sets

In this section, we analyze two data sets. The first data set is the Jander's ant data set, which refers to the movement of ants. The second data set is the data set about ozone concentration and wind direction. We will use only the wind direction part of the dataset for analysis in this section. In the next chapter, this data is analyzed in more detail, and we fit a regression model to the data.

2.7.1 Jander's ant data

We analyze a data set that depicts the orientation of the ants towards a black target when released in a round arena. This experiment was originally conducted by Jander (1957) and later mentioned by Batschelet (1981) and Fisher (1993). The data consists of 100 observations. For this data set, the circular sample mean and the resultant are 3.20 and 0.61.

We plot a circular histogram of Jander's ant data in Figure 2.5. We fit WN, WC and WDE distributions to this data. We compute the posterior mean, standard deviation, 2.5 percentile, median and 97.5 percentile for μ and ρ . The percentiles are computed with 0 radians as the reference point. We have used a class of non-informative priors for (μ, ρ) , which is given in equation (2.2), but present the results only for $a_\rho = 0.5$, in Table 2.10.

At first, while fitting the data, we rejected the first 2000 samples (burn-in period) of the MCMC chain and kept the next 5000 samples after burn-in to do posterior inference.

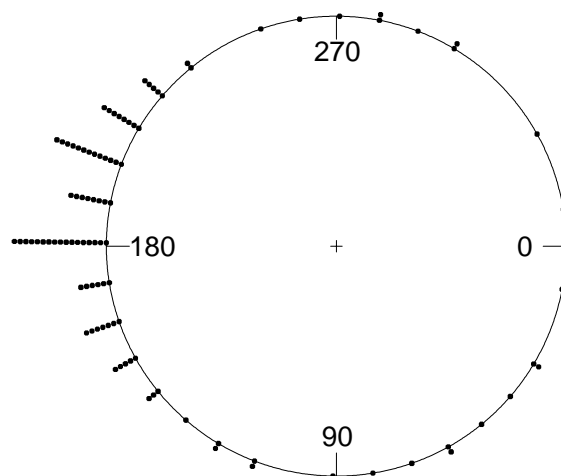


Figure 2.5: Circular plot of the Jander's Ant data

We noted that there were difficulties in achieving convergence using CODA. Therefore, we increased the burn-in time to 10000 samples and kept the next 30000 samples. Initially we use the sample estimates of μ and ρ as our starting values. We then change our starting values for MCMC (for example, $\mu = 0$ and $\rho = 0.99$) and overlay this Markov Chain with the previous Markov Chain. We find that after 3000 samples they both overlap each other and appear to mix very well. Analyzing similar results (not given here), there is no indication of MCMC convergence problem. Summary values for the parameters are given in Table 2.10.

To compare the models we use Gelfand and Ghosh (1998) criteria (GGC). Low values of GGC indicates better fit for a model. From Table 2.10, we see that the WC distribution performs best based on the GGC. WDE distribution estimates are close to the WC distribution estimates. However, the estimates from WN are different from WC and WDE. GGC recommends WDE over WN. Using CODA, we obtained the trace plots and

Table 2.10: Fitting WN, WC and WDE to the Ant data set with $\rho \sim \text{Beta}(0.5, 0.5)$

<i>WN</i>	mean	std dev	2.5 %	50 %	97.5 %
μ	3.14	0.14	2.88	3.14	3.42
ρ	0.54	0.05	0.44	0.54	0.62
GGC	243.97	33.80	183.70	241.95	316.12
<i>WC</i>	mean	std dev	2.5 %	50 %	97.5 %
μ	3.24	0.06	3.12	3.25	3.36
ρ	0.64	0.04	0.57	0.64	0.72
GGC	141.97	11.84	121.44	141.06	167.81
<i>WDE</i>	mean	std dev	2.5 %	50 %	97.5 %
μ	3.21	0.08	3.08	3.19	3.37
ρ	0.62	0.05	0.52	0.62	0.71
GGC	183.43	23.67	142.09	181.62	234.22

the kernel density estimates of the parameters μ and ρ based on the Wrapped Cauchy (WC) distribution, which are shown in Figure 2.6.

Bayesian highest posterior density (HPD) can also be used instead of percentiles. For example, in Figure 2.7, we plot the posterior density of μ while fitting WC to Ant data. Since the posterior density has most its mass close to 3.24 radians (185°), we computed the percentiles with 0 radians as the reference point. The percentiles will change if we change the reference point, but the change will be minimal for the Jander's ant data set, if the reference point is away from 3.24 radians. However, HPD is not affected by the change in the reference point.

This data has been previously analyzed by Sengupta and Pal (2001). They find through hypothesis testing, that WN does not fit this data very well. They recommend a family of symmetric wrapped stable distributions (SWS, refer Mardia and Jupp, 1999, p. 52) or a mixture of SWS and Circular Uniform density for this data. This agrees with the

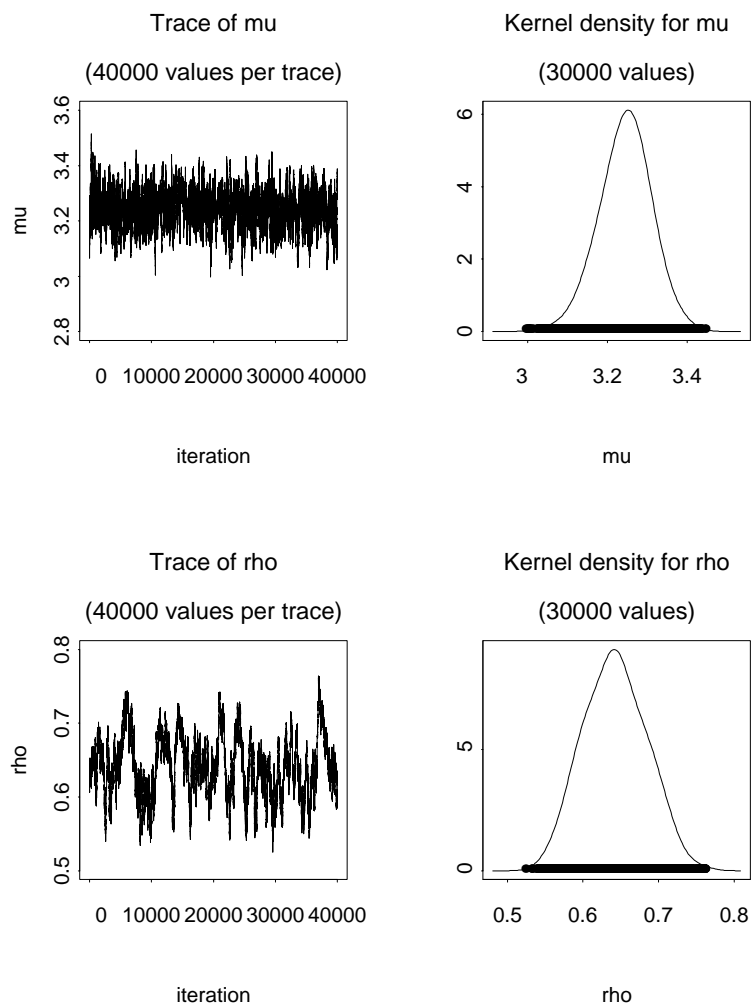


Figure 2.6: Trace plots while fitting WC to Ant data

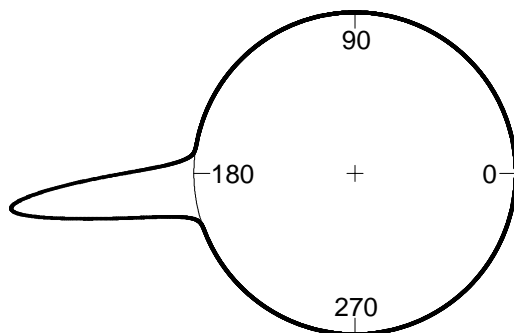


Figure 2.7: Posterior density of μ while fitting WC to Ant data

result recommended by the GGC statistic, where WN is considered to be a weak fit. In addition, we recommend WC, which belongs to the SWS family.

2.7.2 Ozone data set

We analyze a data set containing the measurements of ozone concentration and wind direction. Measurements were taken at 6:00 a.m. at four-day intervals between April 18th and June 29th 1975, at a weather station in Milwaukee. In this section, we will use only the wind direction part of the data. This is shown in Figure 2.8. The circular sample mean and the resultant for wind direction data are 6.57 radians (16°) and 0.52 respectively. For more details under the regression setup, refer to Section 3.6 and Figure 3.1.

We compute the posterior mean, standard deviation, 2.5 percentile, median and 97.5 percentile for μ and ρ . The percentiles are computed with $\pi \approx 3.14$ radians (180°) as the reference point. We rejected the first 5000 samples (burn-in) and kept the next 15000 samples. We used the sample estimates of the mean and the mean resultant as starting

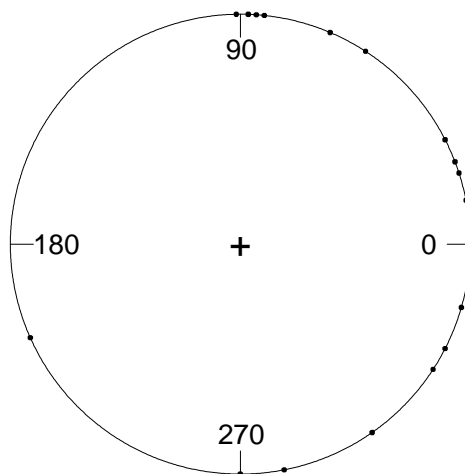


Figure 2.8: Plot of wind directions in the ozone data.

values of μ and ρ . Analyzing results from CODA (not given here), there is no indication of MCMC convergence problem. Summary values for the parameters are given in Table 2.11.

From Table 2.11, we see that the WN distribution performs best based on the Dev. The parameter estimates of μ from WC and WDE are very close to that of WN. The estimate for ρ from WC is lower than that of WN and WDE. We compare this model and the regression model in Section 3.6

2.8 Discussion

In this chapter, we provide a generic method to fit a wide class of continuous densities on the circle using MCMC methodology. We show that using this method, we can easily obtain samples from the joint posterior distribution, and hence, compute the posterior

Table 2.11: Fitting WN, WC and WDE to the Ozone data set with $\rho \sim \text{Beta}(0.5, 0.5)$

<i>WN</i>	mean	std dev	2.5 %	50 %	97.5 %
μ	6.50	0.37	5.77	6.50	7.21
ρ	0.44	0.13	0.13	0.45	0.65
GGC	63.56	22.98	28.73	60.36	116.13
Dev	61.03	2.09	58.94	60.37	66.93
<i>WC</i>	mean	std dev	2.5 %	50 %	97.5 %
μ	6.54	0.53	5.26	6.57	7.54
ρ	0.34	0.17	0.03	0.34	0.64
GGC	59.51	28.96	23.97	51.64	127.78
Dev	63.47	2.65	60.27	62.84	69.83
<i>WDE</i>	mean	std dev	2.5 %	50 %	97.5 %
μ	6.53	0.44	5.86	6.54	7.34
ρ	0.45	0.16	0.10	0.46	0.70
GGC	57.90	25.77	24.89	51.29	120.14
Dev	61.71	2.56	59.00	60.99	68.03

summary values such as the posterior mean, standard deviation, 2.5 percentile, median and 97.5 percentile for μ and ρ . From simulation-based studies, we find that parameter estimates based on WDE are robust to erroneous models. In addition, we find that the posterior distribution is not sensitive to the non-informative class of priors that we proposed.

We also give examples of bimodal density and Wrapped Extreme Value density using this methodology. We extend this to other real life applications, with some explanatory variables in the next chapter.

Chapter 3

Circular Regression

3.1 Introduction

In this chapter, we explore the dependence of the circular response variable on other circular and linear variables. There are many examples of regression involving circular data. Consider the environmental data collected on the wind direction and ozone concentration. This study was performed at a weather station in Milwaukee in 1975 (Johnson and Wehrly, 1977). There seemed to be some evidence of association between the wind direction and ozone concentration. Figure 3.1 shows the joint data plot. The ozone concentrations are plotted as distances from the center and the angles represent the wind directions.

Another interesting biological example refers to the relationship between the orientation of the nests of 50 scrub birds along the bank of a creek bed and the direction of the flow of the creek at the nearest point to the nest (Fisher, 1993). Another example is the study of 31 blue periwinkles performed by Dr. A. Underwood and Ms G. Chapman

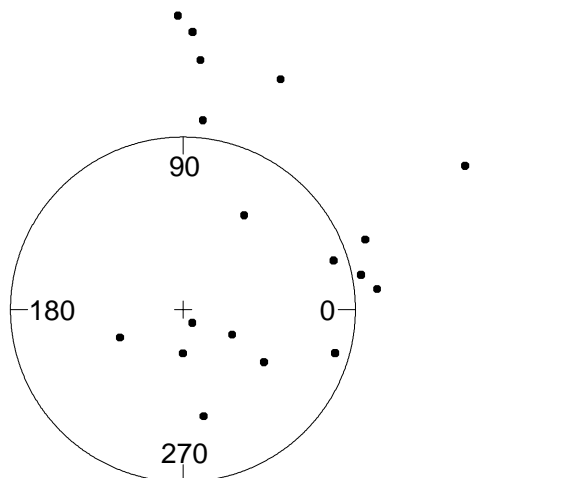


Figure 3.1: Joint data plot of the ozone data. The ozone concentrations are plotted as distances from the center.

(Fisher, 1993). Periwinkles are small snail-like intertidal animals living close to the sea shore. They were transplanted near the seashore from the heights at which they normally live. The position (direction of movement and the distance moved) of 15 periwinkles were studied after 1 day and the rest were studied after 4 days.

3.2 Previous work

We are interested in modeling the variation of the mean direction of a circular or linear response y , in terms of one or more covariates or explanatory variables. The covariates can be linear or circular. When the response variable is linear, and it is regressed on a circular covariate, it is the same as the linear regression model. Standard statistical methods can be applied here. A natural model introduced by Mardia (1976) is

$$X|y \sim N(\beta_0 + \beta_1 \cos(y) + \beta_2 \sin(y), \sigma^2),$$

where y is a circular covariate and β_0 , β_1 , β_2 and σ^2 are unknown parameters. Some extensions of this model were proposed by Johnson and Wehrly (1978).

Our main interest is when the response variable is circular and the covariates are linear or circular. Some of the earlier work in this field has been done by Gould (1969) in which he introduces the regression model,

$$y_i \sim VM(\beta_0 + \beta_1 x_i, \kappa), i = 1, 2, \dots, n,$$

where $VM()$ is the Von Mises distribution. Laycock (1975) describes some optimal designs for this model. The major drawback of this approach is that the likelihood function has infinitely many maxima. Furthermore, β_1 is not identifiable when the covariates x_1, x_2, \dots, x_n are equally spaced.

Johnson and Wehrly (1978) proposed a different class of models in which the response variable completes only a single spiral, as x varies through its range. Fisher and Lee (1992) extended this concept by suggesting the use of a general link function. This link function is a one-to-one function $g()$, which maps the entire real line onto $(-\pi, \pi)$ and satisfies $g(0) = 0$. Some examples are $g(x) = 2 \tan^{-1}(x)$ and $g(x) = 2\pi F(x)$, where $F(x)$ is a distribution function. Using the link function, Fisher (1993) also describes the regression for the mean resultant and regression with mixed models, using Von Mises distribution and MLE techniques.

3.3 Extension of the Data Augmentation Approach for Regression

For the regression problem, we use the data augmentation approach in a Bayesian setup. We extend the theory given in Section 2.2 for regression. Let the observations on the circle be $\mathbf{y} = \{y_1, y_2, y_3, \dots, y_n\}$, $0 \leq y_j < 2\pi$, $j = 1 \dots n$. Let $X = \{x_1, x_2, x_3, \dots, x_n\}$, $j = 1 \dots n$, be the design matrix. We consider the simple regression model,

$$Y_j|X_j \sim WD(\mu_j, \rho),$$

$$\text{where } \mu_j = \beta_0 + \sum_{i=1}^p \beta_i g_i(X_{ij}) \quad (3.1)$$

and $\beta_0, \beta_1, \dots, \beta_p$ and ρ are parameters to be estimated. $g_i()$ s are basis functions and WD represents any wrapped distribution such as WN, WC or WDE. This class of regression models can be extended to generalized additive models (GAM) for circular data (Hastie and Tibshirani, 1986). However, we assume that the functions $g_i()$ s are completely known. Without loss of generality, we can restrict β_0 between 0 and 2π , to avoid identifiability problems. In general, the basis functions $g_i()$ s are some general set of functions that are defined on the centered and scaled covariates. For simplicity in notation, we omit the functions $g_i()$ s, but the covariates are now assumed to be centered and scaled.

Note that in our approach, the infinitely many maxima problem and identifiability problem of the Gould (1969) model do not arise. Moreover, it is not necessary to use a one-to-one link function, which transforms the real line onto $(-\pi, \pi)$ (Johnson and Wehrly, 1978). This is because we are using \mathbf{y} and \mathbf{k} to fit the data rather than just \mathbf{y} , as suggested by Gould (1969) and Johnson and Wehrly (1978). In order to illustrate the theory we

assume that $p = 1$, that is, we use only one covariate. However, it should be noted that the extension to the case of p covariates is straightforward.

A class of non-informative prior for (β_0, β_1, ρ) can be specified as,

$$[\beta_0, \beta_1, \rho] \propto I_{\beta_0}(0, 2\pi) e^{-\frac{\beta_1^2}{2\sigma_{\beta_1}^2}} \rho^{a_\rho-1} (1-\rho)^{a_\rho-1}, \quad a_\rho > 0, \quad (3.2)$$

and using $\rho = h(\sigma)$, the joint density of $(\beta_0, \beta_1, \sigma)$ is given by

$$[\beta_0, \beta_1, \sigma] \propto I_{\beta_0}(0, 2\pi) e^{-\frac{\beta_1^2}{2\sigma_{\beta_1}^2}} h(\sigma)^{a_\rho-1} (1-h(\sigma))^{a_\rho-1} |h'(\sigma)|, \quad a_\rho > 0.$$

Following the same approach described in Section 2.2 and replacing μ by $\beta_0 + \beta_1 x_j$, the joint density of $\mathbf{y}, \mathbf{k}, \beta_0, \beta_1$ and σ is given by

$$\begin{aligned} [\mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \sigma] &\propto [\mathbf{y}|\mathbf{k}, \beta_0, \beta_1, \sigma^2] [\mathbf{k}|\beta_0, \beta_1, \sigma^2] [\beta_0, \beta_1, \sigma^2] \\ &\propto \prod_{j=1}^n \left(\frac{1}{\sigma} f\left(\frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{\sigma}\right) \right) h(\sigma)^{a_\rho-1} (1-h(\sigma))^{a_\rho-1} |h'(\sigma)| \\ &\quad I_{\beta_0}(0, 2\pi) e^{-\frac{\beta_1^2}{2\sigma_{\beta_1}^2}} \\ &\propto \frac{1}{\sigma^{n+n_0}} \prod_{j=1}^n f\left(\frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{\sigma}\right) h(\sigma)^{a_\rho-1+n_1} (1-h(\sigma))^{a_\rho-1} h_1(\sigma) \\ &\quad I_{\beta_0}(0, 2\pi) e^{-\frac{\beta_1^2}{2\sigma_{\beta_1}^2}}, \end{aligned}$$

where $|h'(\sigma)|$ can be factorized as $\frac{1}{\sigma^{n_0}} h(\sigma)^{n_1} h_1(\sigma)$. It is assumed that $h_1(\sigma)$ is invertible.

