Comparing measures of fit for circular distributions

by

Zheng Sun

B.Sc., Simon Fraser University, 2006

A Thesis Submitted in Partial Fulfillment of the

Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Mathematics and Statistics

Comparing measures of fit for circular distributions

by

Zheng Sun

B.Sc., Simon Fraser University, 2006

Supervisory Committee

Dr. William J. Reed, Supervisor

(Department of Mathematics and Statistics)

Dr. Laura Cowen, Departmental Member

(Department of Mathematics and Statistics)

Dr. Mary Lesperance, Departmental Member

(Department of Mathematics and Statistics)

Supervisory Committee

Dr. William J. Reed, Supervisor
(Department of Mathematics and Statistics)

Dr. Laura Cowen, Departmental Member
(Department of Mathematics and Statistics)

Dr. Mary Lesperance, Departmental Member
(Department of Mathematics and Statistics)

ABSTRACT

This thesis shows how to test the fit of a data set to a number of different models, using Watson's U^2 statistic for both grouped and continuous data. While Watson's U^2 statistic was introduced for continuous data, in recent work, the statistic has been adapted for grouped data. However, when using Watson's U^2 for continuous data, the asymptotic distribution is difficult to obtain, particularly, for some skewed circular distributions that contain four or five parameters. Until now, U^2 asymptotic points are worked out only for uniform distribution and the von Mises distribution among all circular distributions. We give U^2 asymptotic points for the wrapped exponential distributions, and we show that U^2 asymptotic points when data are grouped is usually easier to obtain for other more advanced circular distributions.

In practice, all continuous data is grouped into cells whose width is decided by the accuracy of the measurement. It will be found useful to treat such data as grouped with sufficient number of cells in the examples to be analyzed. When the data are

treated as grouped, asymptotic points for U^2 match well with the points when the data are treated as continuous. Asymptotic theory for U^2 adopted for grouped data is given in the thesis. Monte Carlo studies show that, for reasonable sample sizes, the asymptotic points will give good approximations to the p-values of the test.

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Bill Reed for his patience and time for meeting with me frequently throughout this thesis. Professor Reed provided me with extremely valuable advice and an abundance of theoretical help.

I would like to thank Dr. Mary Lesperance for accepting me into the program and providing me with much general advice on data analysis and Dr. Laura Cowen for providing me opportunities to work on real world problems.

I thank Dr. Michael Stephens for introducing me to goodness of fit and directional statistics. I am very grateful for his supports and encouragement.

I am also very appreciative of all the assistance I have received from professors and staff in the department, particularly Dr. Julie Zhu, Dr. Min Tsao, and Dr. Farouk Nathoo. The graduate students in our department have also been very helpful to me. I feel very fortunate to have studied in the Statistics program at the University of Victoria.

To Dad and Mum, thank you for your love and generosity which enables me to study here. To Kelly Wen Zhang, thank you for all the nice things you have done for me and for always being by my side.

Chapter 1

Introduction

In recent work, the fit of various parametric circular models has been compared using likelihood based model selection criteria (AIC, BIC, and maximized log-likelihood) (see e.g. Pewsey (2000), Gatto and Jammalamadaka (2003), Gatto and Jammalamadaka (2006), Pewsey, Lewis, and Jones (2007), Pewsey (2008), Reed and Pewsey (2008)), but no test of the adequacy of the fit of even the best fitting model (by AIC, BIC) seems to have been conducted, when data are continuous. Pewsey (2006), and Reed and Pewsey (2008) used the Pearson χ^2 goodness-of-fit statistic, calculated assuming data were grouped into a certain number of classes, to test fit. However, the statistic requires the expected counts in each class to be greater than 5. One method of doing goodness-of-fit for parametric circular models is by looking at a Q-Q plot of the ordered data $\theta_{(i)}$ against $\hat{F}^{-1}(\frac{i-0.5}{n})$, where $\hat{F}$ is the fitted distribution and n is the number of observations (see Fisher 1995, p. 124), Pewsey, Lewis, and Jones (2007)), but this is a very ad hoc method and F^{-1} can be difficult to obtain for some circular models. The other ad hoc method is by looking at a P-P plot of $F(\theta_{(i)})$ against $(\frac{i-0.5}{n})$.