The full conditional densities of $\mathbf{k}, \beta_0, \beta_1$ and σ are nonstandard densities. Consequently, we introduce auxiliary variables \mathbf{z} and \mathbf{v} . Let $\mathbf{z} = \{z_0, z_1, z_2, z_3\}$ and $\mathbf{v} = \{v_1, v_2, v_3, \dots, v_n\}$, such that

$$[\mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \sigma] \propto \int [\mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}, \mathbf{v}, \sigma] d\mathbf{z} d\mathbf{v}.$$

The joint density of $\mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}, \mathbf{v}$ and σ is given by

$$\begin{aligned} [\mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}, \mathbf{v}, \sigma] &\propto I_{z_0} \left(0, \frac{1}{\sigma^{n+n_0}}\right) \prod_{j=1}^n I_{v_j} \left(0, f\left(\frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{\sigma}\right)\right) I_{z_1} \left(0, h(\sigma)^{a_\rho - 1 + n_1}\right) \\ &\quad \{I_{z_2} \left(0, (1 - h(\sigma))^{a_\rho - 1}\right) I(a_\rho \neq 1) + I(a_\rho = 1)\} I_{z_3} \left(0, h_1(\sigma)\right) \\ &\quad I_{\beta_0} \left(0, 2\pi\right) e^{-\frac{\beta_1^2}{2\sigma^2 \beta_1}}. \end{aligned}$$

Given v_j , we get $\{(gm_l(v_j), gM_l(v_j)), l = 1 \dots q, q \in \mathcal{Z}\}$ such that,

$$\left\{0 < v_j < f\left(\frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{\sigma}\right)\right\} = \bigcup_{l=1}^q \{gm_l(v_j) < \left(\frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{\sigma}\right) < gM_l(v_j)\}$$

is a union of disjoint sets because $gm_1(v_j) < gM_1(v_j) < gm_2(v_j) < gM_2(v_j) < \dots < gm_q(v_j) < gM_q(v_j)$.

The full conditional densities of $\mathbf{k}, \mathbf{v}, \mathbf{z}, \beta_0$ and β_1 are given by

$$v_j | \mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}, \mathbf{v}_{-j}, \sigma \sim U \left[0, f\left(\frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{\sigma}\right)\right]$$

$$\begin{aligned} k_j | \mathbf{y}, \mathbf{k}_{-j}, \beta_0, \beta_1, \mathbf{z}, \mathbf{v}, \sigma &\sim DU \left[\bigcup_{l=1}^q \left\{ \left\lceil \frac{1}{2\pi} (\beta_0 + \beta_1 x_j - y_j + \sigma gm_l(v_j)) \right\rceil, \right. \right. \\ &\quad \left. \left. \left\lfloor \frac{1}{2\pi} (\beta_0 + \beta_1 x_j - y_j + \sigma gM_l(v_j)) \right\rfloor \right\} \right], \end{aligned}$$

where DU stands for Discrete Uniform.

Without loss of generality, let $l_o \in \{1, 2, \dots, q\}$ be such that,

$$\left\lceil \frac{1}{2\pi} (\beta_0 + \beta_1 x_j - y_j + \sigma gm_{l_o}(v_j)) \right\rceil \leq k_j \leq \left\lfloor \frac{1}{2\pi} (\beta_0 + \beta_1 x_j - y_j + \sigma gM_{l_o}(v_j)) \right\rfloor.$$

$$z_0 | \mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}_{-0}, \mathbf{v}, \sigma \sim U \left[0, \frac{1}{\sigma^{n+n_0}}\right]$$

$$z_1 | \mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}_{-1}, \mathbf{v}, \sigma \sim U \left[0, h(\sigma)^{a_\rho - 1 + n_1}\right]$$

$$z_2 | \mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}_{-2}, \mathbf{v}, \sigma \sim \begin{cases} U \left[0, (1 - h(\sigma))^{a_\rho - 1}\right], & a_\rho \neq 1 \\ z_2 \text{ is not needed, } & a_\rho = 1 \end{cases}$$

$$z_3 | \mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}_{-3}, \mathbf{v}, \sigma \sim \begin{cases} U[0, h_1(\sigma)], & h_1(\sigma) \neq 1 \\ z_3 \text{ is not needed, } & h_1(\sigma) \equiv 1 \end{cases}$$

$$\beta_0 | \mathbf{y}, \mathbf{k}, \beta_1, \mathbf{z}, \mathbf{v}, \sigma \sim U[m_{\beta_0}, M_{\beta_0}],$$

$$\text{where } m_{\beta_0} = \max_{j=1}^n [y_j + 2\pi k_j - \beta_1 x_j - \sigma g M_{l_0}(v_j)] \bigvee 0$$

$$M_{\beta_0} = \min_{j=1}^n [y_j + 2\pi k_j - \beta_1 x_j - \sigma g m_{l_0}(v_j)] \bigwedge (2\pi)$$

$$\beta_1 | \mathbf{y}, \mathbf{k}, \mathbf{z}, \beta_0, \mathbf{v}, \sigma \sim U[m_{\beta_1}, M_{\beta_1}],$$

$$\text{where } m_{\beta_1} = \max_{j=1}^n m_{\beta_1, j} \bigvee 0,$$

$$M_{\beta_1} = \min_{j=1}^n M_{\beta_1, j} \bigwedge (2\pi),$$

where

$$m_{\beta_1, j} = \begin{cases} \left[\frac{y_j + 2\pi k_j - \beta_0 - \sigma g M_{l_0}(v_j)}{x_j} \right], & x_j > 0 \\ \left[\frac{y_j + 2\pi k_j - \beta_0 - \sigma g m_{l_0}(v_j)}{x_j} \right], & x_j < 0 \end{cases}$$

$$M_{\beta_1, j} = \begin{cases} \left[\frac{y_j + 2\pi k_j - \beta_0 - \sigma g m_{l_0}(v_j)}{x_j} \right], & x_j > 0 \\ \left[\frac{y_j + 2\pi k_j - \beta_0 - \sigma g M_{l_0}(v_j)}{x_j} \right], & x_j < 0. \end{cases}$$

Since $h(\sigma)$ is a monotone decreasing function of σ and is invertible, h^{-1} is also a monotone function. This enables us invert the relationship between z_1 , z_2 and σ and obtain some parts of the distribution for σ .

Also, since $h_1(\sigma)$ is invertible, given z_3 , we get $\{(gm_l^*(z_3), gM_l^*(z_3)), l = 1 \dots q, q \in \mathcal{Z}\}$ such that, $\{0 < z_3 < h_1(\sigma)\} = \bigcup_{l=1}^q \{gm_l^*(z_3) < \sigma < gM_l^*(z_3)\}$ is a union of disjoint sets because $gm_1^*(z_3) < gM_1^*(z_3) < gm_2^*(z_3) < gM_2^*(z_3) < \dots < gm_q^*(z_3) < gM_q^*(z_3)$. If $h_1(\sigma)$ is not easily invertible, then factorize $h_1(\sigma)$ into functions, whose inverse can be easily computed. Then, this method can also be easily extended, similar to the extension given for

the density in Section 2.2.

$$\sigma|\mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}, \mathbf{v} \sim U \left[(m_\sigma, M_\sigma) \cap \bigcup_{l=1}^q \{gm_l^*(z_3), gM_l^*(z_3)\} \right],$$

$$m_\sigma = \begin{cases} \max_{j=1}^n h_m \left(\frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gm_{l_0}(v_j)}, \frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gM_{l_0}(v_j)} \right) \vee h^{-1} \left(1 - z_2^{\frac{1}{a_\rho - 1}} \right), & a_\rho > 1 \\ \max_{j=1}^n h_m \left(\frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gm_{l_0}(v_j)}, \frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gM_{l_0}(v_j)} \right), & a_\rho \leq 1 \end{cases}$$

$$M_\sigma = \begin{cases} \min_{j=1}^n h_M \left(\frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gm_{l_0}(v_j)}, \frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gM_{l_0}(v_j)} \right) \wedge \frac{1}{z_0^{\frac{1}{n+n_0}}} \wedge h^{-1} \left(z_1^{\frac{1}{a_\rho - 1 + n_1}} \right), \\ \text{where } (a_\rho \geq 1) \cup ((a_\rho < 1) \cap (z_2 \leq 1)) \\ \left[\min_{j=1}^n h_M \left(\frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gm_{l_0}(v_j)}, \frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gM_{l_0}(v_j)} \right) \wedge \frac{1}{z_0^{\frac{1}{n+n_0}}} \wedge h^{-1} \left(z_1^{\frac{1}{a_\rho - 1 + n_1}} \right) \right. \\ \left. \wedge h^{-1} \left(1 - \frac{1}{z_2^{\frac{1}{1-a_\rho}}} \right) \right], \text{ where } ((a_\rho < 1) \cap (z_2 > 1)), \end{cases}$$

where $h_m \left(\frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gm_{l_0}(v_j)}, \frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gM_{l_0}(v_j)} \right)$ and $h_M \left(\frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gm_{l_0}(v_j)}, \frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gM_{l_0}(v_j)} \right)$

are defined as follows.

$$h_m \left(\frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gm_{l_0}(v_j)}, \frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gM_{l_0}(v_j)} \right) = \begin{cases} \frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gM_{l_0}(v_j)}, & gm_{l_0}(v_j) \geq 0 \\ \frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gm_{l_0}(v_j)}, & gM_{l_0}(v_j) \leq 0 \\ \left[\frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gm_{l_0}(v_j)} \vee \frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gM_{l_0}(v_j)} \right], & \\ gm_{l_0}(v_j) < 0 < gM_{l_0}(v_j) \end{cases}$$

$$h_M \left(\frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gm_{l_0}(v_j)}, \frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gM_{l_0}(v_j)} \right) = \begin{cases} \frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gm_{l_0}(v_j)}, & gm_{l_0}(v_j) \geq 0 \\ \frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gM_{l_0}(v_j)}, & gM_{l_0}(v_j) \leq 0 \\ [h_M \text{ is not needed,} \\ gm_{l_0}(v_j) < 0 < gM_{l_0}(v_j)] \end{cases}$$

For symmetric unimodal densities,

$$h_m \left(\frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gM_{I_0}(v_j)}, \frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gM_{I_0}(v_j)} \right) = \frac{|y_j - \beta_0 - \beta_1 x_j + 2\pi k_j|}{gM_{I_0}(v_j)}$$

and h_M is not needed.

We illustrate the above technique for several popular distributions such as Wrapped Normal, Wrapped Cauchy and Wrapped Double Exponential distributions. In Appendix B, we provide more details on the exact form of the full conditional distributions for these three wrapped distributions. We study the performance of the proposed method by several simulation experiments.

3.4 Simulation studies

For our simulation studies, we generate samples of size $n = 31$ from Wrapped Normal (WN), Wrapped Cauchy (WC) and Wrapped Double Exponential (WDE) distributions with parameters set at $\beta_0 = 1$, $\rho = 0.8$ and varying $\beta_1 = -1.5, 0, 1.5$. The observations for the covariate \mathbf{x} , were the distance measurements from the movement of periwinkle data (Fisher, 1993). The covariates were centered and scaled as mentioned previously. We fit each of the three distributions with priors given in equation (3.2) with $a_\rho = 0.5$ and $\sigma_{\beta_1} = 5$. We compute the posterior mean, standard deviation, 2.5 percentile, median and 97.5 percentile for β_0 , β_1 and ρ for each simulation. The percentiles are computed with 0 radians as the reference point. The simulation standard errors for each of these summary values are also computed. We also compute the coverage probability for the 95% posterior interval given by the 2.5 and 97.5 percentile of the posterior distribution. In each simulation, we choose the burn-in period to be 2000 samples (i.e. throw away first 2000 samples from the

Table 3.1: Regressing WN model to WN data

$\beta_1 = -1.5$	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.01	0.13	0.76	1.01	1.27	0.94
s.e.	0.20	0.10	0.14	0.24	0.28	0.01
β_1	-1.48	0.19	-1.84	-1.48	-1.10	0.95
s.e.	0.26	0.17	0.52	0.25	0.56	0.01
ρ	0.78	0.05	0.66	0.79	0.87	0.93
s.e.	0.06	0.02	0.09	0.06	0.03	0.01
$\beta_1 = 0$	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	0.99	0.13	0.75	0.99	1.24	0.96
s.e.	0.12	0.02	0.12	0.12	0.13	0.01
β_1	0.00	0.17	-0.34	0.00	0.35	0.95
s.e.	0.17	0.03	0.17	0.17	0.18	0.01
ρ	0.78	0.05	0.66	0.79	0.86	0.93
s.e.	0.05	0.01	0.08	0.05	0.04	0.01
$\beta_1 = 1.5$	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.01	0.12	0.77	1.01	1.26	0.97
s.e.	0.11	0.02	0.12	0.11	0.12	0.01
β_1	1.49	0.17	1.16	1.49	1.83	0.95
s.e.	0.16	0.02	0.17	0.16	0.17	0.01
ρ	0.79	0.05	0.67	0.79	0.87	0.94
s.e.	0.05	0.01	0.07	0.05	0.03	0.01

MCMC chain) and keep 5000 samples after burn-in, to obtain posterior summary values. All summary values in Table 3.1, Table 3.2 and Table 3.3 are based on these final 5000 samples. We repeat the entire procedure 500 times to see the frequentist performance of the proposed Bayes method. In SAS, on a pentium III machine, on an average it took about 30 minutes to perform the entire simulation for a given wrapped distribution.

From Table 3.1, Table 3.2 and Table 3.3, we see that the proposed method performs very well in terms of maintaining the nominal coverage probability. Also, the posterior mean and median serve as good point estimates of the parameters.

In order to study the sensitivity of the sampling distribution, we generated sam-

Table 3.2: Regressing WC model to WC data

$\beta_1 = -1.5$	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.00	0.07	0.86	1.00	1.13	0.95
s.e.	0.07	0.02	0.08	0.07	0.08	0.01
β_1	-1.50	0.10	-1.70	-1.50	-1.30	0.96
s.e.	0.11	0.04	0.14	0.11	0.12	0.01
ρ	0.78	0.05	0.67	0.79	0.86	0.93
s.e.	0.05	0.01	0.08	0.05	0.04	0.01
$\beta_1 = 0$	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.00	0.07	0.86	1.00	1.13	0.98
s.e.	0.06	0.02	0.07	0.06	0.07	0.01
β_1	-0.01	0.10	-0.21	-0.00	0.19	0.96
s.e.	0.10	0.04	0.14	0.10	0.12	0.01
ρ	0.78	0.05	0.67	0.79	0.86	0.93
s.e.	0.05	0.01	0.08	0.05	0.03	0.01
$\beta_1 = 1.5$	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.01	0.07	0.88	1.01	1.14	0.95
s.e.	0.23	0.02	0.22	0.23	0.24	0.01
β_1	1.49	0.10	1.29	1.49	1.69	0.93
s.e.	0.15	0.05	0.21	0.15	0.15	0.01
ρ	0.78	0.05	0.67	0.79	0.87	0.91
s.e.	0.06	0.01	0.08	0.06	0.04	0.01

Table 3.3: Regressing WDE model to WDE data

$\beta_1 = -1.5$	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.00	0.11	0.79	1.00	1.21	0.94
s.e.	0.11	0.03	0.12	0.11	0.13	0.01
β_1	-1.50	0.16	-1.81	-1.50	-1.19	0.94
s.e.	0.16	0.04	0.19	0.16	0.17	0.01
ρ	0.78	0.06	0.63	0.78	0.88	0.95
s.e.	0.06	0.01	0.09	0.06	0.04	0.01
$\beta_1 = 0$	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.00	0.11	0.79	1.00	1.22	0.94
s.e.	0.11	0.03	0.12	0.11	0.12	0.01
β_1	0.01	0.16	-0.30	0.01	0.32	0.93
s.e.	0.16	0.05	0.18	0.16	0.19	0.01
ρ	0.78	0.06	0.63	0.79	0.88	0.92
s.e.	0.06	0.01	0.09	0.06	0.04	0.01
$\beta_1 = 1.5$	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.00	0.11	0.79	1.00	1.21	0.96
s.e.	0.10	0.02	0.11	0.10	0.11	0.01
β_1	1.50	0.16	1.19	1.50	1.80	0.94
s.e.	0.16	0.04	0.18	0.16	0.19	0.01
ρ	0.78	0.06	0.64	0.79	0.88	0.95
s.e.	0.06	0.01	0.09	0.06	0.04	0.01

ples of size $n = 31$ from WN, WC and WDE with parameters $\beta_0 = 1$, $\beta_1 = 1.5$ and $\rho = 0.8$. We then fitted Wrapped Normal, Wrapped Cauchy and Wrapped Double Exponential distributions to the three datasets. We repeated the method 500 times to see the frequentist performance of the Bayes method for erroneous models. As before, we report the posterior mean, standard deviation, and two equal tail percentiles along with the Monte Carlo standard error. The results are tabulated in Table 3.4, Table 3.5 and Table 3.6.

Comparing the results in Table 3.4, Table 3.5 and Table 3.6, we see that the model selection criteria GGC and deviance work well and select the right distribution. A brief description of GGC and Dev is given in Section 2.3. While fitting the WN, WC and WDE distributions, the right distribution has the lowest GGC and Dev value. Also, in this case β_0 , β_1 and ρ have been well estimated. In general, the location parameters β_0 and β_1 are estimated well even when the models are erroneous.

In Table 3.4, Table 3.5 and Table 3.6, we see WN model estimates the location parameters well, even when the distribution is not correctly specified. However, the lower than nominal coverage probability of ρ , indicates that WN is not robust in estimating the mean resultant length of the distribution. In the case of WC model, the location parameters are estimated well, but the mean resultant is not well estimated. One of the reasons is that we use a high mean resultant value ($\rho = 0.8$) for simulation and Cauchy (or WC) being a flat distribution is unable to fit the simulated data well. For WDE model, even when the distribution is incorrect, the parameters β_0 , β_1 and ρ have been well estimated. Therefore, we see that the WDE model is robust in estimating β_0 , β_1 and ρ .

We generate data from the Von Mises distribution on the circle $(0, 2\pi)$. To generate

Table 3.4: Regressing WN, WC and WDE models to WN data with $\beta_1 = 1.5$

<i>WN</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.04	0.15	0.77	1.03	1.36	0.95
s.e.	0.40	0.19	0.37	0.39	0.71	0.01
β_1	1.47	0.17	1.13	1.47	1.79	0.97
s.e.	0.69	0.33	1.03	0.71	0.77	0.01
ρ	0.77	0.06	0.64	0.78	0.86	0.93
s.e.	0.07	0.03	0.13	0.07	0.05	0.01
GGC	152.22	44.33	70.12	150.67	243.12	
s.e.	23.69	2.33	20.80	24.05	25.61	
Dev	64.28	2.96	61.02	63.43	71.72	
s.e.	9.35	2.20	8.98	9.24	10.87	
<i>WC</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.03	0.14	0.75	1.04	1.31	0.89
s.e.	0.44	0.15	0.31	0.49	0.56	0.01
β_1	1.49	0.17	1.16	1.49	1.80	0.92
s.e.	0.23	0.35	0.87	0.16	0.59	0.01
ρ	0.65	0.07	0.50	0.65	0.77	0.27
s.e.	0.07	0.01	0.09	0.07	0.05	0.02
GGC	160.22	44.21	78.62	158.59	251.18	
s.e.	23.23	2.14	20.17	23.56	25.15	
Dev	71.05	2.70	67.88	70.37	77.93	
s.e.	9.06	0.72	8.76	9.07	9.24	
<i>WDE</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.02	0.14	0.76	1.01	1.30	0.92
s.e.	0.30	0.16	0.29	0.30	0.52	0.01
β_1	1.47	0.16	1.14	1.48	1.76	0.94
s.e.	0.40	0.33	1.05	0.32	0.24	0.01
ρ	0.75	0.07	0.60	0.76	0.86	0.96
s.e.	0.06	0.02	0.10	0.07	0.04	0.01
GGC	152.09	43.83	71.43	150.38	242.38	
s.e.	24.82	2.11	21.85	25.20	26.34	
Dev	66.83	3.01	62.57	66.35	74.08	
s.e.	9.05	1.48	9.01	9.19	9.63	

Table 3.5: Regressing WN, WC and WDE models to WC data with $\beta_1 = 1.5$

<i>WN</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.04	0.16	0.72	1.04	1.36	0.96
s.e.	0.43	0.18	0.33	0.49	0.59	0.01
β_1	1.46	0.18	1.12	1.47	1.79	0.95
s.e.	0.46	0.26	0.89	0.39	0.34	0.01
ρ	0.74	0.06	0.60	0.75	0.83	0.59
s.e.	0.12	0.02	0.16	0.11	0.08	0.02
GGC	152.27	45.18	69.09	150.53	245.07	
s.e.	29.70	2.86	25.44	30.00	33.67	
Dev	67.07	2.72	63.95	66.38	74.11	
s.e.	17.23	0.77	17.10	17.25	17.46	
<i>WC</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.00	0.07	0.87	1.00	1.15	0.94
s.e.	0.11	0.11	0.08	0.09	0.29	0.01
β_1	1.49	0.08	1.31	1.50	1.66	0.95
s.e.	0.11	0.25	0.71	0.07	0.42	0.01
ρ	0.78	0.05	0.67	0.79	0.87	0.95
s.e.	0.06	0.02	0.08	0.06	0.03	0.01
GGC	131.97	40.72	58.76	129.87	217.31	
s.e.	26.52	2.76	22.36	26.92	29.29	
Dev	50.36	2.75	47.16	49.66	57.35	
s.e.	14.43	1.07	14.21	14.45	14.68	
<i>WDE</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.04	0.11	0.82	1.03	1.25	0.97
s.e.	0.35	0.18	0.10	0.37	0.61	0.01
β_1	1.47	0.15	1.17	1.48	1.73	0.96
s.e.	0.31	0.46	1.22	0.20	0.59	0.01
ρ	0.79	0.06	0.66	0.80	0.89	0.83
s.e.	0.10	0.03	0.13	0.11	0.07	0.02
GGC	138.21	42.93	60.00	136.20	227.33	
s.e.	29.32	2.70	25.36	29.71	32.63	
Dev	56.45	3.14	51.85	56.04	63.88	
s.e.	15.91	1.36	16.19	16.01	15.92	

Table 3.6: Regressing WN, WC and WDE models to WDE data with $\beta_1 = 1.5$

<i>WN</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.07	0.15	0.77	1.08	1.36	0.94
s.e.	0.57	0.19	0.43	0.63	0.71	0.01
β_1	1.42	0.18	1.09	1.43	1.75	0.95
s.e.	0.62	0.39	1.15	0.74	0.32	0.01
ρ	0.76	0.06	0.63	0.77	0.85	0.77
s.e.	0.10	0.02	0.14	0.10	0.06	0.02
GGC	151.30	44.52	69.09	149.70	242.65	
s.e.	26.19	2.37	22.72	26.54	28.48	
Dev	65.28	2.74	62.05	64.64	72.19	
s.e.	13.34	1.01	12.91	13.47	13.47	
<i>WC</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.04	0.11	0.82	1.04	1.27	0.92
s.e.	0.39	0.13	0.32	0.41	0.55	0.01
β_1	1.45	0.14	1.17	1.45	1.70	0.93
s.e.	0.56	0.31	1.01	0.57	0.61	0.01
ρ	0.70	0.06	0.56	0.71	0.81	0.62
s.e.	0.07	0.02	0.10	0.07	0.05	0.02
GGC	149.83	43.18	70.78	148.06	239.19	
s.e.	24.85	2.31	21.35	25.26	26.85	
Dev	64.20	2.79	60.95	63.53	71.35	
s.e.	11.46	1.15	11.12	11.59	12.08	
<i>WDE</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.02	0.12	0.79	1.02	1.25	0.95
s.e.	0.29	0.15	0.25	0.28	0.46	0.01
β_1	1.48	0.13	1.22	1.48	1.73	0.95
s.e.	0.16	0.26	0.61	0.15	0.42	0.01
ρ	0.77	0.06	0.63	0.78	0.87	0.93
s.e.	0.08	0.02	0.10	0.08	0.06	0.01
GGC	145.64	43.43	66.02	143.85	235.47	
s.e.	26.09	2.26	22.75	26.42	28.13	
Dev	62.65	3.01	58.32	62.18	69.92	
s.e.	11.56	1.23	11.75	11.59	11.75	

Table 3.7: Regressing WN, WC and WDE models to VM data with $\beta_1 = 1.5$

<i>WN</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.01	0.13	0.75	1.02	1.27	0.96
s.e.	0.22	0.08	0.13	0.26	0.26	0.01
β_1	1.50	0.13	1.23	1.50	1.76	0.95
s.e.	0.13	0.09	0.35	0.13	0.14	0.01
ρ	0.78	0.05	0.65	0.78	0.86	0.90
s.e.	0.06	0.02	0.10	0.06	0.04	0.01
GGC	152.23	44.17	70.28	150.78	242.70	
s.e.	25.17	2.23	22.01	25.46	27.24	
Dev	63.79	2.68	60.71	63.11	70.71	
s.e.	9.81	0.58	9.67	9.81	10.05	
<i>WC</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.00	0.13	0.76	1.00	1.25	0.90
s.e.	0.15	0.07	0.15	0.14	0.28	0.01
β_1	1.50	0.14	1.23	1.50	1.78	0.91
s.e.	0.19	0.21	0.31	0.14	0.61	0.01
ρ	0.67	0.07	0.52	0.67	0.78	0.35
s.e.	0.05	0.02	0.08	0.05	0.04	0.02
GGC	158.75	43.94	77.63	157.13	249.15	
s.e.	24.75	2.17	21.31	25.17	26.55	
Dev	68.72	2.75	65.60	67.98	75.73	
s.e.	8.50	1.34	8.43	8.43	9.16	
<i>WDE</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.02	0.13	0.77	1.02	1.28	0.92
s.e.	0.26	0.13	0.17	0.30	0.44	0.01
β_1	1.46	0.14	1.17	1.47	1.73	0.93
s.e.	0.53	0.24	0.95	0.51	0.54	0.01
ρ	0.76	0.07	0.60	0.77	0.87	0.96
s.e.	0.07	0.02	0.10	0.06	0.04	0.01
GGC	151.90	43.65	71.51	150.24	241.81	
s.e.	26.70	2.35	23.56	27.09	28.50	
Dev	65.34	3.04	61.02	64.79	72.71	
s.e.	9.54	1.64	9.55	9.53	10.21	

the data from the Von Mises distribution, we use the algorithm given by Best and Fisher (1978). As before, we set $\beta_0 = 1$, $\beta_1 = 1.5$ and $\rho = 0.8$. We fit Wrapped Normal, Wrapped Cauchy and Wrapped Double Exponential distributions to Von Mises data and repeat the procedure 500 times to study frequentist performance.

In Table 3.7, we see that WN is robust in estimating the β_0 , β_1 and ρ , when the data are generated from the Von Mises distribution. This is expected as WN closely approximates the Von Mises distribution. In the case of WC model, the low coverage probability of ρ indicates that WC is not robust in estimating the mean resultant length of the distribution. However, the location parameter is estimated well. For WDE model, the parameters μ and ρ have been well estimated. The WDE model is robust in estimating the β_0 , β_1 and ρ . GGC selects WDE model as the best fit for VM data. However, Dev selects WN model as the best fit for VM data.

We extend the linear regression model by introducing an additional categorical covariate. The results are similar and are shown in Appendix D. We introduce regression diagnostics plots for circular data in the next section.

3.5 Regression Diagnostic plots

In this section, we illustrate the working of the regression method through a simulation study. Let $\mathbf{u} = \{u_1, u_2, \dots, u_n\}$ be a sample of size $n = 50$, generated from WN with parameters $\beta_0 = 3$, $\beta_1 = -6.5$ and $\rho = 0.8$. The original mean $\boldsymbol{\mu} = \beta_0 + \beta_1 \mathbf{x}$, is represented by the line drawn in Figure 3.2. The observations for covariate \mathbf{x} , were the distance measurements from the movement of periwinkle data (Fisher, 1993). The covariates were

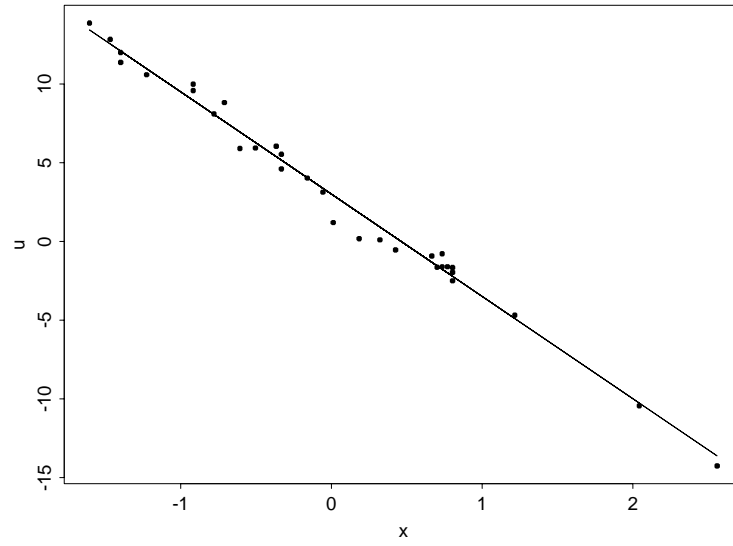


Figure 3.2: Plot of the original values.

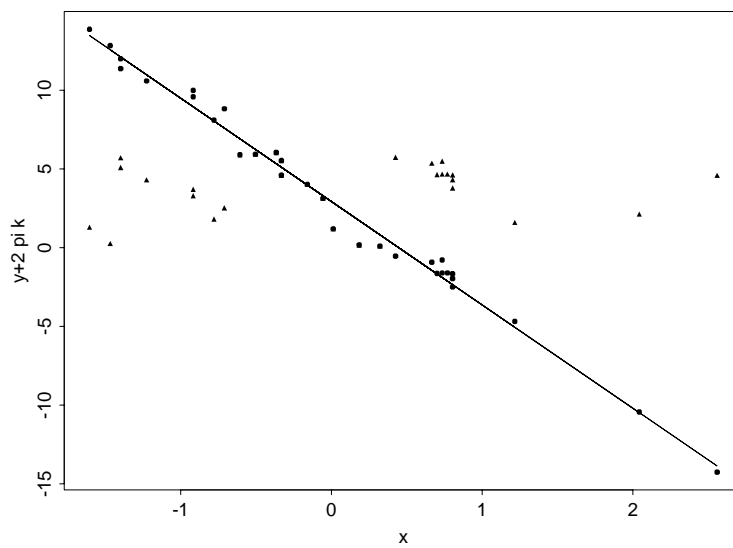


Figure 3.3: Regression Plot of the sample and fitted values.

centered and scaled. In Figure 3.2, the original sample on the real line, \mathbf{u} is plotted as dots along the y-axis and the covariates \mathbf{x} , is plotted along the x-axis. We then wrap the values of \mathbf{u} onto the circle, and obtain values of $\mathbf{y} = \{y_1, y_2, \dots, y_n\}$ between 0 and 2π . That is, $y_i = u_i(\text{mod } 2\pi)$, $1 \leq i \leq n$. \mathbf{y} values are shown as triangles in Figure 3.3.

We consider \mathbf{y} as the sample circular data. We fit a linear WN regression model to \mathbf{y} . We compute the Bayesian posterior mean values, $\hat{\beta}_0$ and $\hat{\beta}_1$ for β_0 and β_1 respectively. The line in Figure 3.3 represents the estimated line $\hat{\beta}_0 + \hat{\beta}_1\mathbf{x}$. In addition, let \hat{k}_i be the Bayesian posterior mean value of the unknown wrapping numbers, $k_i, 1 \leq i \leq n$. Let $\hat{\mathbf{k}} = \{\hat{k}_1, \hat{k}_2, \dots, \hat{k}_n\}$. In Figure 3.3, we plot the values of $\mathbf{y} + 2\pi\hat{\mathbf{k}}$ as dots along the y-axis and the covariates \mathbf{x} , along the x-axis.

Thus, from Figure 3.2 and Figure 3.3, we see that the two plots are very similar. We can now compute residuals, defined as $\mathbf{y} + 2\pi\hat{\mathbf{k}} - \hat{\beta}_0 - \hat{\beta}_1\mathbf{x}$. Therefore, we can perform regression diagnostic methods (Chatterjee, Hadi & Price, 1999) for circular data to analyze the performance of the fitted model. This is significant, as this is the first time regression diagnostics methods have been performed on circular data. In the next section, we analyze a real data set, and show a residual plot in Figure 3.7.

3.6 Application to real data sets

We analyze a data set containing the measurements of ozone concentration and wind direction. Measurements were taken at 6:00 a.m. at four-day intervals between April 18th and June 29th 1975, at a weather station in Milwaukee (Fisher, 1993). We want to study the relationship between the ozone concentration and the wind direction. For this

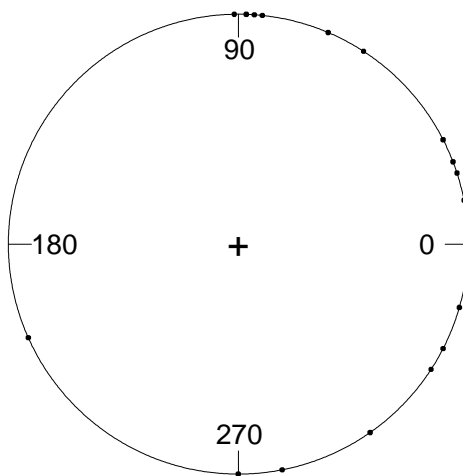


Figure 3.4: Plot of wind directions in the ozone data.

data set, the sample mean and the standard deviation for ozone concentration are 51.2 and 29.37 respectively. The circular sample mean and the resultant for wind direction data are 6.57 radians (16°) and 0.52 respectively.

Figure 3.1 shows the joint data plot of the ozone concentration and wind direction. Figure 3.4 shows the circular histogram plot of the wind directions. We fit WN, WC and WDE distributions to this data. We compute the posterior mean, standard deviation, 2.5 percentile, median and 97.5 percentile for β_0 , β_1 and ρ . The percentiles are computed with $\pi \approx 3.14$ radians (180°) as the reference point. We have used a class of non-informative priors for (β_0, β_1, ρ) , which is given in equation (3.2), but present the results only for the case where $a_\rho = 0.5$, as given in Table 3.8.

While fitting the data, we rejected the first 2000 samples (burn-in period) of the MCMC chain and kept the next 5000 samples after burn-in to perform posterior inference.

There were difficulties in achieving convergence using CODA. Therefore, we increased the burn-in time to 5000 samples and kept the next 15000 samples. We used the sample estimates of the mean and mean resultant as starting values of β_0 and ρ respectively. The starting value of β_1 is 0. We then change our starting values for MCMC (for example, $\beta_0 = 0$ and $\rho = 0.98$) and overlay this Markov Chain with the previous Markov Chain. We find that after 3000 samples they both overlap each other and look random. Analyzing similar results (not given here), there is no indication of MCMC convergence problem. Summary values for the parameters are given in Table 3.8.

To compare the models we use GGC and Dev. Low values of GGC and Dev indicates better fit for a model. From Table 3.8, we see that the WN distribution performs best based on the GGC. WDE estimates are close to WC estimates. However, the estimates from WN are different from WC and WDE. Dev value for WC is slightly less than WN, but the GGC value of WN is significantly lower. We recommend the WN for this data set. Using CODA, we obtained the trace plots and the kernel density estimates of the parameters β_0 and β_1 based on the Wrapped Normal (WN) distribution shown in Figure 3.5.

This data has been previously analyzed by Fisher (1993). He finds that there is some association between the two variables. We find that both parameters β_0 and β_1 are significant for the linear model. This is consistent with the results of Fisher (1993).

We would also like to compare this regression model with the previous model we fitted in Section 2.7.2. In the previous model, we used only the wind direction and did not use the ozone concentration data. Comparing the results in Table 3.8 with Table 2.11, we see that for all the regression models WN, WC and WDE, the GGC and Dev values are

Table 3.8: Regressing WN, WC and WDE models to the Ozone data set with $\rho \sim \text{Beta}(0.5, 0.5)$

<i>WN</i>	mean	std dev	2.5 %	50 %	97.5 %
β_0	6.39	0.24	5.94	6.39	6.88
β_1	0.81	0.27	0.32	0.80	1.42
ρ	0.62	0.11	0.35	0.63	0.79
GGC	44.16	21.51	16.47	39.30	98.16
Dev	51.65	3.03	48.26	50.87	59.71
<i>WC</i>	mean	std dev	2.5 %	50 %	97.5 %
β_0	6.55	0.22	6.14	6.55	6.93
β_1	1.04	0.32	0.42	1.04	1.60
ρ	0.58	0.13	0.25	0.61	0.76
GGC	54.27	23.47	18.55	50.93	108.66
Dev	51.53	3.65	47.82	50.45	61.66
<i>WDE</i>	mean	std dev	2.5 %	50 %	97.5 %
β_0	6.43	0.21	5.99	6.43	6.83
β_1	1.02	0.34	0.31	1.04	1.66
ρ	0.60	0.12	0.35	0.61	0.80
GGC	52.13	22.73	18.82	48.62	104.81
Dev	53.93	2.67	50.42	53.32	60.68

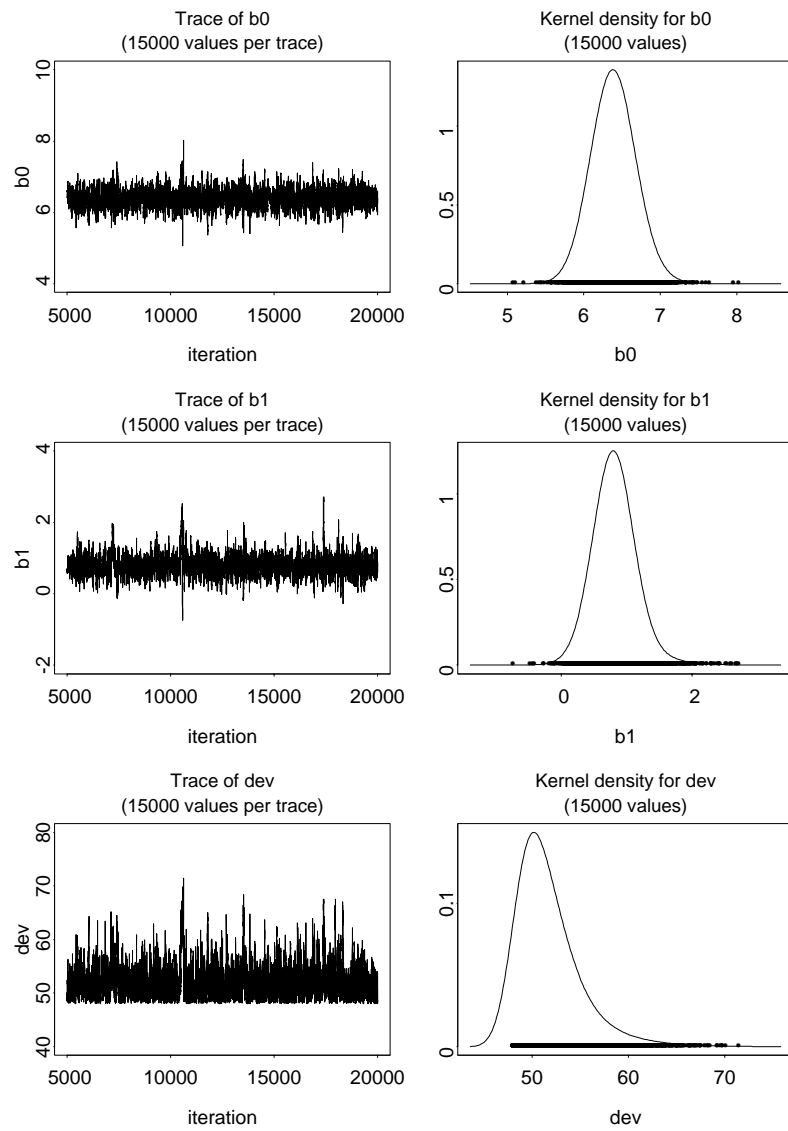


Figure 3.5: Trace plots while fitting WN to the Ozone data

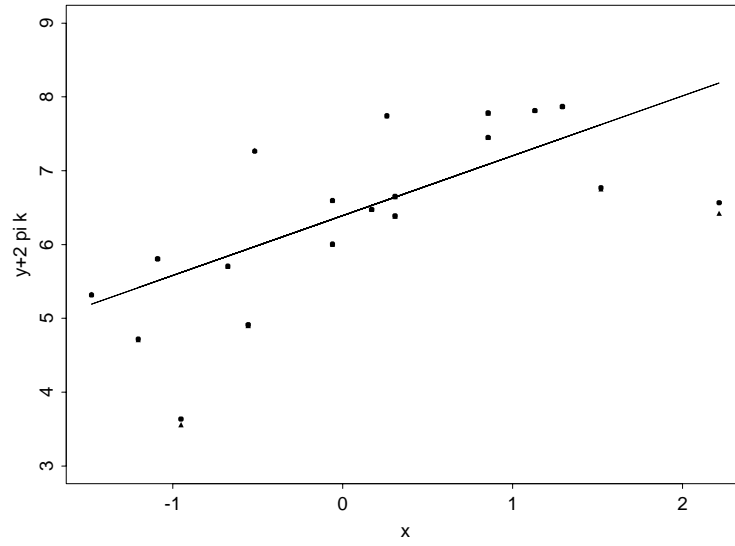


Figure 3.6: Regression Plot of the fitted values.