It is the purpose of this thesis to rectify this fact by examing the use of Watson's

U^2 statistic for assessing the goodness of fit of parametric circular models with several parameters, and of comparing its use with the results of model selection criteria on a number of real data sets. While Watson's U^2 statistic was introduced for continuous data, in recent work, the statistic has been adapted for grouped data and we provide asymptotic theory for U^2 adopted for grouped data. When using Watson's U^2 for continuous data, the asymptotic null distribution is difficult to obtain, particularly, for some skewed circular distributions that contain four or five parameters. Until now, U^2 asymptotic percentage points have been worked out for only two circular distributions, namely the uniform distribution and the von Mises distribution (Lockhart and Stephens, 1985). We outline how asymptotic percentage points of the null distribution for U^2 can be found for any parametric model, and as an example, we then find these percentage points for the wrapped exponential model.

In practice, all continuous data are in essence grouped into cells whose width is determined by the accuracy of the measurements. Rather than treating data as continuous, a more practical alternative when using U^2 is to treat the data as grouped data with large number of cells because of the relative ease of obtaining percentage points. When the data are treated as grouped, asymptotic points for U^2 match well with the points when the data are treated as continuous for uniform, von Mises and wrapped exponential distributions, and they also match well in all cases with percentage points obtained by parametric bootstrap procedures.

We use these methods to test for adequacy of fit for various models using a number of data sets (both grouped and ungrouped).

1.1 Circular data

Circular data arise in various ways. Two of the most common correspond to two circular measuring instruments: the compass and the clock. Typical types of data measured by the compass include wind directions, ocean current directions, directions and orientations of birds and animals, and orientation of geological phenomena such as rock cores and fractures. Typical data measured by the clock includes arrival times of patients in an emergency clinic, incidences of a disease throughout the year, and the number of tourists (daily or monthly) in a city within a year, where the calendar is regarded as a one-year clock. As can be seen, circular or directional data arise in many scientific fields, including Biology, Geology, Geography, Meteorology, and Physics.

Research on directional data can be dated back to the 18th century. In 1734 Daniel Bernoulli proposed to use the resultant length of normal vectors to test for uniformity of unit vectors on the sphere. Lord Rayleigh (1880) studied the distribution of the resultant length of normal vectors and developed Rayleigh's one sample test. Von Mises (1918) introduced a distribution on the circle by using a characterization analogous to Gauss's characterization of the normal distribution on a line. Later, interest was renewed in spherical and circular data by R. A. Fisher (1953), and Watson and William (1956). Stephens studied the small sample theory (1962) and the distributions of various goodness-of-fit statistics (1964) (1965). Batschelet's (1981) influential book introduced the field to biologists. The books by Mardia (1972), N. I. Fisher (1993), Mardia and Jupp (1999) and Jammalamadaka and SenGupta (2001) provide many statistical methods for analyzing circular data.

1.2 Content of the thesis

The main purpose of this work is to apply goodness-of-fit methods to parametric circular models, where they have seldom been used before. By treating continuous data as grouped data with many cells, it is demonstrated by Monte Carlo studies that the asymptotic null distribution of Watson's U^2 statistic can be found accurately. The structure of this thesis is as follows:

1. Chapter 2 and 3 give an introduction to circular statistics and parametric circular distributions

2. Chapter 4 fits parametric models using maximum likelihood estimation for grouped and ungrouped data

3. Chapter 5 introduces the Watson's goodness-of-fit statistic

4. Chapter 6 gives asymptotic theory

5. Chapter 7 compares asymptotic percentage points with parametric bootstrap percentage points

6. Chapter 8 presents applications to several data sets

7. Chapter 9 are conclusions and future work.

Chapter 2

Circular data and sample statistics

A directional observation can be regarded as a vector $\vec{OP}$ from the center O to a point P on a unit circle. Once an initial direction and an orientation of the circle have been chosen, each directional observation can be specified by an angle from the initial direction.