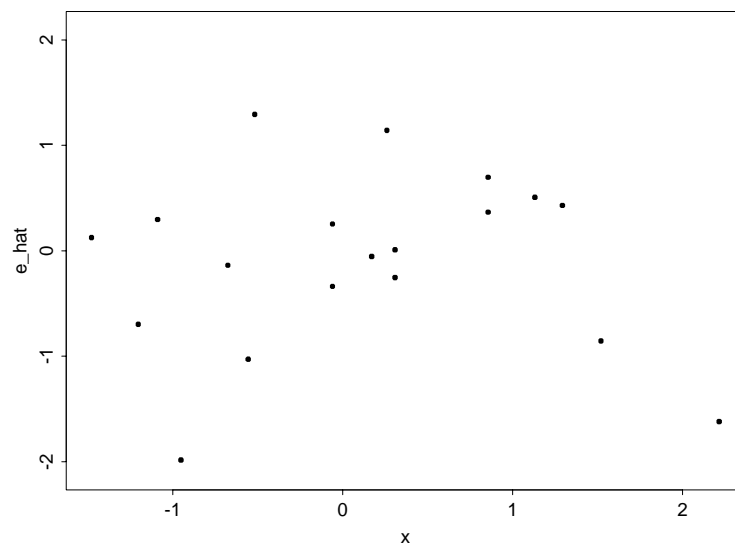


Figure 3.7: Plot of the residuals.

low. The standard deviation for the parameters are also low. This shows that the regression model is a better fit for the ozone data, which is expected. We also give the regression plot for the data in Figure 3.6 and the plot of the corresponding residuals in Figure 3.7. The plots show that the linear model is a good model for this data.

3.7 Discussion

In this chapter, we have extended the methods of parameter estimation to the regression model. We have shown the advantages of the wrapping method to the other methods. We can work with a much larger class of families instead of restricting ourselves to Von Mises and Projected Normal distributions. The posterior summary statistics for β_0 , β_1 and ρ can be easily computed. We are able to illustrate how the circular regression coefficients can be interpreted in terms of the linear regression model setup. Our method is able to provide regression diagnostics for circular data. This is significant, as this is the first time regression diagnostics methods have been performed on circular data. We agree with the findings of the previous chapter that the parameter estimates based on WDE model are robust for erroneous models. More simulation results with an additional categorical covariate are given in Appendix D. We extend this method to time series analysis in the next chapter.

Chapter 4

Circular Time Series

4.1 Introduction

Let us look at circular random variables that are time dependent or with a one dimensional spatial structure. Many examples occur in the fields of meteorology and oceanography. A popular example is the hourly or daily measurements of wind direction. A study was performed at a site on Black Mountain, ACT, Australia (Cameron, 1983). 72 observations were collected on wind direction to calibrate three anemometers (instrument used to measure wind speed). The objective was to fit the data by a simple model, to forecast future wind patterns. More details about this data is given in Section 4.5.

the one dimensional spatial model occurs in structural geology. A common example is mapping the orientation of rock structures at regular intervals to look for an overall trend, or for sudden changes in the general trend. 63 observations of median direction of face cleat were collected at the coal mine, in Wallsend Boreland Colliery, NSW, Australia (Shepherd and Fisher, 1981, 1982). These observations are shown in Figure 4.1. Face cleat is the major

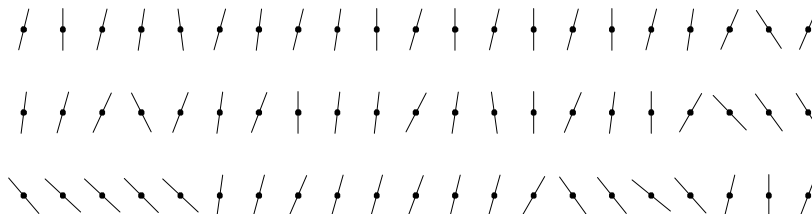


Figure 4.1: Plot of the median direction of face cleat collected at Wallsend Boreland Colliery, NSW in Australia.

joint or cleavage system in a coal seam. The data were collected at 20-meter intervals along a coal seam. At each 20-meter interval, five observations were taken and their median was used as a sample. In many coal mines, a shift from the preferred direction is an indicator of the possibility of hazardous mining conditions, that can occur due to forthcoming faults. A simple model is needed to analyze the data, so that data can be gathered and immediately analyzed, to give timely warning about potential hazards. The distinction of this data from the usual time series data is that it can be analyzed in both directions.

4.2 Previous work

Many methods have been suggested to implement time series models for circular data. Mardia and Jupp (1999) suggest a Projected Normal process approach. They consider $\{(X_t, Y_t)\}_{t=1,2,\dots}$ as a process on a plane. Let $\{\theta_t\}_{t=1,2,\dots}$ be the corresponding radial projection onto the unit circle, then $X_t = R_t \cos(\theta_t)$ and $Y_t = R_t \sin(\theta_t)$. Therefore, if $\{(X_t, Y_t)\}_{t=1,2,\dots}$ is a stationary bivariate Gaussian process, then θ_t has the Projected Normal distribution. They consider $\{R_t\}_{t=1,2,\dots}$ as not observed and use EM algorithm to fit the missing data problem.

Direct Linked process is another method to analyze time series data. This is an extension of the method suggested by Johnson and Wehrly (1978) and Fisher and Lee (1992) for regression using link functions. The link function is a one-to-one function $g(\cdot)$, which maps the entire real line onto $(-\pi, \pi)$ and satisfies $g(0) = 0$. Some examples are $g(x) = 2 \tan^{-1}(x)$ and $g(x) = 2\pi F(x)$, where $F(x)$ is a distribution function. The general idea is to transform the circular data to the real line using the link function and perform the time series analysis on the real line. This method works only for concentrated series, where the variance is small.

Circular Autoregressive process, CAR(p) is another approach that is constructed using conditional distributions. This approach is associated with the Von Mises model. If $g(\cdot)$ is a link function, then CAR(p) is defined as $y_t | (y_{t-1}, \dots, y_{t-p}) \sim VM(\mu_t, \kappa)$, where $\mu_t = \mu + g(\alpha_1 g^{-1}(y_{t-1} - \mu) + \dots + \alpha_p g^{-1}(y_{t-p} - \mu))$ and $VM(\cdot)$ is the Von Mises distribution. Accardi, Cabrera and Watson (1987) considered the case $p = 1$ and $\alpha_1 = 1$, where the model reduces to $y_t | y_{t-1} \sim VM(y_{t-1}, \kappa)$.

4.3 Extension of the Data Augmentation Approach for Time Series

We will be using the wrapped approach. Breckling (1989) gave the definitions for the Wrapped Autoregressive process, $WAR(p)$. Let $\{U_t\}_{t=1,2,\dots}$ be a process on the line. Then, the corresponding process on the circle by wrapping U_t is defined by

$$Y_t = U_t \pmod{2\pi}$$

$$\text{or } U_t = Y_t + 2\pi k_t,$$

where k_t is an unobserved integer. Fisher and Lee (1994) used the EM algorithm to fit such processes. However, the E-step involves the ratio of large infinite sums, which needs to be approximated at each step. This makes the algorithm computationally inefficient. In addition, the standard errors of the MLEs have to be evaluated based on large-sample theory. Coles (1998) also did work on WAR(p) using MCMC.

For the time series problem, we use the data augmentation approach in a Bayesian setup. We extend the theory given in Section 3.3 for time series. Let the observations on the circle be $\mathbf{y} = \{y_1, y_2, y_3, \dots, y_n\}$, $0 \leq y_j < 2\pi$, $j = 1 \dots n$. We define the regressor variables from the regression setup, $\mathbf{x} = \{x_1, x_2, x_3, \dots, x_n\}$, $j = 1 \dots n$ as $x_1 = 0$ and $x_j = y_{j-1} + 2\pi k_{j-1} - \beta_0$, $j = 2 \dots n$ to obtain the WAR(1) setup. Therefore, for the simple WAR(1) problem,

$$Y_j | X_j \sim WD(\mu_j, \rho),$$

$$\text{where } \mu_j = \beta_0 + \beta_1 x_j$$

$$\text{and } x_j = y_{j-1} + 2\pi k_{j-1} - \beta_0.$$

This can be extended to WAR(p), where $\mu_j = \beta_0 + \sum_{i=1}^p \beta_i (y_{j-i} + 2\pi k_{j-1} - \beta_0)$ and $\beta_0, \beta_1, \dots, \beta_p$ and ρ are parameters to be estimated. WD represents any wrapped distribution such as WN or WDE. Without loss of generality, we can restrict β_0 between 0 and 2π to avoid identifiability problems. The variance of Y_1 is defined slightly differently from the others to satisfy stationarity conditions.

It is not necessary to use a one-to-one link function, which transforms the real line onto $(-\pi, \pi)$ (Direct Linked Processes). This is because we are using \mathbf{y} and \mathbf{k} to fit the data rather than just \mathbf{y} . In order to illustrate the theory we assume that $p = 1$. However,

the extension to WAR(p) is straightforward.

A class of non-informative priors for (β_0, β_1, ρ) can be specified as,

$$[\beta_0, \beta_1, \rho] \propto I_{\beta_0}(0, 2\pi) I_{\beta_1}(-1, 1) \rho^{a_\rho-1} (1-\rho)^{a_\rho-1}, \quad a_\rho > 0 \quad (4.1)$$

and using $\rho = h(\sigma)$, the joint density of $(\beta_0, \beta_1, \sigma)$ is given by

$$[\beta_0, \beta_1, \sigma] \propto I_{\beta_0}(0, 2\pi) I_{\beta_1}(-1, 1) h(\sigma)^{a_\rho-1} (1-h(\sigma))^{a_\rho-1} |h'(\sigma)|, \quad a_\rho > 0.$$

Following the same approach described in Section 2.2 and replacing μ by $\beta_0 + \beta_1 x_j$,

the joint density of $\mathbf{y}, \mathbf{k}, \beta_0, \beta_1$ and σ is given by

$$\begin{aligned} [\mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \sigma] &\propto [\mathbf{y}|\mathbf{k}, \beta_0, \beta_1, \sigma^2] [\mathbf{k}|\beta_0, \beta_1, \sigma^2] [\beta_0, \beta_1, \sigma^2] \\ &\propto \frac{\sqrt{1-\beta_1^2}}{\sigma} f\left(\frac{y_1-\beta_0+2\pi k_1}{\sigma/\sqrt{1-\beta_1^2}}\right) \prod_{j=2}^n \left(\frac{1}{\sigma} f\left(\frac{y_j-\beta_0-\beta_1 x_j+2\pi k_j}{\sigma}\right)\right) \\ &\quad h(\sigma)^{a_\rho-1} (1-h(\sigma))^{a_\rho-1} |h'(\sigma)| I_{\beta_0}(0, 2\pi) I_{\beta_1}(-1, 1) \\ &\propto \frac{\sqrt{1-\beta_1^2}}{\sigma^{n+n_0}} f\left(\frac{y_1-\beta_0+2\pi k_1}{\sigma/\sqrt{1-\beta_1^2}}\right) \prod_{j=2}^n f\left(\frac{y_j-\beta_0-\beta_1 x_j+2\pi k_j}{\sigma}\right) \\ &\quad h(\sigma)^{a_\rho-1+n_1} (1-h(\sigma))^{a_\rho-1} h_1(\sigma) I_{\beta_0}(0, 2\pi) I_{\beta_1}(-1, 1), \end{aligned}$$

where $|h'(\sigma)|$ can be factorized as $\frac{1}{\sigma^{n_0}} h(\sigma)^{n_1} h_1(\sigma)$. It is assumed that $h_1(\sigma)$ is invertible.

The full conditional densities of $\mathbf{k}, \beta_0, \beta_1$ and σ are nonstandard densities. Therefore, we introduce auxiliary variables \mathbf{z} and \mathbf{v} . Let $\mathbf{z} = \{z_0, z_1, z_2, z_3, z_4\}$ and $\mathbf{v} = \{v_1, v_2, v_3, \dots, v_n\}$, such that

$$[\mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \sigma] \propto \int [\mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}, \mathbf{v}, \sigma] d\mathbf{z} d\mathbf{v}.$$

The joint density of $\mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}, \mathbf{v}$ and σ is given by

$$[\mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}, \mathbf{v}, \sigma] \propto I_{v_1}\left(0, f\left(\frac{y_1-\beta_0+2\pi k_1}{\sigma/\sqrt{1-\beta_1^2}}\right)\right) \prod_{j=2}^n I_{v_j}\left(0, f\left(\frac{y_j-\beta_0-\beta_1 x_j+2\pi k_j}{\sigma}\right)\right)$$

$$\begin{aligned}
& I_{z_0} \left(0, \frac{1}{\sigma^{n+n_0}} \right) I_{z_1} \left(0, h(\sigma)^{a_\rho-1+n_1} \right) I_{z_3} (0, h_1(\sigma)) \\
& \{ I_{z_2} \left(0, (1-h(\sigma))^{a_\rho-1} \right) I(a_\rho \neq 1) + I(a_\rho = 1) \} \\
& I_{z_4} \left(0, \sqrt{1-\beta_1^2} \right) I_{\beta_0} (0, 2\pi) I_{\beta_1} (-1, 1).
\end{aligned}$$

The implementation is similar to that described in Section 3.3. The main difference is that the full conditional of k_j is dependent on \mathbf{k}_{-j} . The full conditionals for the Wrapped Normal model are given in Appendix C.

We establish a relationship between the first order autocorrelation coefficient of the circle and autocorrelation coefficient of the real line. The definition for the correlation coefficient is not unique in the circular sample space. Mardia and Jupp (1999) provide a summary of various correlation coefficients. We will use the correlation coefficient defined by Fisher and Lee (1983). Given two circular random variables, Y_1 and Y_2 , the correlation coefficient is given by

$$\gamma = 2[R^2(Y_1 - Y_2) - R^2(Y_1 + Y_2)] / \sqrt{[1 - R^2(2Y_1)][1 - R^2(2Y_2)]}, \quad (4.2)$$

where $R^2(Y) \equiv [E(\cos Y)]^2 + [E(\sin Y)]^2$ is the square of the mean resultant for circular variable Y . Using this definition of correlation coefficient, Fisher and Lee (1994) showed that if U_t is a Gaussian AR(p) process, the circular autocorrelation function for the Wrapped Autoregressive Normal WARN(p) process, $Y_t = U_t(\text{mod } 2\pi)$ is given by

$$\gamma(k) = \frac{\sinh[2\gamma_k \sigma_0^2]}{\sinh[2\sigma_0^2]},$$

where γ_k is the k-lag autocorrelation function of U_t , and σ_0^2 is the marginal variance of U_t .

The marginal variance of U_t is

$$\sigma_0^2 = \text{Var}(U_t) = \sigma^2 / (1 - \beta_1 \gamma_1 - \dots - \beta_p \gamma_p),$$

where $\sigma^2, \beta_1, \dots, \beta_p$ are the AR(p) parameters of U_t . For the WARN(1) process this simplifies to

$$\gamma(1) = \frac{\sinh[2\beta_1\sigma_0^2]}{\sinh[2\sigma_0^2]},$$

$$\text{where } \sigma_0^2 = \sigma^2/(1 - \beta_1^2).$$

We studied the relation between β_1 and $\gamma(1)$ for WARN(1) and found that they have similar values for small values of σ . As the value of σ increases, circular autocorrelation function, $\gamma(1)$ tends to 0 irrespective of the value of β_1 . This is expected as the value of σ increases, we obtain a Uniform distribution on the circle. The relationship is shown in Figure 4.2 for $\rho = 0.8$ and $\rho = 0.5$ for WARN(1) process. The variable along the x-axis is the autocorrelation function and the variable along the y-axis is the circular autocorrelation function. For $\rho = 0.5$, the line is more flat because the value of σ is larger.