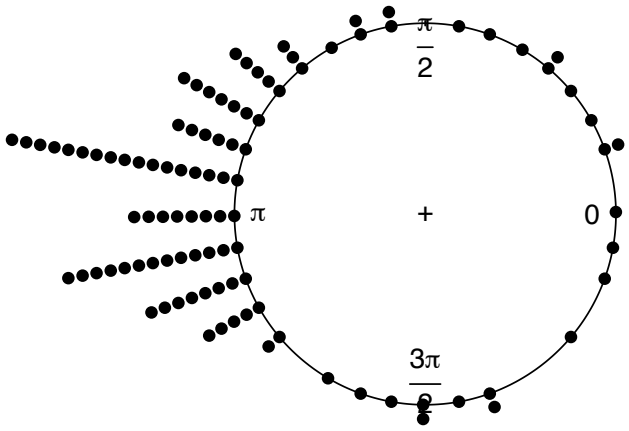


Figure 2.1: Circular data plot of orientations of 100 ants, measured in radians.

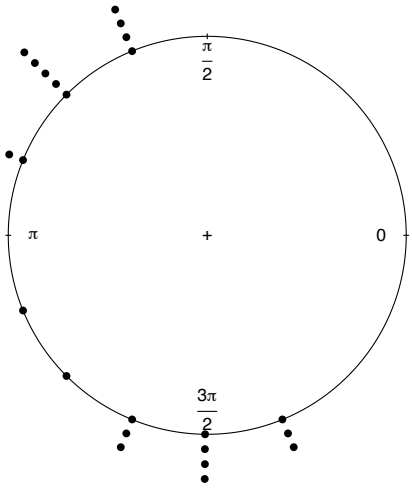


Figure 2.2: Circular data plot of 23 fish reacting to artificial sunlight, measured in radians.

A typical plot of circular observations is shown in Figure 2.1. The plot shows directions chosen by 100 ants in response to an evenly illuminated black target that was placed at π (180°) are recorded. The data are taken from Fisher (1995, p.243) and are a random sample as part of a larger sample from Jander (1957, Figure 18A). Data are collected in an arena, each ant was placed individually into the arena and the optical orientation of the ant was recorded. The ants tended to run toward the stimulus. These data will be analyzed in section 8.1.

The data of Figure 2.2 comes from a study of fish reacting to sunlight (at two times of day), or to an overcast sky (no sunlight), and finally to an articial light which does not move with the sun. The data set was given by Braemer (1960) and quoted by

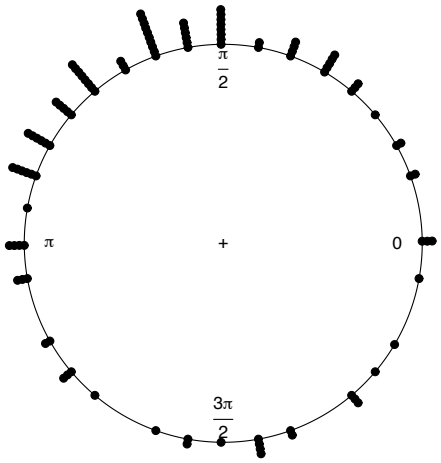


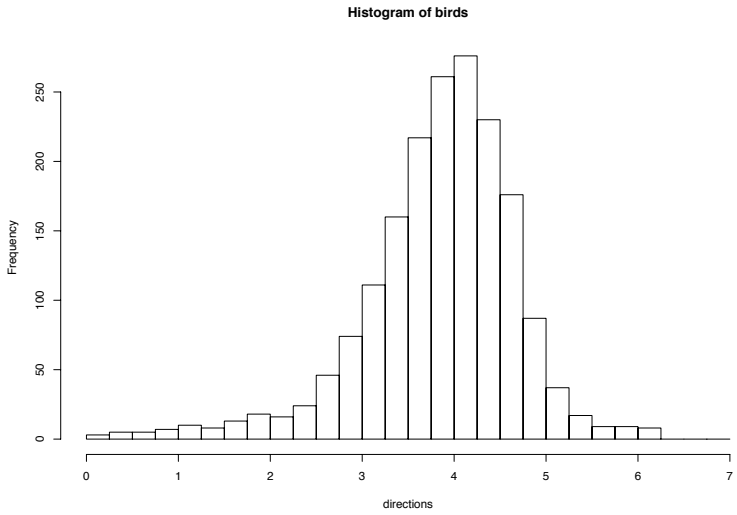
Figure 2.3: Circular data plot of 104 cross-bed measurements from Himalayan molasse, measured in radians.

Schmidt-Koenig (1975). The directions taken by 23 fish in the final group (articial light) were measured anti-clockwise from the x-axis, and the numbers in k = 16 cells were 0,0,0,0,0,4,5,2,0,1,1,3,4,3,0,0. The boundaries for the first cell are $-11.25°$ and $11.25°$. These data will be analyzed in section 8.2.

The circular plot in Figure 2.3 shows 104 measurements of directions of sand stone rocks from Himalayan molasse in Pakistan, taken from Fisher (1995, pp. 250-251) and are originally reported by Wells (1990); these data will be analyzed in section 8.3.

A histogram of the headings of 1827 flight directions recorded at an observational post near Stuttgart during the autumn migratory period of 1987, and reported in

Figure 2.4: Histogram of the headings of 1827 migrating birds in Germany, direction measured clockwise from north in radians .



Bruderer and Jenni (1990) is presented in Figure 2.4. The directions are measured clockwise from the north to the nearest degree. Like the sandstone data, they are grouped data with cell width 1°. These data will be analyzed in section 8.4.

Due to the geometry of the sample space, circular data cannot be modelled using standard statistical techniques. For example, the sample mean of a data set on the circle (circular mean) is not the usual sample mean (linear mean). To see this, consider a data set with two observations $y_1 = 1°$ and $y_2 = 359°$, the usual sample mean $\frac{1°+359°}{2} = 180°$, but the sample mean direction is 0°-right opposite! To define the circular mean, one needs to combine all the observations as unit vectors. The usual way to combine unit vectors is vector addition: the direction of the resultant vector will be defined as the mean direction of the individual vectors and the length of the resultant vector will be defined as the mean of the vectors. Suppose a sample is given by n unit vectors OP_i, $i = 1, \cdots, n$, from the center O of a circle with radius 1, to points P_i on the circumference of the circle. Let θ_i be the angular coordinate of OP_i; and let

$\bar{S} = \frac{1}{n}\sum_{i=1}^{n}\sin(\theta_i)$ and $\bar{C} = \frac{1}{n}\sum_{i=1}^{n}\cos(\theta_i)$. Then the sample mean direction is defined as

$$\overline{\theta} = \begin{cases} \tan^{-1}(\frac{\bar{S}}{\bar{C}}) & \bar{C} \geq 0 \\ \tan^{-1}(\frac{\bar{S}}{\bar{C}}) + \pi & \bar{C} < 0, \end{cases} \tag{2.1}$$

where the function $\tan^{-1}()$ is the inverse tangent function which takes values in $[-\pi/2, \pi/2]$. and the sample mean resultant length is defined as

$$\bar{R} = \sqrt{\bar{S}^2 + \bar{C}^2},\ \bar{R} \in [0, 1]. \tag{2.2}$$

The size of $\bar{R}$ will depend on the spread of the data. If the sample can be regarded as randomly scattered around the circle, $\bar{R} \to 0$ as $n \to \infty$ and if the sample is concentrated at $\bar{\theta}$, $\bar{R} \to 1$ as $n \to \infty$.