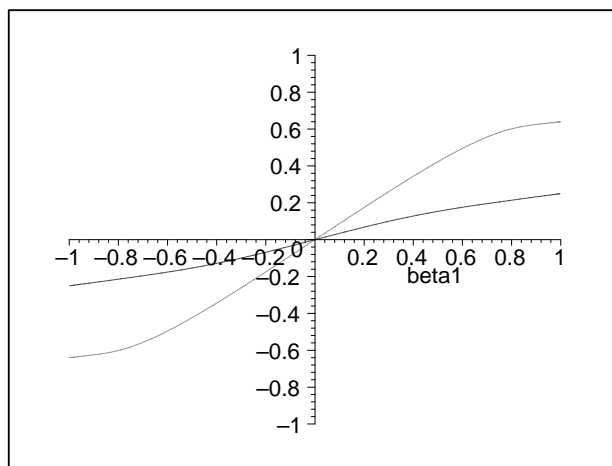


Figure 4.2: Relation between circular autocorrelation function and autocorrelation function

The corresponding sample circular correlation for (4.2) was given by Fisher and Lee (1983). They defined the sample circular correlation between $\mathbf{y}_1 = \{y_{11}, y_{12}, \dots, y_{1n}\}$

and $\mathbf{y}_2 = \{y_{21}, y_{22}, \dots, y_{2n}\}$ as

$$\hat{\gamma} = \frac{\sum \sin(y_{1i} - y_{1j}) \sin(y_{2i} - y_{2j})}{\sqrt{\sum \sin^2(y_{1i} - y_{1j})} \sqrt{\sum \sin^2(y_{2i} - y_{2j})}}, \quad (4.3)$$

where each summation is over the range $1 \leq i < j \leq n$. We use this to compute the sample circular autocorrelation at lag k . Given time series data, $\mathbf{y} = \{y_1, y_2, \dots, y_n\}$, define $\mathbf{y}_1 = \{y_1, y_2, \dots, y_{n-k}\}$ and $\mathbf{y}_2 = \{y_{k+1}, y_{k+2}, \dots, y_n\}$. The sample circular correlation between \mathbf{y}_1 and \mathbf{y}_2 gives the k -lag sample circular autocorrelation for \mathbf{y} .

Subsequently we study the performance of the proposed method by simulation experiments. We illustrate the above technique for the Wrapped Normal distribution.

4.4 Simulation studies

For our simulation studies, we generate samples of size $n = 50$ from the Wrapped Autoregressive Normal WARN(1) distribution, with parameters set at $\beta_0 = \pi = 3.14$, $\beta_1 = 0.7$ and $\rho = 0.8$. We fit WARN(1) with a prior given in equation (4.1) with $a_\rho = 0.5$. We compute the posterior mean, standard deviation, 2.5 percentile, median and 97.5 percentile for β_0 , β_1 and ρ for each simulation. The percentiles are computed with 0 radians as the reference point. The simulation standard errors for each of these summary values are also computed. We also compute the coverage probability for the 95% posterior interval given by the 2.5 and 97.5 percentile of the posterior distribution. In each simulation, we choose the burn-in period to be 2000 samples (i.e. throw away first 2000 samples from the MCMC chain) and then keep 5000 samples after burn-in, to obtain posterior summary values. All summary values in Table 4.1 are based on these final 5000 samples. We repeat the entire procedure 500 times to see the frequentist performance of the proposed Bayes method.

Table 4.1: Regressing WARN(1) model to WARN(1) data

	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	3.13	0.42	2.28	3.13	3.99	0.96
s.e.	0.31	0.17	0.50	0.31	0.49	0.01
β_1	0.67	0.11	0.44	0.67	0.89	0.96
s.e.	0.12	0.02	0.15	0.12	0.08	0.01
ρ	0.79	0.04	0.70	0.79	0.85	0.95
s.e.	0.04	0.01	0.05	0.04	0.03	0.01

Using SAS, on a pentium III machine, on an average it took about 2 hours to perform the entire simulation for a given wrapped distribution.

From Table 4.1, we see that the proposed method performs well in terms of maintaining the nominal coverage probability. Also, the posterior mean and median can serve as a good point estimates of the parameters.

4.5 Application to real data sets

We analyze a data set containing the hourly measurements of wind direction. This study was conducted over three days at Black Mountain, ACT, Australia. The wind directions are plotted in Figure 4.3. The observations on wind direction were collected to calibrate three anemometers. The circular sample mean and the resultant for wind direction are 5.08 radians (291°) and 0.68 respectively.

We fit WARN(1) distribution to the wind direction data. We compute the posterior mean, standard deviation, 2.5 percentile, median and 97.5 percentile for β_0 , β_1 , $\gamma(1)$ and ρ . The percentiles are computed with 0 radians as the reference point. We have used a class of non-informative priors for (β_0, β_1, ρ) , which is given in equation (4.1). The results are

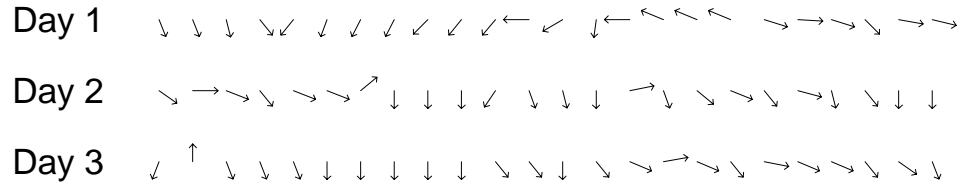


Figure 4.3: Plot of hourly wind directions collected over three days.

Table 4.2: Fitting WARN(1) to the wind direction data set with $\rho \sim \text{Beta}(0.5, 0.5)$

$a_\rho = 0.5$	mean	std dev	2.5 %	50 %	97.5 %
β_0	5.00	0.24	4.53	5.00	5.50
β_1	0.54	0.11	0.33	0.54	0.75
$\gamma(1)$	0.37	0.07	0.23	0.37	0.50
ρ	0.72	0.04	0.63	0.72	0.79

given in Table 4.2. The full conditionals obtained during the implementation of MCMC are given in Appendix C.

While fitting the wind direction data, we generated 7000 samples. Using Heidelberger and Welch Stationarity test (1983) and other tests in CODA software, we find that the burn-in period for the chain is 700. Therefore, we reject the first 2100 samples (burn-in period) and keep the remaining 4900 samples to perform posterior inference. Summary values for the parameters are given in Table 4.2. Using CODA, we also obtained the trace plots and the kernel density estimates of the parameters β_0 , β_1 , $\gamma(1)$ and ρ based on WARN(1) distribution, which are shown in Figure 4.4.

The wind direction data has been previously analyzed by Fisher (1993) by two different methods using Von Mises models. The parameter estimates for his models are ($\beta_0 = 5.08$, $\beta_1 = 0.5$ and $\rho = 0.68$) and ($\beta_0 = 5.05$, $\beta_1 = 0.68$ and $\rho = 0.76$). Therefore,

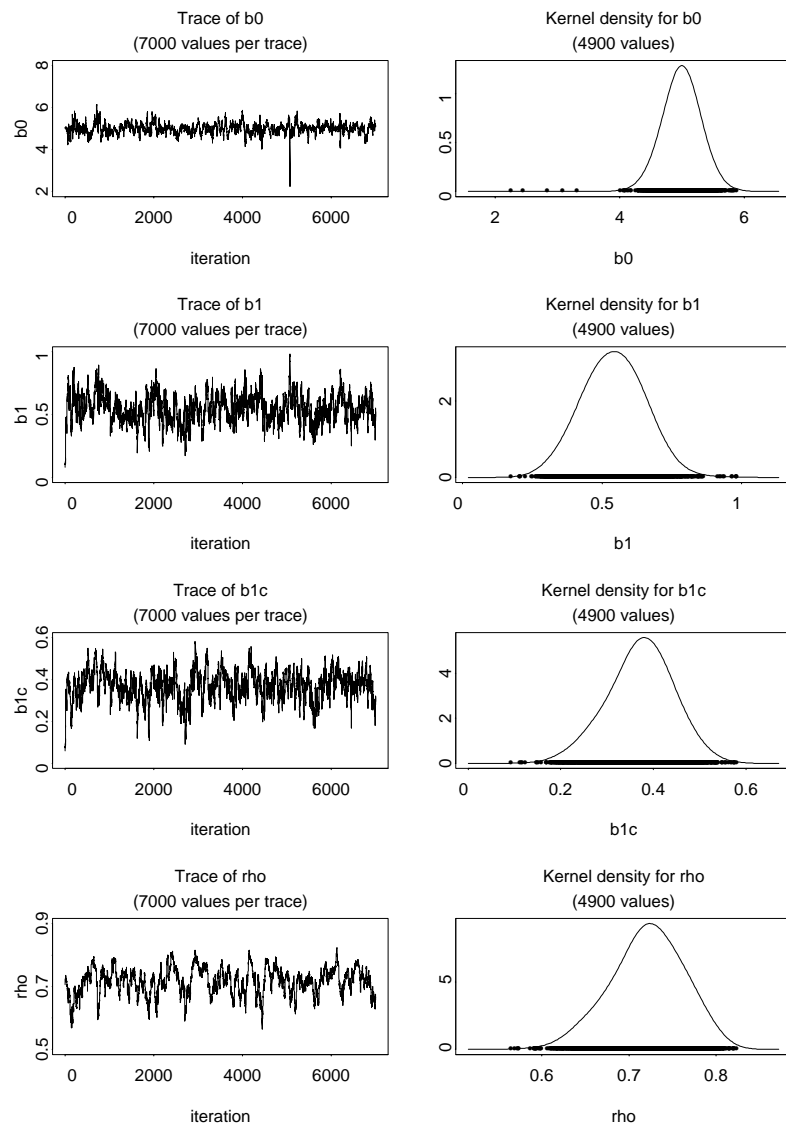


Figure 4.4: Trace plots while fitting WARN(1) to Wind direction data using time series model

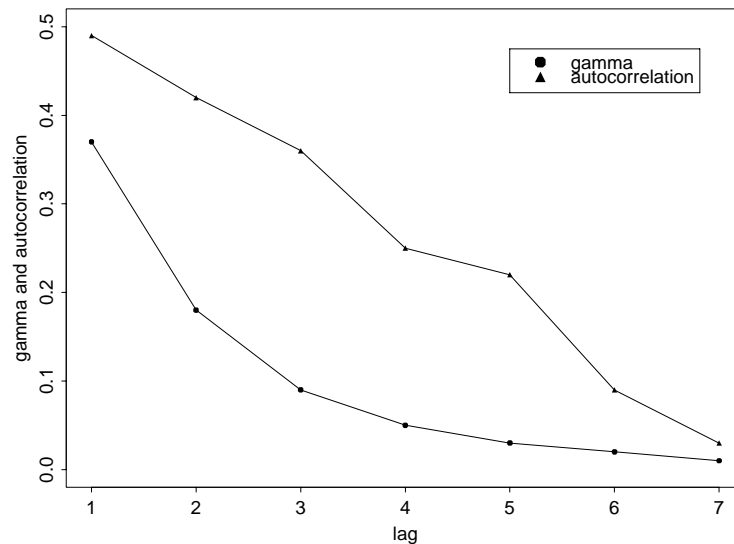


Figure 4.5: Plot of gamma, circular autocorrelation for WARN(1) model and sample circular autocorrelation at lags 1 to 7.

the results we obtained are consistent with the results of Fisher (1993). This is expected as Von Mises and Wrapped Normal densities are very similar.

We computed the circular autocorrelation functions, $\gamma(1), \gamma(2), \dots, \gamma(7)$ for the wind direction data set. We also computed the sample circular autocorrelation functions at lag 1 to lag 7 given by (4.3). This is shown in Figure 4.5. From the figure, we see that the circular autocorrelation for WARN(1) model is smaller than the sample circular autocorrelation, which suggests that we should try to fit higher order WARN models.

4.6 Discussion

In this chapter, we extended the regression methods from the previous chapter to fit time series models. Though we have discussed only WAR(1) in detail, this method can be easily extended to any WAR(p) model. The posterior summary statistics for β_0, β_1 and

ρ are easily obtained. Previously, Von Mises distribution was the most popular distribution to analyze circular time series data. Now, we can work with a much larger class of circular distributions. In addition, as we are using the Bayesian setup, we can easily compute other statistics such as circular autocorrelation functions and their standard errors.

Bibliography

- [1] Accardi, L., Cabrera, J. & Watson, G. S. (1987) Some stationary Markov processes in discrete time for unit vectors. *Metron*, **45**, 115-133.

- [2] Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. In *Proceedings of the Second International Symposium on Information Theory*, B. N. Petrov and F. Csaki (eds), 267-281. Akademiai Kiado, Budapest.

- [3] Best, D. J. & Fisher, N. I. (1978) Efficient simulation of the Von Mises distribution. *Appl. Statist.*, **28**, 152-157.

- [4] Batschelet, E. (1981) *Circular statistics in biology.*, Academic Press, London.

- [5] Breckling, J. (1989) *Analysis of directional time series: Applications to wind speed and direction.*, Lecture Notes in Statistics **61**, Springer-Verlag, Berlin.

- [6] Cameron, M. A. (1983) The comparison of time series recorders. *Technometrics*, **25**, 9-22.

- [7] Chatterjee, S., Hadi A. S. & Price B. (1999) *Regression Analysis by Example.*, Wiley, New York.

- [8] Coles, S. (1998) Inference for circular distributions and processes. *Statistics and Computing*, **8**, 105-113.
- [9] Damien, P., Wakefield, J. & Walker, S. (1999) Gibbs sampling for Bayesian non-conjugate hierarchical models by using auxiliary variables. *J. R. Statist. Soc. B*, **61**, 331-344.
- [10] Damien, P. & Walker, S. (1999) A full Bayesian analysis of circular data using von Mises distribution. *The Canadian Journal of Statistics*, **27**, 291-298.
- [11] Fisher, N. I. (1993) *Statistical analysis of circular data.*, Cambridge Univ. Press, Cambridge.
- [12] Fisher, N. I. & Lee, A.J. (1983) A correlation coefficient for circular data. *Biometrika*, **70**, 327-332.
- [13] Fisher, N. I. & Lee, A.J. (1992) Regression models for an angular response. *Biometrics*, **48**, 665-677.
- [14] Fisher, N. I. & Lee, A.J. (1994) Time series analysis of circular data. *J. R. Statist. Soc. B*, **56**, 327-339.
- [15] Fraser, D. A. S. (1979) *Inference and Linear Models.*, McGraw-Hill, New York.
- [16] Gelfand, A. E. & Ghosh, S. K. (1998) Model choice: A minimum posterior predictive loss function. *Biometrika*, **85**, 1-11.
- [17] Gordon, A. D., Jupp, P. E. & Byrne, R. W. (1989) Construction and assessment of mental maps *British Journal of Mathematical & Statistical Psychology*, **42**, 169-182.

- [18] Gould, A. L. (1969) A regression technique for angular variates. *Biometrics*, **25**, 683-700.
- [19] Greenwood, J. A. & Durand, D. (1955) The distribution of length and components of the sum of n random unit vectors. *Ann. Math. Statist.*, **26**, 233-246.
- [20] Guttorp, P. & Lockhart, R. A. (1988) Finding the location of a signal: A Bayesian analysis. *J. Am. Statist. Assoc.*, **83**, 322-330.
- [21] Harrison, D. & Kanji, G. K. (1988) Development of analysis of variance for circular data. *Journal of Applied Statistics*, **15**, 197-224.
- [22] Harrison, D., Kanji, G. K. & Gadsden, R.J. (1986) Analysis of variance for circular data. *Journal of Applied Statistics*, **13**, 123-138.
- [23] Hastie, T.J. & Tibshirani, R. J. (1986) Generalized Additive Models. *Statistical Science*, **1**, 297-318.
- [24] Heidelberger, P. & Welch, P. (1983) Simulation run length control in the presence of an initial transient. *Operations Research*, **31**, 1109-1144.
- [25] Higdon, D. M. (1998) Auxiliary variable methods for Markov Chain Monte Carlo with applications. *J. Am. Statist. Assoc.*, **93**, 585-595.
- [26] Jander, R. (1957) Die optische Richtungsorientierung der roten Waldameise (*Formica rufa L.*). *Zeitschrift Fur Vergleichende Physiologie*, **40**, 162-238.
- [27] Johnson, R. A. & Wehrly T. (1977) Measures and models for angular correlation and angular-linear correlation. *J. R. Statist. Soc. B*, **39**, 222-229.

- [28] Johnson, R. A. & Wehrly T. (1978) Some angular-linear distribution and related regression models. *J. Am. Statist. Assoc.*, **73**, 602-606.
- [29] Kent, J. T. & Tyler, D. E. (1988) Maximum likelihood estimation for the wrapped Cauchy distribution. *Journal of Applied Statistics*, **15**, 247-254.
- [30] Laycock, P. J. (1975) Optimal design: regression models for directions. *Biometrika*, **62**, 305-311.
- [31] Lee, J. A. H. (1962) Seasonal variation in the clinical onset of leukaemia in young people. *British Medical Journal*, **1**, 1737-1738.
- [32] Lévy, P. (1939) L'addition des variables aléatoires définies sur une circonférence. *Bulletin de la Société mathématique de France*, **67**, 1-41.
- [33] Mardia, K. V. (1972) *Statistics of directional data.*, Academic Press, London.
- [34] Mardia, K. V. (1976) Linear-circular correlation coefficients and rhythmometry. *Biometrika*, **63**, 403-405.
- [35] Mardia, K. V. & Jupp, P. E. (1999) *Directional statistics.*, Wiley, Chichester.
- [36] McCullagh, P. & Nelder, J. A. (1989) *Generalized Linear Models, 2nd ed.*, Chapman and Hall, New York.
- [37] Monahan, J. F. & Boos, D. D.,(1992) Proper likelihoods for bayesian analysis. *Biometrika*, **79**, 271-278.
- [38] Pincus, H. J. (1953) The analysis of aggregates of orientation data in the earth sciences. *Journal of Geology*, **61**, 482-509.

- [39] Presnell, B., Morrison, S. P. & Littell, R. C. (1998) Projected multivariate linear models for directional data. *J. Am. Statist. Assoc.*, **93**, 1068-1077.
- [40] Proschan, M. A. & Follman, D. A. (1997) A restricted test of circadian rhythm. *J. Am. Statist. Assoc.*, **92**, 717-724.
- [41] Rayleigh, Lord (1919) On the problem of vibrations, and of random flights in one, two or three dimensions. *Philosophical Magazine*, **37**, 321-347.
- [42] Ross, H. E., Crickmar, S. D., Sills, N. V. & Owen, E. P. (1969) Orientation to the vertical in free divers. *Aerospace Medicine*, **40**, 728-732.
- [43] Schwartz, G (1978) Estimating the dimension of a model. *Annals of Statistics*, **6**, 461-464.
- [44] Sengupta, A & Pal, C (2001) On optimal tests for isotropy against the symmetric wrapped stable-circular uniform mixture family. *Journal of Applied Statistics*, **28**, 129-143.
- [45] Shepherd, J. & Fisher N. I. (1981) A rapid method of mapping fractures in mines. *Australian Coal Miner*, **3(8)**, 24-33.
- [46] Shepherd, J. & Fisher N. I. (1982) Rapid method of mapping fracture trends in collieries. *Trans. Soc. Min. Eng. AIME*, **270**, 1931-1932.
- [47] Stephens, M. A. (1969) *Techniques for directional data*. Technical Report 150, Department of Statistics, Stanford University.

- [48] Tanner, M. A. & Wong W. H. (1987) The calculation of posterior distributions by data augmentation. *J. Am. Statist. Assoc.*, **82**, 528-540.
- [49] van Dyk, D. A. & Meng, X. L. (2001) The art of data augmentation. *Journal of Computational and Graphical Statistics*, **10**, 1-50.
- [50] Von Mises, R. (1918) Über die “Ganzzahligkeit” der Atomgewicht und verwandte Fragen. *Physikalische Zeitschrift*, **19**, 490-500.
- [51] Watson, G. S. (1983) *Statistics on spheres* University of Arkansas Lecture Notes in the Mathematical Sciences, **6**.