The pth sample central trigonometric moments about zero direction are defined as

$$a_p = \frac{1}{n}\sum_{i=1}^{n}\cos(p\theta_i) \text{ and } b_p = \frac{1}{n}\sum_{i=1}^{n}\sin(p\theta_i).$$

Hence $\bar{C}$ and $\bar{S}$ are the first sample trigonometric moments about the zero direction.

Chapter 3

Parametric distributions on circle

A circular distribution is a probability distribution around a unit circle from 0 to 2π, and it assigns probabilities or probability densities to different directions. Circular distributions are important because they can provide good summaries of some data sets. The probability density function (p.d.f.) of a continuous circular distribution has the following three properties:

1. $f(\theta) \geq 0$
2. $\int_0^{2\pi} f(\theta)\,\mathrm{d}\theta = 1$
3. $f(\theta) = f(\theta + 2\pi k)$ for any integer k.

The cumulative density function (c.d.f.) F is defined as

$$F(\omega) = P(0 \leq \theta \leq \omega) = \int_0^{\omega} f(\theta)\mathrm{d}\theta,\ \omega \in [0, 2\pi),$$

and by definition, $F(0) = 0$ and $F(2\pi) = 1$.

The characteristic function (char. f.) of a circular random variable (r.v.) Θ is defined

as

$$\phi(t) = E(e^{it\Theta}),$$

and the value of the above function evaluated at integer p is called the pth trigonometric moment of Θ

$$\phi(p) = E(e^{ip\Theta}) = \int_0^{2\pi} e^{ip\Theta} \mathrm{d}F(\Theta),\ p = 0, \pm 1, \pm 2, ...$$

By Euler's formula, it follows that

$$E(e^{ip\Theta}) = E\{\cos(p\Theta) + i\sin(p\Theta)\} = \int_0^{2\pi} \cos(p\Theta)\mathrm{d}F(\Theta) + i\int_0^{2\pi} \sin(p\Theta)\mathrm{d}F(\Theta);$$

therefore, the pth trigonometric moment can be written as:

$$\phi(p) = \alpha_p + i\beta_p,$$

where

$$\alpha_p = E\{\cos(p\Theta)\} = \int_0^{2\pi} \cos(p\Theta)\mathrm{d}F(\Theta) \tag{3.1}$$

and

$$\beta_p = E\{\sin(p\Theta)\} = \int_0^{2\pi} \sin(p\Theta)\mathrm{d}F(\Theta). \tag{3.2}$$

Note that α_p and β_p are the population analogues for the sample trigonometric moments a_p and b_p. When $p = 1, \phi(1) = \alpha_1 + i\beta_1 = \rho e^{i\mu}$, where $\mu = \arctan\frac{\beta_1}{\alpha_1}$ is the mean direction and $\rho = \sqrt{\alpha_1^2 + \beta_1^2}$ is the mean resultant length. The quantities μ and ρ are the population analogues for $\bar{\theta}$ and $\bar{R}$. The mean direction ρ lies within [0,1], and it characterizes the spread of the distribution. If $\rho = 0$, the distribution can be regarded as uniformly distributed around the circle and if $\rho = 1$, the distribution is

concentrated at μ. In the next few sections, a number of parametric distributions which will be used in later chapters are introduced.

3.1 The uniform distribution

With this model, all directions on the unit circle are chosen with equal probability, that is, data are treated as having no preferred direction and the mean direction is undefined. The distribution has the probability density function (p.d.f.)

$$f(\theta) = \frac{1}{2\pi},\ \theta \in [0, 2\pi),$$

cumulative density function (c.d.f.)

$$F(\theta) = \frac{\theta}{2\pi},\ \theta \in [0, 2\pi),$$

and characteristic function (char.f.)

$$\phi_p = \begin{cases} 1, & p = 0 \\ 0, & p \neq 0. \end{cases}$$

Tests for randomness:

There are many statistics available to test whether the population from which the sample is drawn is compatible with a uniform distribution (i.e. testing the null hypothesis that the parent population is uniformly distributed). The most well known are the Watson's U^2 statistic (Watson, 1961, Stephens, 1964a), the Rayleigh's R statistic. The other tests are Kuiper's test (Kuiper, 1960, Stephens, 1965a), the V test (Greenwood and Durand, 1955), the Hodges' and Ajne's test (Hodges, 1955,

Ajne, 1968), the range test (Laubscher and Rudolph, 1968), and the Rao's spacing test (Rao, 1969, 1976).