Appendix A

Explicit full conditionals

A.1 Wrapped Normal Distribution

$$f(x) = \sqrt{\frac{1}{2\pi}} e^{-\frac{x^2}{2}}$$

$$h(\sigma) = e^{-\frac{\sigma^2}{2}}$$

$$h^{-1}(\rho) = \sqrt{-2\sigma^2 \log \rho}$$

$$n_0 = -1$$

$$n_1 = 1$$

$$gM_1(v_j) = \sqrt{-2 \log v_j}, \quad q = 1$$

$$gm_1(v_j) = -gM_1(v_j)$$

$$h_1(\sigma) \equiv 1, \text{ so } z_3, gm_l^*(z_3) \text{ and } gM_l^*(z_3) \text{ are not needed.}$$

$$h_m \left(\frac{y_j - \mu + 2\pi k_j}{g m_{I_0}(v_j)}, \frac{y_j - \mu + 2\pi k_j}{g M_{I_0}(v_j)} \right) = \frac{|y_j - \mu + 2\pi k_j|}{g M_{I_0}(v_j)}, h_M \text{ is not needed.}$$

The full conditional densities of \mathbf{k} , \mathbf{v} , \mathbf{z} , μ and σ^2 are given by

$$\begin{aligned} [k_j | \mathbf{y}, \mathbf{k}_{-j}, \mu, \mathbf{z}, \mathbf{v}, \sigma] &\propto I_{v_j} \left(0, e^{-\frac{(y_j - \mu + 2\pi k_j)^2}{2\sigma^2}} \right) \\ k_j | \mathbf{y}, \mathbf{k}_{-j}, \mu, \mathbf{z}, \mathbf{v}, \sigma &\sim DU \left[\left[\frac{1}{2\pi} \left(\mu - y_j - \sqrt{-2\sigma^2 \log v_j} \right) \right], \right. \\ &\quad \left. \left[\frac{1}{2\pi} \left(\mu - y_j + \sqrt{-2\sigma^2 \log v_j} \right) \right] \right], \end{aligned}$$

where DU stands for Discrete Uniform.

$$\begin{aligned} v_j | \mathbf{y}, \mathbf{k}, \mu, \mathbf{z}, \mathbf{v}_{-j}, \sigma &\sim U \left[0, e^{-\frac{(y_j - \mu + 2\pi k_j)^2}{2\sigma^2}} \right] \\ z_0 | \mathbf{y}, \mathbf{k}, \mu, \mathbf{z}_{-0}, \mathbf{v}, \sigma &\sim U \left[0, \frac{1}{\sigma^{n-1}} \right] \\ z_1 | \mathbf{y}, \mathbf{k}, \mu, \mathbf{z}_{-1}, \mathbf{v}, \sigma &\sim U \left[0, e^{-\frac{a_\rho \sigma^2}{2}} \right] \\ z_2 | \mathbf{y}, \mathbf{k}, \mu, \mathbf{z}_{-2}, \mathbf{v}, \sigma &\sim \begin{cases} U \left[0, \left(1 - e^{-\frac{\sigma^2}{2}} \right)^{a_\rho - 1} \right] & a_\rho \neq 1 \\ z_2 \text{ is not needed} & a_\rho = 1 \end{cases} \\ \mu | \mathbf{y}, \mathbf{k}, \mathbf{z}, \mathbf{v}, \sigma &\sim U [m_\mu, M_\mu], \end{aligned}$$

$$\begin{aligned} \text{where } m_\mu &= \max_{j=1}^n \left[y_j + 2\pi k_j - \sqrt{-2\sigma^2 \log v_j} \right] \vee 0 \\ M_\mu &= \min_{j=1}^n \left[y_j + 2\pi k_j + \sqrt{-2\sigma^2 \log v_j} \right] \wedge (2\pi) \end{aligned}$$

$$\sigma | \mathbf{y}, \mathbf{k}, \mu, \mathbf{z}, \mathbf{v} \sim U [m_\sigma, M_\sigma],$$

$$\text{where } m_\sigma = \begin{cases} \max_{j=1}^n \left[\frac{|y_j - \mu + 2\pi k_j|}{\sqrt{-2 \log v_j}} \right] \vee \sqrt{-2 \log \left(1 - z_2^{\frac{1}{a_\rho - 1}} \right)}, a_\rho > 1 \\ \max_{j=1}^n \left[\frac{|y_j - \mu + 2\pi k_j|}{\sqrt{-2 \log v_j}} \right], a_\rho \leq 1 \end{cases}$$

$$M_\sigma = \begin{cases} \frac{1}{z_0^{\frac{1}{n-1}}} \wedge \sqrt{\frac{-2 \log z_1}{a_\rho}}, (a_\rho \geq 1) \cup ((a_\rho < 1) \cap (z_2 \leq 1)) \\ \frac{1}{z_0^{\frac{1}{n-1}}} \wedge \sqrt{\frac{-2 \log z_1}{a_\rho}} \wedge \sqrt{-2 \log(1 - \frac{1}{z_2^{\frac{1}{1-a_\rho}}})}, ((a_\rho < 1) \cap (z_2 > 1)) \end{cases}$$

A.2 Wrapped Cauchy Distribution

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$$

$$h(\sigma) = e^{-\sigma}$$

$$h^{-1}(\rho) = -\log(\rho)$$

$$n_0 = 0$$

$$n_1 = 1$$

$$gM_1(v_j) = \sqrt{\frac{1-v_j}{v_j}}, q = 1$$

$$gm_1(v_j) = -gM_1(v_j)$$

$$h_1(\sigma) \equiv 1, \text{ so } z_3, gm_l^*(z_3) \text{ and } gM_l^*(z_3) \text{ are not needed.}$$

$$h_m \left(\frac{y_j - \mu + 2\pi k_j}{gm_{l_0}(v_j)}, \frac{y_j - \mu + 2\pi k_j}{gM_{l_0}(v_j)} \right) = \frac{|y_j - \mu + 2\pi k_j|}{gM_{l_0}(v_j)}, h_M \text{ is not needed.}$$

The full conditional densities of \mathbf{k} , \mathbf{v} , \mathbf{z} , μ and σ^2 are given by

$$[k_j | \mathbf{y}, \mathbf{k}_{-j}, \mu, \mathbf{z}, \mathbf{v}, \sigma] \propto I_{v_j} \left(0, \frac{1}{1 + \left(\frac{y_j - \mu + 2\pi k_j}{\sigma} \right)^2} \right)$$

$$k_j | \mathbf{y}, \mathbf{k}_{-j}, \mu, \mathbf{z}, \mathbf{v}, \sigma \sim DU \left[\left[\frac{1}{2\pi} \left(\mu - y_j - \sigma \sqrt{\frac{1-v_j}{v_j}} \right) \right], \left[\frac{1}{2\pi} \left(\mu - y_j + \sigma \sqrt{\frac{1-v_j}{v_j}} \right) \right] \right],$$

where DU stands for Discrete Uniform.

$$\begin{aligned}
v_j | \mathbf{y}, \mathbf{k}, \mu, \mathbf{z}, \mathbf{v}_{-j}, \sigma &\sim U \left[0, \frac{1}{1 + \left(\frac{y_j - \mu + 2\pi k_j}{\sigma} \right)^2} \right] \\
z_0 | \mathbf{y}, \mathbf{k}, \mu, \mathbf{z}_{-0}, \mathbf{v}, \sigma &\sim U \left[0, \frac{1}{\sigma^n} \right] \\
z_1 | \mathbf{y}, \mathbf{k}, \mu, \mathbf{z}_{-1}, \mathbf{v}, \sigma &\sim U [0, e^{-a_\rho \sigma}] \\
z_2 | \mathbf{y}, \mathbf{k}, \mu, \mathbf{z}_{-2}, \mathbf{v}, \sigma &\sim \begin{cases} U [0, (1 - e^{-\sigma})^{a_\rho - 1}] & a_\rho \neq 1 \\ z_2 \text{ is not needed} & a_\rho = 1 \end{cases} \\
\mu | \mathbf{y}, \mathbf{k}, \mathbf{z}, \mathbf{v}, \sigma &\sim U [m_\mu, M_\mu],
\end{aligned}$$

$$\begin{aligned}
\text{where } m_\mu &= \max_{j=1}^n \left[y_j + 2\pi k_j - \sigma \sqrt{\frac{1-v_j}{v_j}} \right] \vee 0 \\
M_\mu &= \min_{j=1}^n \left[y_j + 2\pi k_j + \sigma \sqrt{\frac{1-v_j}{v_j}} \right] \wedge (2\pi)
\end{aligned}$$

$$\sigma | \mathbf{y}, \mathbf{k}, \mu, \mathbf{z}, \mathbf{v} \sim U [m_\sigma, M_\sigma],$$

$$\begin{aligned}
\text{where } m_\sigma &= \begin{cases} \max_{j=1}^n \left[|y_j - \mu + 2\pi k_j| \sqrt{\frac{v_j}{1-v_j}} \right] \vee \left(-\log(1 - z_1^{\frac{1}{a_\rho - 1}}) \right), & a_\rho > 1 \\ \max_{j=1}^n \left[|y_j - \mu + 2\pi k_j| \sqrt{\frac{v_j}{1-v_j}} \right], & a_\rho \leq 1 \end{cases} \\
M_\sigma &= \begin{cases} \frac{1}{z_0^n} \wedge -\frac{\log z_2}{a_\rho}, & (a_\rho \geq 1) \cup ((a_\rho < 1) \cap (z_1 \leq 1)) \\ \frac{1}{z_0^n} \wedge -\frac{\log z_2}{a_\rho} \wedge -\log \left(1 - \frac{1}{z_1^{1-a_\rho}} \right), & ((a_\rho < 1) \cap (z_1 > 1)) \end{cases}
\end{aligned}$$

A.3 Wrapped Double Exponential Distribution

$$f(x) = \frac{1}{2} e^{-x}$$

$$h(\sigma) = \frac{1}{1 + \sigma^2}$$

$$h^{-1}(\rho) = \sqrt{\frac{1-\rho}{\rho}}$$

$$n_0 = 1 - 2a_\rho$$

$$n_1 = 2a_\rho$$

$$gM_1(v_j) = \log v_j, \quad q = 1$$

$$gm_1(v_j) = -gM_1(v_j)$$

$$h_1(\sigma) \equiv 1, \text{ so } z_3, gm_l^*(z_3) \text{ and } gM_l^*(z_3) \text{ are not needed.}$$

$$h_m \left(\frac{y_j - \mu + 2\pi k_j}{gm_{l_0}(v_j)}, \frac{y_j - \mu + 2\pi k_j}{gM_{l_0}(v_j)} \right) = \frac{|y_j - \mu + 2\pi k_j|}{gM_{l_0}(v_j)}, \quad h_M \text{ is not needed.}$$

Since $(1 - h(\sigma)) = \sigma^2 h(\sigma)$, z_2 is not required.

The full conditional densities of \mathbf{k} , \mathbf{v} , \mathbf{z} , μ and σ^2 are given by

$$[k_j | \mathbf{y}, \mathbf{k}_{-j}, \mu, \mathbf{z}, \mathbf{v}, \sigma] \propto I_{v_j} \left(0, e^{-\frac{|y_j - \mu + 2\pi k_j|}{\sigma}} \right)$$

$$k_j | \mathbf{y}, \mathbf{k}_{-j}, \mu, \mathbf{z}, \mathbf{v}, \sigma \sim DU \left[\left[\frac{1}{2\pi} (\mu - y_j + \sigma \log v_j) \right], \left[\frac{1}{2\pi} (\mu - y_j - \sigma \log v_j) \right] \right],$$

where DU stands for Discrete Uniform.

$$v_j | \mathbf{y}, \mathbf{k}, \mu, \mathbf{z}, \mathbf{v}_{-j}, \sigma \sim U \left[0, e^{-\frac{|y_j - \mu + 2\pi k_j|}{\sigma}} \right]$$

$$z_0 | \mathbf{y}, \mathbf{k}, \mu, \mathbf{z}_{-0}, \mathbf{v}, \sigma \sim U \left[0, \frac{1}{\sigma^{n-2a_\rho+1}} \right]$$

$$z_1 | \mathbf{y}, \mathbf{k}, \mu, \mathbf{z}_{-1}, \mathbf{v}, \sigma \sim U \left[0, \frac{1}{(1+\sigma^2)^{2a_\rho}} \right]$$

$$\mu | \mathbf{y}, \mathbf{k}, \mathbf{z}, \mathbf{v}, \sigma \sim U [m_\mu, M_\mu],$$

$$\text{where } m_\mu = \max_{j=1}^n [y_j + 2\pi k_j + \sigma \log v_j] \bigvee 0$$

$$M_\mu = \min_{j=1}^n [y_j + 2\pi k_j - \sigma \log v_j] \bigwedge (2\pi)$$

$$\sigma | \mathbf{y}, \mathbf{k}, \mu, \mathbf{z}, \mathbf{v} \sim U [m_\sigma, M_\sigma],$$

$$\text{where } m_\sigma = \max_{j=1}^n \left[\frac{|y_j - \mu + 2\pi k_j|}{-\log v_j} \right]$$

$$M_\sigma = \frac{1}{z_0^{n-2a_\rho+1}} \bigwedge_{z_1} \sqrt{\frac{\frac{1}{2a_\rho}}{\frac{1-z_1}{2a_\rho}}}$$

A.4 Wrapped Extreme Value Distribution

$$f(x) = \frac{1}{\sigma} e^{-\left(\frac{x-\mu}{\sigma}\right)} e^{-e^{-\left(\frac{x-\mu}{\sigma}\right)}}, \quad -\infty < x < \infty$$

$$\rho = \sqrt{\frac{\pi\sigma}{\sinh(\pi\sigma)}}$$

The prior for (μ, σ) is given by

$$[\mu, \sigma] \propto I_\mu(0, 2\pi) \sigma^{-2(a+1)} e^{-b/\sigma^2}, \quad a > 0, b > 0$$

The joint density of $\mathbf{y}, \mathbf{k}, \mu$ and σ is given by

$$[\mathbf{y}, \mathbf{k}, \mu, \sigma] \propto [\mathbf{y}|\mathbf{k}, \mu, \sigma^2] [\mathbf{k}|\mu, \sigma^2] [\mu, \sigma^2]$$

$$\propto \prod_{j=1}^n \left(\frac{1}{\sigma} e^{-\left(\frac{y_j-\mu+2\pi k_j}{\sigma}\right)} e^{-e^{-\left(\frac{y_j-\mu+2\pi k_j}{\sigma}\right)}} \right) I_\mu(0, 2\pi) \sigma^{-2(a+1)} e^{-b/\sigma^2}$$

The joint density of $\mathbf{y}, \mathbf{k}, \mu, \mathbf{z}, \mathbf{v}, \mathbf{w}$ and σ is given by,

$$[\mathbf{y}, \mathbf{k}, \mu, \mathbf{z}, \mathbf{v}, \mathbf{w}, \sigma] \propto I_{z_0} \left(0, \frac{1}{\sigma^{n+2a+2}}\right) \prod_{j=1}^n I_{v_j} \left(0, e^{-\left(\frac{y_j-\mu+2\pi k_j}{\sigma}\right)}\right) I_{w_j} \left(0, e^{-e^{-\left(\frac{y_j-\mu+2\pi k_j}{\sigma}\right)}}\right)$$

$$I_{z_1} \left(0, e^{-b/\sigma^2}\right) I_\mu(0, 2\pi)$$

The full conditional densities of $\mathbf{k}, \mathbf{v}, \mathbf{z}, \mu$ and σ^2 are given by

$$k_j | \mathbf{y}, \mathbf{k}_{-j}, \mu, \mathbf{z}, \mathbf{v}, \mathbf{w}, \sigma \sim DU \left[\left[\frac{1}{2\pi} (\mu - y_j - \sigma \log(-\log w_j)) \right], \right.$$

$$\left. \left[\frac{1}{2\pi} (\mu - y_j - \sigma \log v_j) \right] \right]$$

where DU stands for Discrete Uniform.

$$\begin{aligned}
v_j | \mathbf{y}, \mathbf{k}, \mu, z, \mathbf{v}_{-j}, \mathbf{w}, \sigma &\sim U \left[0, e^{-\left(\frac{y_j - \mu + 2\pi k_j}{\sigma}\right)} \right] \\
w_j | \mathbf{y}, \mathbf{k}, \mu, z, \mathbf{v}, \mathbf{w}_{-j}, \sigma &\sim U \left[0, e^{-e^{-\left(\frac{y_j - \mu + 2\pi k_j}{\sigma}\right)}} \right] \\
z_0 | \mathbf{y}, \mathbf{k}, \mu, z_{-0}, \mathbf{v}, \mathbf{w}, \sigma &\sim U \left[0, \frac{1}{\sigma^{n+2a+2}} \right] \\
z_1 | \mathbf{y}, \mathbf{k}, \mu, z_{-1}, \mathbf{v}, \mathbf{w}, \sigma &\sim U \left[0, e^{-b/\sigma^2} \right] \\
\mu | \mathbf{y}, \mathbf{k}, z, \mathbf{v}, \mathbf{w}, \sigma &\sim U [m_\mu, M_\mu],
\end{aligned}$$

$$\text{where } m_\mu = \max_{j=1}^n [y_j + 2\pi k_j + \sigma \log v_j] \sqrt{0}$$

$$M_\mu = \min_{j=1}^n [y_j + 2\pi k_j + \sigma \log(-\log w_j)] \wedge (2\pi)$$

$$\sigma | \mathbf{y}, \mathbf{k}, \mu, z, \mathbf{v}, \mathbf{w} \sim U [m_\sigma, M_\sigma],$$

$$\text{where } m_\sigma = \sqrt{\frac{-b}{\log z_1}} \vee \max_{\{j:w_j < e^{-1}\}} \left[\frac{(y_j - \mu + 2\pi k_j)}{-\log(-\log w_j)} \right] \vee \max_{\{j:v_j < 1\}} \left[\frac{(y_j - \mu + 2\pi k_j)}{-\log v_j} \right]$$

$$M_\sigma = \frac{1}{z_0^{n+2a+2}} \wedge \min_{\{j:w_j > e^{-1}\}} \left[\frac{(y_j - \mu + 2\pi k_j)}{-\log(-\log w_j)} \right] \vee \min_{\{j:v_j > 1\}} \left[\frac{(y_j - \mu + 2\pi k_j)}{-\log v_j} \right]$$

A.5 Bimodal (Wrapped Beta) Distribution

$$f(x) = \frac{3}{4\sigma} \left[1 - \left(\frac{x-\mu}{\sigma} \right)^2 \right], \quad \mu - \sigma < x < \mu + \sigma$$

$$\rho = \left| \frac{-3}{\sigma^3} (-\sin(\sigma) + \cos(\sigma)\sigma) \right|$$

The prior for (μ, σ) is given by

$$[\mu, \sigma] \propto I_\mu(0, 2\pi) \sigma^{-2(a+1)} e^{-b/\sigma^2}, \quad a > 0, \quad b > 0$$

The joint density of $\mathbf{y}, \mathbf{k}, \mu$ and σ is given by

$$\begin{aligned} [\mathbf{y}, \mathbf{k}, \mu, \sigma] &\propto [\mathbf{y}|\mathbf{k}, \mu, \sigma^2] [\mathbf{k}|\mu, \sigma^2] [\mu, \sigma^2] \\ &\propto \prod_{j=1}^n \left(\frac{3}{4\sigma} \left[1 - \left(\frac{y_j - \mu + 2\pi k_j}{\sigma} \right)^2 \right] \right) I_\mu(0, 2\pi) \sigma^{-2(a+1)} e^{-b/\sigma^2} \end{aligned}$$

The joint density of $\mathbf{y}, \mathbf{k}, \mu, \mathbf{z}, \mathbf{v}$ and σ is given by,

$$[\mathbf{y}, \mathbf{k}, \mu, \mathbf{z}, \mathbf{v}, \sigma] \propto I_{z_0} \left(0, \frac{1}{\sigma^{n+2a+2}} \right) \prod_{j=1}^n I_{v_j} \left(0, 1 - \left(\frac{y_j - \mu + 2\pi k_j}{\sigma} \right)^2 \right) I_{z_1} \left(0, e^{-b/\sigma^2} \right) I_\mu(0, 2\pi)$$

The full conditional densities of $\mathbf{k}, \mathbf{v}, \mathbf{z}, \mu$ and σ^2 are given by,

$$\begin{aligned} k_j | \mathbf{y}, \mathbf{k}_{-j}, \mu, \mathbf{z}, \mathbf{v}, \sigma &\sim DU \left[\left[\frac{1}{2\pi} (\mu - y_j - \sigma \sqrt{1 - v_j}) \right], \right. \\ &\quad \left. \left[\frac{1}{2\pi} (\mu - y_j + \sigma \sqrt{1 - v_j}) \right] \right], \end{aligned}$$

where DU stands for Discrete Uniform.

$$v_j | \mathbf{y}, \mathbf{k}, \mu, \mathbf{z}, \mathbf{v}_{-j}, \sigma \sim U \left[0, 1 - \left(\frac{y_j - \mu + 2\pi k_j}{\sigma} \right)^2 \right]$$

$$z_0 | \mathbf{y}, \mathbf{k}, \mu, \mathbf{z}_{-0}, \mathbf{v}, \sigma \sim U \left[0, \frac{1}{\sigma^{n+2a+2}} \right]$$

$$z_1 | \mathbf{y}, \mathbf{k}, \mu, \mathbf{z}_{-1}, \mathbf{v}, \sigma \sim U \left[0, e^{-b/\sigma^2} \right]$$

$$\mu | \mathbf{y}, \mathbf{k}, \mathbf{z}, \mathbf{v}, \sigma \sim U [m_\mu, M_\mu],$$

$$\text{where } m_\mu = \max_{j=1}^n [y_j + 2\pi k_j - \sigma \sqrt{1 - v_j}] \bigvee 0$$

$$M_\mu = \min_{j=1}^n [y_j + 2\pi k_j + \sigma \sqrt{1 - v_j}] \bigwedge (2\pi)$$

$$\sigma | \mathbf{y}, \mathbf{k}, \mu, \mathbf{z}, \mathbf{v} \sim U [m_\sigma, M_\sigma],$$

$$\text{where } m_\sigma = \sqrt{\frac{-b}{\log z_1}} \bigvee \max_{j=1}^n \left[\frac{|y_j - \mu + 2\pi k_j|}{\sqrt{1 - v_j}} \right]$$

$$M_\sigma = \frac{1}{z_0^{n+2a+2}}$$

Appendix B

Explicit full conditionals for regression

B.1 Wrapped Normal Distribution

$$f(x) = \sqrt{\frac{1}{2\pi}} e^{-\frac{x^2}{2}}$$

$$h(\sigma) = e^{-\frac{\sigma^2}{2}}$$

$$h^{-1}(\rho) = \sqrt{-2\sigma^2 \log \rho}$$

$$n_0 = -1$$

$$n_1 = 1$$

$$gM_1(v_j) = \sqrt{-2 \log v_j}, \quad q = 1$$

$$gm_1(v_j) = -gM_1(v_j)$$

$h_1(\sigma) \equiv 1$, so z_3 , $gm_l^*(z_3)$ and $gM_l^*(z_3)$ are not needed.

$$h_m \left(\frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gm_{l_0}(v_j)}, \frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gM_{l_0}(v_j)} \right) = \frac{|y_j - \beta_0 - \beta_1 x_j + 2\pi k_j|}{gM_{l_0}(v_j)}, h_M \text{ is not needed.}$$

The full conditional densities of \mathbf{k} , \mathbf{v} , \mathbf{z} , β_0 , β_1 and σ^2 are given by

$$\begin{aligned} [k_j | \mathbf{y}, \mathbf{k}_{-j}, \beta_0, \beta_1, \mathbf{z}, \mathbf{v}, \sigma] &\propto I_{v_j} \left(0, e^{-\frac{(y_j - \beta_0 - \beta_1 x_j + 2\pi k_j)^2}{2\sigma^2}} \right) \\ k_j | \mathbf{y}, \mathbf{k}_{-j}, \beta_0, \beta_1, \mathbf{z}, \mathbf{v}, \sigma &\sim DU \left[\left[\frac{1}{2\pi} \left(\beta_0 + \beta_1 x_j - y_j - \sqrt{-2\sigma^2 \log v_j} \right) \right], \right. \\ &\quad \left. \left[\frac{1}{2\pi} \left(\beta_0 + \beta_1 x_j - y_j + \sqrt{-2\sigma^2 \log v_j} \right) \right] \right], \end{aligned}$$

where DU stands for Discrete Uniform.

$$\begin{aligned} v_j | \mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}, \mathbf{v}_{-j}, \sigma &\sim U \left[0, e^{-\frac{(y_j - \beta_0 - \beta_1 x_j + 2\pi k_j)^2}{2\sigma^2}} \right] \\ z_0 | \mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}_{-0}, \mathbf{v}, \sigma &\sim U \left[0, \frac{1}{\sigma^{n-1}} \right] \\ z_1 | \mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}_{-1}, \mathbf{v}, \sigma &\sim U \left[0, e^{-\frac{a_\rho \sigma^2}{2}} \right] \\ z_2 | \mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}_{-2}, \mathbf{v}, \sigma &\sim \begin{cases} U \left[0, \left(1 - e^{-\frac{\sigma^2}{2}} \right)^{a_\rho - 1} \right] & a_\rho \neq 1 \\ z_2 \text{ is not needed} & a_\rho = 1 \end{cases} \\ \beta_0 | \mathbf{y}, \mathbf{k}, \beta_1, \mathbf{z}, \mathbf{v}, \sigma &\sim U [m_{\beta_0}, M_{\beta_0}], \end{aligned}$$

$$\text{where } m_{\beta_0} = \max_{j=1}^n \left[y_j + 2\pi k_j - \beta_1 x_j - \sqrt{-2\sigma^2 \log v_j} \right] \vee 0$$

$$M_{\beta_0} = \min_{j=1}^n \left[y_j + 2\pi k_j - \beta_1 x_j + \sqrt{-2\sigma^2 \log v_j} \right] \wedge (2\pi)$$

$$\beta_1 | \mathbf{y}, \mathbf{k}, \mathbf{z}, \beta_0, \mathbf{v}, \sigma \sim U [m_{\beta_1}, M_{\beta_1}],$$

$$\text{where } m_{\beta_1} = \max_{j=1}^n m_{\beta_1, j} \vee 0,$$

$$M_{\beta_1} = \min_{j=1}^n M_{\beta_1, j} \wedge (2\pi),$$

$$\text{where } m_{\beta_1, j} = \begin{cases} \left[\frac{y_j + 2\pi k_j - \beta_0 - \sqrt{-2\sigma^2 \log v_j}}{x_j} \right], & x_j > 0 \\ \left[\frac{y_j + 2\pi k_j - \beta_0 + \sqrt{-2\sigma^2 \log v_j}}{x_j} \right], & x_j < 0 \end{cases}$$

$$M_{\beta_1, j} = \begin{cases} \left[\frac{y_j + 2\pi k_j - \beta_0 + \sqrt{-2\sigma^2 \log v_j}}{x_j} \right], & x_j > 0 \\ \left[\frac{y_j + 2\pi k_j - \beta_0 - \sqrt{-2\sigma^2 \log v_j}}{x_j} \right], & x_j < 0 \end{cases}$$

$$\sigma | \mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}, \mathbf{v} \sim U [m_\sigma, M_\sigma],$$

$$\text{where } m_\sigma = \begin{cases} \max_{j=1}^n \left[\frac{|y_j - \beta_0 - \beta_1 x_j + 2\pi k_j|}{\sqrt{-2 \log v_j}} \right] \vee \sqrt{-2 \log \left(1 - \frac{1}{z_2^{a_\rho - 1}}\right)}, & a_\rho > 1 \\ \max_{j=1}^n \left[\frac{|y_j - \beta_0 - \beta_1 x_j + 2\pi k_j|}{\sqrt{-2 \log v_j}} \right], & a_\rho \leq 1 \end{cases}$$

$$M_\sigma = \begin{cases} \frac{1}{z_0^{n-1}} \wedge \sqrt{\frac{-2 \log z_1}{a_\rho}}, & (a_\rho \geq 1) \cup ((a_\rho < 1) \cap (z_2 \leq 1)) \\ \frac{1}{z_0^{n-1}} \wedge \sqrt{\frac{-2 \log z_1}{a_\rho}} \wedge \sqrt{-2 \log \left(1 - \frac{1}{z_2^{1-a_\rho}}\right)}, & ((a_\rho < 1) \cap (z_2 > 1)) \end{cases}$$

B.2 Wrapped Cauchy Distribution

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$$

$$h(\sigma) = e^{-\sigma}$$

$$h^{-1}(\rho) = -\log(\rho)$$

$$n_0 = 0$$

$$n_1 = 1$$

$$gM_1(v_j) = \sqrt{\frac{1-v_j}{v_j}}, \quad q = 1$$

$$gm_1(v_j) = -gM_1(v_j)$$

$$h_1(\sigma) \equiv 1, \text{ so } z_3, gm_l^*(z_3) \text{ and } gM_l^*(z_3) \text{ are not needed.}$$

$$h_m \left(\frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{g m_{I_0}(v_j)}, \frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{g M_{I_0}(v_j)} \right) = \frac{|y_j - \beta_0 - \beta_1 x_j + 2\pi k_j|}{g M_{I_0}(v_j)}, h_M \text{ is not needed.}$$

The full conditional densities of \mathbf{k} , \mathbf{v} , \mathbf{z} , β_0 , β_1 and σ^2 are given by

$$\begin{aligned} [k_j | \mathbf{y}, \mathbf{k}_{-j}, \beta_0, \beta_1, \mathbf{z}, \mathbf{v}, \sigma] &\propto I_{v_j} \left(0, \frac{1}{1 + \left(\frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{\sigma} \right)^2} \right) \\ k_j | \mathbf{y}, \mathbf{k}_{-j}, \beta_0, \beta_1, \mathbf{z}, \mathbf{v}, \sigma &\sim DU \left[\left[\frac{1}{2\pi} \left(\beta_0 + \beta_1 x_j - y_j - \sigma \sqrt{\frac{1-v_j}{v_j}} \right) \right], \right. \\ &\quad \left. \left[\frac{1}{2\pi} \left(\beta_0 + \beta_1 x_j - y_j + \sigma \sqrt{\frac{1-v_j}{v_j}} \right) \right] \right], \end{aligned}$$

where DU stands for Discrete Uniform.

$$\begin{aligned} v_j | \mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}, \mathbf{v}_{-j}, \sigma &\sim U \left[0, \frac{1}{1 + \left(\frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{\sigma} \right)^2} \right] \\ z_0 | \mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}_{-0}, \mathbf{v}, \sigma &\sim U \left[0, \frac{1}{\sigma^n} \right] \\ z_1 | \mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}_{-1}, \mathbf{v}, \sigma &\sim U \left[0, e^{-a_\rho \sigma} \right] \\ z_2 | \mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}_{-2}, \mathbf{v}, \sigma &\sim \begin{cases} U \left[0, (1 - e^{-\sigma})^{a_\rho - 1} \right] & a_\rho \neq 1 \\ z_2 \text{ is not needed} & a_\rho = 1 \end{cases} \\ \beta_0 | \mathbf{y}, \mathbf{k}, \beta_1, \mathbf{z}, \mathbf{v}, \sigma &\sim U [m_{\beta_0}, M_{\beta_0}], \end{aligned}$$

$$\begin{aligned} \text{where } m_{\beta_0} &= \max_{j=1}^n \left[y_j + 2\pi k_j - \beta_1 x_j - \sigma \sqrt{\frac{1-v_j}{v_j}} \right] \vee 0 \\ M_{\beta_0} &= \min_{j=1}^n \left[y_j + 2\pi k_j - \beta_1 x_j + \sigma \sqrt{\frac{1-v_j}{v_j}} \right] \wedge (2\pi) \end{aligned}$$

$$\beta_1 | \mathbf{y}, \mathbf{k}, \mathbf{z}, \beta_0, \mathbf{v}, \sigma \sim U [m_{\beta_1}, M_{\beta_1}],$$

$$\begin{aligned} \text{where } m_{\beta_1} &= \max_{j=1}^n m_{\beta_1, j} \vee 0, \\ M_{\beta_1} &= \min_{j=1}^n M_{\beta_1, j} \wedge (2\pi), \end{aligned}$$

$$\text{where } m_{\beta_1, j} = \begin{cases} \left[\frac{y_j + 2\pi k_j - \beta_0 - \sigma \sqrt{\frac{1-v_j}{v_j}}}{x_j} \right], & x_j > 0 \\ \left[\frac{y_j + 2\pi k_j - \beta_0 + \sigma \sqrt{\frac{1-v_j}{v_j}}}{x_j} \right], & x_j < 0 \end{cases}$$

$$M_{\beta_1, j} = \begin{cases} \left[\frac{y_j + 2\pi k_j - \beta_0 + \sigma \sqrt{\frac{1-v_j}{v_j}}}{x_j} \right], & x_j > 0 \\ \left[\frac{y_j + 2\pi k_j - \beta_0 - \sigma \sqrt{\frac{1-v_j}{v_j}}}{x_j} \right], & x_j < 0 \end{cases}$$

$$\sigma | \mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}, \mathbf{v} \sim U [m_\sigma, M_\sigma],$$

$$\text{where } m_\sigma = \begin{cases} \left[\max_{j=1}^n \left[|y_j - \beta_0 - \beta_1 x_j + 2\pi k_j| \sqrt{\frac{v_j}{1-v_j}} \right] \right. \\ \left. \vee \left(-\log \left(1 - z_1^{\frac{1}{a_\rho - 1}} \right) \right) \right], & a_\rho > 1 \\ \max_{j=1}^n \left[|y_j - \beta_0 - \beta_1 x_j + 2\pi k_j| \sqrt{\frac{v_j}{1-v_j}} \right], & a_\rho \leq 1 \end{cases}$$

$$M_\sigma = \begin{cases} \frac{1}{z_0^n} \wedge -\frac{\log z_2}{a_\rho}, & (a_\rho \geq 1) \cup ((a_\rho < 1) \cap (z_1 \leq 1)) \\ \frac{1}{z_0^n} \wedge -\frac{\log z_2}{a_\rho} \wedge -\log \left(1 - \frac{1}{z_1^{\frac{1}{1-a_\rho}}} \right), & ((a_\rho < 1) \cap (z_1 > 1)) \end{cases}$$

B.3 Wrapped Double Exponential Distribution

$$f(x) = \frac{1}{2} e^{-x}$$

$$h(\sigma) = \frac{1}{1+\sigma^2}$$

$$h^{-1}(\rho) = \sqrt{\frac{1-\rho}{\rho}}$$

$$n_0 = 1 - 2a_\rho$$

$$n_1 = 2a_\rho$$

$$gM_1(v_j) = \log v_j, \quad q = 1$$

$$gm_1(v_j) = -gM_1(v_j)$$

$$h_1(\sigma) \equiv 1, \text{ so } z_3, gm_l^*(z_3) \text{ and } gM_l^*(z_3) \text{ are not needed.}$$

$$h_m \left(\frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gm_{i_0}(v_j)}, \frac{y_j - \beta_0 - \beta_1 x_j + 2\pi k_j}{gM_{i_0}(v_j)} \right) = \frac{|y_j - \beta_0 - \beta_1 x_j + 2\pi k_j|}{gM_{i_0}(v_j)}, \quad h_M \text{ is not needed.}$$

Since $(1 - h(\sigma)) = \sigma^2 h(\sigma)$, z_2 is not required.

The full conditional densities of \mathbf{k} , \mathbf{v} , \mathbf{z} , β_0 , β_1 and σ^2 are given by

$$\begin{aligned} [k_j | \mathbf{y}, \mathbf{k}_{-j}, \beta_0, \beta_1, \mathbf{z}, \mathbf{v}, \sigma] &\propto I_{v_j} \left(0, e^{-\frac{|y_j - \beta_0 - \beta_1 x_j + 2\pi k_j|}{\sigma}} \right) \\ k_j | \mathbf{y}, \mathbf{k}_{-j}, \beta_0, \beta_1, \mathbf{z}, \mathbf{v}, \sigma &\sim DU \left[\left[\frac{1}{2\pi} (\beta_0 + \beta_1 x_j - y_j + \sigma \log v_j) \right], \right. \\ &\quad \left. \left[\frac{1}{2\pi} (\beta_0 + \beta_1 x_j - y_j - \sigma \log v_j) \right] \right], \end{aligned}$$

where DU stands for Discrete Uniform.

$$v_j | \mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}, \mathbf{v}_{-j}, \sigma \sim U \left[0, e^{-\frac{|y_j - \beta_0 - \beta_1 x_j + 2\pi k_j|}{\sigma}} \right]$$

$$z_0 | \mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}_{-0}, \mathbf{v}, \sigma \sim U \left[0, \frac{1}{\sigma^{n-2a_\rho+1}} \right]$$

$$z_1 | \mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}_{-1}, \mathbf{v}, \sigma \sim U \left[0, \frac{1}{(1+\sigma^2)^{2a_\rho}} \right]$$

$$\beta_0 | \mathbf{y}, \mathbf{k}, \beta_1, \mathbf{z}, \mathbf{v}, \sigma \sim U [m_{\beta_0}, M_{\beta_0}],$$

$$\text{where } m_{\beta_0} = \max_{j=1}^n [y_j + 2\pi k_j - \beta_1 x_j + \sigma \log v_j] \bigvee 0$$

$$M_{\beta_0} = \min_{j=1}^n [y_j + 2\pi k_j - \beta_1 x_j - \sigma \log v_j] \bigwedge (2\pi)$$

$$\beta_1 | \mathbf{y}, \mathbf{k}, \mathbf{z}, \beta_0, \mathbf{v}, \sigma \sim U [m_{\beta_1}, M_{\beta_1}],$$

$$\text{where } m_{\beta_1} = \max_{j=1}^n m_{\beta_1,j} \bigvee 0,$$

$$M_{\beta_1} = \min_{j=1}^n M_{\beta_1,j} \bigwedge (2\pi),$$

$$\begin{aligned}
\text{where } m_{\beta_1, j} &= \begin{cases} \left[\frac{y_j + 2\pi k_j - \beta_0 + \sigma \log v_j}{x_j} \right], & x_j > 0 \\ \left[\frac{y_j + 2\pi k_j - \beta_0 - \sigma \log v_j}{x_j} \right], & x_j < 0 \end{cases} \\
M_{\beta_1, j} &= \begin{cases} \left[\frac{y_j + 2\pi k_j - \beta_0 - \sigma \log v_j}{x_j} \right], & x_j > 0 \\ \left[\frac{y_j + 2\pi k_j - \beta_0 + \sigma \log v_j}{x_j} \right], & x_j < 0 \end{cases} \\
\sigma | \mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}, \mathbf{v} &\sim U [m_\sigma, M_\sigma],
\end{aligned}$$

$$\text{where } m_\sigma = \max_{j=1}^n \left[\frac{|y_j - \beta_0 - \beta_1 x_j + 2\pi k_j|}{-\log v_j} \right]$$

$$M_\sigma = \frac{1}{\frac{1}{n-2a_\rho+1} z_0} \bigwedge \sqrt{\frac{\frac{1}{2a_\rho}}{1-z_1} \frac{1}{z_1 \frac{1}{2a_\rho}}}$$

Appendix C

Explicit full conditionals for Time series

C.1 Wrapped Normal Distribution

$$f(x) = \sqrt{\frac{1}{2\pi}} e^{-\frac{x^2}{2}}$$

$$h(\sigma) = e^{-\frac{\sigma^2}{2}}$$

$$n_0 = -1$$

$$n_1 = 1$$

$$x_1 = 0$$

$$x_j = y_{j-1} + 2\pi k_{j-1} - \beta_0, j = 2 \dots n$$

$h_1(\sigma) \equiv 1$, so z_3 is not needed. The full conditional densities of \mathbf{k} , \mathbf{v} , \mathbf{z} , β_0 , β_1 and σ^2 are

given by

$$k_j | \mathbf{y}, \mathbf{k}_{-j}, \beta_0, \beta_1, \mathbf{z}, \mathbf{v}, \sigma \sim DU [m_{k_j}, M_{k_j}], j = 1, \dots, n,$$

where DU stands for Discrete Uniform.

$$\begin{aligned} m_{k_j} &= m1_{k_j} \vee m2_{k_j} \\ m1_{k_j} &= \begin{cases} \left\lfloor \frac{1}{2\pi} \left(\beta_0 + \beta_1 x_j - y_j - \sqrt{-2\sigma^2 \log v_j / (1 - \beta_1^2)} \right) \right\rfloor & j = 1 \\ \left\lfloor \frac{1}{2\pi} \left(\beta_0 + \beta_1 x_j - y_j - \sqrt{-2\sigma^2 \log v_j} \right) \right\rfloor & j > 1 \end{cases} \\ m2_{k_j} &= \begin{cases} \left\lfloor \frac{1}{2\pi\beta_1} \left(-\beta_0 - \beta_1(y_j - \beta_0) + y_{j+1} + 2\pi k_{j+1} - \sqrt{-2\sigma^2 \log v_{j+1}} \right) \right\rfloor & , \beta_1 > 0 \\ \left\lfloor \frac{1}{2\pi\beta_1} \left(-\beta_0 - \beta_1(y_j - \beta_0) + y_{j+1} + 2\pi k_{j+1} + \sqrt{-2\sigma^2 \log v_{j+1}} \right) \right\rfloor & , \beta_1 < 0 \\ m2_{k_j} \text{ is not required} & , (\beta_1 = 0) \cup (j = n) \end{cases} \\ M_{k_j} &= M1_{k_j} \wedge M2_{k_j} \\ M1_{k_j} &= \begin{cases} \left\lfloor \frac{1}{2\pi} \left(\beta_0 + \beta_1 x_j - y_j + \sqrt{-2\sigma^2 \log v_j / (1 - \beta_1^2)} \right) \right\rfloor & j = 1 \\ \left\lfloor \frac{1}{2\pi} \left(\beta_0 + \beta_1 x_j - y_j + \sqrt{-2\sigma^2 \log v_j} \right) \right\rfloor & j > 1 \end{cases} \\ M2_{k_j} &= \begin{cases} \left\lfloor \frac{1}{2\pi\beta_1} \left(-\beta_0 - \beta_1(y_j - \beta_0) + y_{j+1} + 2\pi k_{j+1} + \sqrt{-2\sigma^2 \log v_{j+1}} \right) \right\rfloor & , \beta_1 > 0 \\ \left\lfloor \frac{1}{2\pi\beta_1} \left(-\beta_0 - \beta_1(y_j - \beta_0) + y_{j+1} + 2\pi k_{j+1} - \sqrt{-2\sigma^2 \log v_{j+1}} \right) \right\rfloor & , \beta_1 < 0 \\ M2_{k_j} \text{ is not required} & , (\beta_1 = 0) \cup (j = n) \end{cases} \end{aligned}$$

$$v_1 | \mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}, \mathbf{v}_{-1}, \sigma \sim U \left[0, e^{-\frac{(y_1 - \beta_0 + 2\pi k_1)^2 (1 - \beta_1^2)}{2\sigma^2}} \right]$$

$$v_j | \mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}, \mathbf{v}_{-j}, \sigma \sim U \left[0, e^{-\frac{(y_j - \beta_0 - \beta_1 x_j + 2\pi k_j)^2}{2\sigma^2}} \right]$$

$$z_0 | \mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}_{-0}, \mathbf{v}, \sigma \sim U \left[0, \frac{1}{\sigma^{n-1}} \right]$$

$$z_1 | \mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}_{-1}, \mathbf{v}, \sigma \sim U \left[0, e^{-\frac{a_\rho \sigma^2}{2}} \right]$$

$$z_2|\mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}_{-2}, \mathbf{v}, \sigma \sim \begin{cases} U \left[0, \left(1 - e^{-\frac{\sigma^2}{2}} \right)^{a_\rho - 1} \right] & a_\rho \neq 1 \\ z_2 \text{ is not needed} & a_\rho = 1 \end{cases}$$

$$z_4|\mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}_{-4}, \mathbf{v}, \sigma \sim U \left[0, \sqrt{1 - \beta_1^2} \right]$$

$$\beta_0|\mathbf{y}, \mathbf{k}, \beta_1, \mathbf{z}, \mathbf{v}, \sigma \sim U [m_{\beta_0}, M_{\beta_0}],$$

$$\begin{aligned} \text{where } m_{\beta_0} &= \left[y_1 + 2\pi k_1 - \sqrt{-2\sigma^2 \log v_1 / (1 - \beta_1^2)} \right] \bigvee \\ & \max_{j=2}^n \left[y_j + 2\pi k_j - \beta_1(y_{j-1} + 2\pi k_{j-1}) - \sqrt{-2\sigma^2 \log v_j} \right] / (1 - \beta_1) \bigvee 0 \\ M_{\beta_0} &= \left[y_1 + 2\pi k_1 - \sqrt{-2\sigma^2 \log v_1 / (1 - \beta_1^2)} \right] \bigwedge \\ & \min_{j=2}^n \left[y_j + 2\pi k_j - \beta_1(y_{j-1} + 2\pi k_{j-1}) + \sqrt{-2\sigma^2 \log v_j} \right] / (1 - \beta_1) \bigwedge (2\pi) \end{aligned}$$

$$\beta_1|\mathbf{y}, \mathbf{k}, \mathbf{z}, \beta_0, \mathbf{v}, \sigma \sim \begin{cases} U [m_{\beta_1}, M_{\beta_1}] & , a_{\beta_1} \leq 0 \\ U \{ [m_{\beta_1}, M_{\beta_1}] \cap ([-1, -\sqrt{a_{\beta_1}}] \cup [\sqrt{a_{\beta_1}}, 1]) \} & , a_{\beta_1} > 0, \end{cases}$$

$$\text{where } m_{\beta_1} = \max_{j=2}^n m_{\beta_1, j} \bigvee -\sqrt{1 - z_4^2},$$

$$M_{\beta_1} = \min_{j=2}^n M_{\beta_1, j} \bigwedge \sqrt{1 - z_4^2},$$

$$\text{where } m_{\beta_1, j} = \begin{cases} \left[\frac{y_j + 2\pi k_j - \beta_0 - \sqrt{-2\sigma^2 \log v_j}}{x_j} \right], & x_j > 0 \\ \left[\frac{y_j + 2\pi k_j - \beta_0 + \sqrt{-2\sigma^2 \log v_j}}{x_j} \right], & x_j < 0 \end{cases}$$

$$M_{\beta_1, j} = \begin{cases} \left[\frac{y_j + 2\pi k_j - \beta_0 + \sqrt{-2\sigma^2 \log v_j}}{x_j} \right], & x_j > 0 \\ \left[\frac{y_j + 2\pi k_j - \beta_0 - \sqrt{-2\sigma^2 \log v_j}}{x_j} \right], & x_j < 0 \end{cases}$$

$$a_{\beta_1} = 1 + (2\sigma^2 \log v_1) / (y_1 - \beta_0 + 2\pi k_1)^2$$

$$\sigma|\mathbf{y}, \mathbf{k}, \beta_0, \beta_1, \mathbf{z}, \mathbf{v} \sim U [m_\sigma, M_\sigma],$$

$$\text{where } m_\sigma = \begin{cases} \max_{j=1}^n \left[\frac{|y_j - \beta_0 - \beta_1 x_j + 2\pi k_j| \sqrt{1 - \beta_1^2} I(j=1)}{\sqrt{-2 \log v_j}} \right] \bigvee \sqrt{-2 \log \left(1 - z_2^{\frac{1}{a_\rho - 1}} \right)}, & a_\rho > 1 \\ \max_{j=1}^n \left[\frac{|y_j - \beta_0 - \beta_1 x_j + 2\pi k_j| \sqrt{1 - \beta_1^2} I(j=1)}{\sqrt{-2 \log v_j}} \right], & a_\rho \leq 1 \end{cases}$$

$$M_\sigma = \begin{cases} \frac{1}{z_0^{\frac{1}{n-1}}} \wedge \sqrt{\frac{-2 \log z_1}{a_\rho}}, (a_\rho \geq 1) \cup ((a_\rho < 1) \cap (z_2 \leq 1)) \\ \frac{1}{z_0^{\frac{1}{n-1}}} \wedge \sqrt{\frac{-2 \log z_1}{a_\rho}} \wedge \sqrt{-2 \log(1 - \frac{1}{z_2^{1-a_\rho}})}, ((a_\rho < 1) \cap (z_2 > 1)) \end{cases}$$

Appendix D

Regression with categorical covariate

Regression with categorical covariate is a special case of the model given by (3.1). In Section 3.3, we considered the model $\mu_j = \beta_0 + \beta_1 x_j$ in detail. In this appendix, we extend this model with a categorical covariate, w_j . w_j is 0 or 1. The model is $\mu_j = \beta_0 + \beta_1 x_j + \beta_2 w_j$. Without loss of generality, we can restrict β_2 between β_0 and $\beta_0 + 2\pi$, to avoid identifiability problems. Therefore, we assume a Uniform prior for β_2 on $(\beta_0, \beta_0 + 2\pi)$. The general theory is the same as that given in Section 3.3.

In order to study the sensitivity of the sampling distribution, we generated samples of size $n = 31$ from WN, WC and WDE with parameters $\beta_0 = 1$, $\beta_1 = 1.5$, $\beta_2 = 2$ and $\rho = 0.8$. We use the same data generating process as used in Section 3.4. w is sampled from Bernoulli(0.5). We then fitted Wrapped Normal, Wrapped Cauchy and Wrapped Double Exponential distributions to the three datasets. We repeated the method 500 times to see

the frequentist performance of the Bayes method for erroneous models. As before, we report the posterior mean, standard deviation, and two equal tail percentiles along with the Monte Carlo standard error. The percentiles are computed with 0 radians as the reference point. The results are tabulated in Table D.1, Table D.2 and Table D.3.

Comparing the results in Table D.1, Table D.2 and Table D.3, we see that the model selection criteria GGC and Dev work well and select the correct distribution. Also, in this case β_0 , β_1 , β_2 and ρ have been well estimated. In general, the location parameters β_0 , β_1 and β_2 are estimated well, even when the models are incorrect. The results are very similar to the results in Section 3.4.

In Table D.1, Table D.2 and Table D.3, we see WN model estimates the location parameters well, even when the distribution is not correct. However, the lower than nominal coverage probability of ρ indicates that WN is not robust in estimating the mean resultant length of the distribution. This is seen clearly in the case of regressing WN to WC data. In the case of WC model, the location parameters are estimated well, but the mean resultant is not well estimated. The coverage probabilities are very low. One reason for this could be that we are using a high mean resultant value ($\rho = 0.8$) for simulation, and Cauchy (or WC) being a flat distribution is unable to fit the simulated data well. For WDE model, even when the distribution is erroneous, the parameters β_0 , β_1 , β_2 and ρ have been well estimated. Therefore, WDE model is robust in estimating the β_0 , β_1 , β_2 and ρ .

Table D.1: Regressing WN, WC and WDE models to WN data with categorical covariate

<i>WN</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.11	0.13	0.85	1.11	1.36	0.93
s.e.	0.66	0.22	0.52	0.70	0.86	0.01
β_1	1.39	0.12	1.14	1.40	1.60	0.93
s.e.	0.84	0.32	1.25	0.85	0.62	0.01
β_2	1.88	0.18	1.54	1.88	2.24	0.94
s.e.	0.69	0.27	1.02	0.72	0.58	0.01
ρ	0.77	0.03	0.71	0.78	0.82	0.90
s.e.	0.12	0.03	0.14	0.12	0.10	0.01
GGC	374.86	67.13	248.72	373.17	510.72	
s.e.	60.01	9.18	54.60	60.40	68.02	
Dev	207.84	3.80	203.39	207.15	216.44	
s.e.	29.55	7.49	28.14	30.45	33.94	
<i>WC</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.07	0.13	0.84	1.07	1.31	0.85
s.e.	0.56	0.19	0.51	0.56	0.75	0.02
β_1	1.43	0.12	1.19	1.43	1.65	0.87
s.e.	0.62	0.38	1.24	0.67	0.42	0.02
β_2	1.92	0.17	1.59	1.93	2.25	0.89
s.e.	0.54	0.23	0.83	0.54	0.56	0.01
ρ	0.64	0.04	0.57	0.65	0.71	0.01
s.e.	0.08	0.02	0.09	0.09	0.08	0.00
GGC	416.54	69.02	286.61	414.75	556.34	
s.e.	49.31	5.88	45.30	49.59	54.02	
Dev	230.48	3.26	226.40	229.80	238.17	
s.e.	23.55	4.75	22.81	23.82	25.32	
<i>WDE</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.04	0.12	0.83	1.04	1.27	0.87
s.e.	0.45	0.16	0.42	0.45	0.63	0.02
β_1	1.45	0.11	1.23	1.46	1.65	0.89
s.e.	0.46	0.34	1.02	0.48	0.40	0.01
β_2	1.95	0.16	1.64	1.95	2.26	0.89
s.e.	0.47	0.21	0.71	0.47	0.46	0.01
ρ	0.76	0.04	0.68	0.77	0.83	0.89
s.e.	0.08	0.04	0.12	0.08	0.07	0.01
GGC	374.18	66.99	248.94	372.19	510.43	
s.e.	51.47	9.10	45.80	51.79	59.84	
Dev	214.36	4.14	208.86	213.60	223.75	
s.e.	23.02	7.32	21.33	23.31	29.16	

Table D.2: Regressing WN, WC and WDE models to WC data with categorical covariate

<i>WN</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.13	0.16	0.83	1.12	1.44	0.92
s.e.	0.71	0.25	0.61	0.72	0.97	0.01
β_1	1.36	0.16	1.06	1.37	1.64	0.94
s.e.	0.85	0.43	1.44	0.87	0.65	0.01
β_2	1.88	0.23	1.45	1.88	2.31	0.94
s.e.	0.75	0.34	1.14	0.78	0.71	0.01
ρ	0.72	0.04	0.65	0.72	0.78	0.45
s.e.	0.14	0.03	0.15	0.14	0.11	0.02
GGC	388.33	70.11	256.82	386.45	530.26	
s.e.	69.52	9.13	63.33	70.09	76.25	
Dev	229.87	3.71	225.20	229.27	237.94	
s.e.	37.00	7.18	35.77	38.04	38.66	
<i>WC</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.08	0.10	0.92	1.07	1.28	0.93
s.e.	0.53	0.27	0.48	0.52	0.89	0.01
β_1	1.37	0.14	1.12	1.38	1.63	0.94
s.e.	0.85	0.56	1.74	0.87	0.80	0.01
β_2	1.92	0.13	1.65	1.92	2.17	0.94
s.e.	0.53	0.34	1.05	0.53	0.61	0.01
ρ	0.77	0.03	0.71	0.77	0.82	0.83
s.e.	0.13	0.04	0.15	0.13	0.11	0.02
GGC	336.94	62.42	221.92	334.61	465.21	
s.e.	71.86	10.58	62.98	72.60	81.89	
Dev	169.76	4.17	164.75	169.05	179.16	
s.e.	43.77	9.51	41.58	44.43	49.13	
<i>WDE</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.07	0.09	0.90	1.07	1.24	0.95
s.e.	0.52	0.19	0.44	0.52	0.75	0.01
β_1	1.41	0.10	1.23	1.40	1.59	0.97
s.e.	0.99	0.40	1.48	1.05	0.79	0.01
β_2	1.94	0.12	1.70	1.95	2.18	0.97
s.e.	0.50	0.26	0.87	0.49	0.52	0.01
ρ	0.80	0.03	0.73	0.80	0.85	0.71
s.e.	0.11	0.03	0.14	0.12	0.10	0.02
GGC	337.03	64.59	217.25	334.74	469.81	
s.e.	68.01	7.13	60.32	68.59	75.87	
Dev	184.23	3.91	178.59	183.72	193.37	
s.e.	37.86	5.83	36.59	38.40	40.98	

Table D.3: Regressing WN, WC and WDE models to WDE data with categorical covariate

<i>WN</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.10	0.14	0.84	1.10	1.35	0.93
s.e.	0.65	0.21	0.52	0.68	0.83	0.01
β_1	1.38	0.11	1.17	1.37	1.59	0.93
s.e.	0.89	0.27	1.28	0.90	0.51	0.01
β_2	1.90	0.18	1.55	1.89	2.25	0.92
s.e.	0.66	0.27	1.00	0.70	0.53	0.01
ρ	0.76	0.03	0.70	0.76	0.81	0.73
s.e.	0.11	0.02	0.13	0.12	0.10	0.02
GGC	373.34	67.43	246.41	371.54	510.14	
s.e.	60.86	4.04	55.62	61.16	65.09	
Dev	214.12	3.15	210.11	213.47	221.66	
s.e.	31.43	3.29	30.34	31.64	32.31	
<i>WC</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.03	0.09	0.86	1.03	1.22	0.90
s.e.	0.39	0.16	0.36	0.39	0.59	0.01
β_1	1.51	0.09	1.32	1.51	1.68	0.93
s.e.	0.53	0.39	0.86	0.54	0.83	0.01
β_2	1.97	0.12	1.72	1.97	2.21	0.91
s.e.	0.40	0.21	0.70	0.40	0.39	0.01
ρ	0.71	0.03	0.64	0.71	0.76	0.17
s.e.	0.08	0.02	0.08	0.08	0.07	0.02
GGC	380.28	66.32	256.09	378.36	515.30	
s.e.	50.39	5.28	45.75	50.78	54.02	
Dev	207.52	3.12	203.42	206.82	214.91	
s.e.	25.44	4.15	24.31	25.61	26.31	
<i>WDE</i>	mean	std dev	2.5 %	50 %	97.5 %	cov prob
β_0	1.04	0.10	0.85	1.04	1.25	0.91
s.e.	0.40	0.18	0.34	0.40	0.67	0.01
β_1	1.42	0.10	1.23	1.42	1.61	0.94
s.e.	0.74	0.33	1.21	0.73	0.64	0.01
β_2	1.96	0.14	1.68	1.96	2.24	0.92
s.e.	0.40	0.25	0.78	0.42	0.36	0.01
ρ	0.78	0.04	0.70	0.78	0.84	0.89
s.e.	0.10	0.03	0.13	0.10	0.08	0.01
GGC	357.14	66.04	234.04	355.24	491.89	
s.e.	58.02	8.90	51.51	59.13	65.88	
Dev	201.95	4.14	196.25	201.51	211.34	
s.e.	28.80	7.55	27.37	29.97	33.56	