Method of simulation:

Uniform random variables (r.v.'s) can be obtained from standard packages in any statistical software. After generating uniform r.v.'s U on the interval $[0,1]$ (e.g. in program R, use command "runif"), one needs to transform U to the interval $[0,2\pi]$ by simple multiplication to obtain values of Θ.

3.2 The von Mises (VM) distribution

This distribution was introduced by von Mises (1918). It has been named the circular normal distribution by Gumbel, Greenwood, and Durand (1953) since it plays a similar role to that of the normal distribution on a line. The von Mises distribution is the most commonly used distribution for symmetric and unimodal samples of circular data. The p.d.f. of the VM distribution is

$$f(\theta;\mu,\kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa\cos(\theta-\mu)\},\ \theta \in [0,2\pi),\ \kappa \in [0,\infty)$$

where

$$I_0(\kappa) = \frac{1}{2\pi}\int_0^{2\pi} \exp\{\kappa\cos(\theta)\}\,\mathrm{d}\theta$$

is a modified Bessel function of order zero. Modified Bessel functions need to be evaluated numerically, for example, using command "I.0" in R package "circular". Also Abramowitz and Stegun (1970) give polynomial approximations to $I_0(\kappa)$. The

c.d.f. of the VM distribution is

$$F(\theta;\mu,\kappa) = \frac{1}{I_0(\kappa)}\int_0^{\theta} \exp\{\kappa\cos(u)\}\,\mathrm{d}u,\ \theta\in[0,2\pi),\ \kappa\in[0,\infty);$$

and the char. f. of the VM distribution is

$$\phi_p = e^{ip\mu}\frac{I_p(\kappa)}{I_0(\kappa)},\ p=0,\pm1,\pm2,\cdots.$$

The distribution has a maximum value at $\theta = \mu$ and it is symmetric around μ, which is therefore the modal and mean direction. The parameter κ is a concentration parameter. As $\kappa \to 0$, the distribution degenerates into a uniform distribution; as $\kappa \to \infty$, the distribution concentrates in the direction of μ. For $k = 4$, over 99% of the probability lies in the arc $[\mu - \frac{\pi}{2}, \mu + \frac{\pi}{2}]$ (see Figure 3.1). Generalizations of the von Mises distributions for modeling unimodal or bimodal, and for modeling symmetric or skewed data were proposed by Cox (1975) and Yfantis and Borgman (1982). Later, Gatto and Jammalamadaka (2006) presented generalization which allowed for modeling multimodal data.

Watson's U^2 statistic (Lockhart and Stephens, 1985) and Cox's test (Barndorff-Neilson and Cox, 1980) can be used to test the hypothesis that the population from which the sample is drawn follows a von Mises distribution. There are many other procedures developed under the assumption that samples are drawn from von Mises distributions (for example, tests and confidence intervals for μ with the concentration parameter κ assumed both known and unknown, tests and confidence intervals for κ, two sample procedures (Watson and William, 1956, Stephens, 1972), and multi-sample tests (Stephens 1972)).

Method of simulation:

A method of generating von Mises pseudo r.v.'s is given in Fisher (1993); the algorithm was developed by Best and Fisher (1979). A different algorithm which is more efficient when κ is changing from call to call was suggested by Dagpunar (1990). The command "rvonmises" in *R* package "circular" written by Agostinelli, Lund and Southworth (2007) was used to generate VM pseudo r.v.'s.

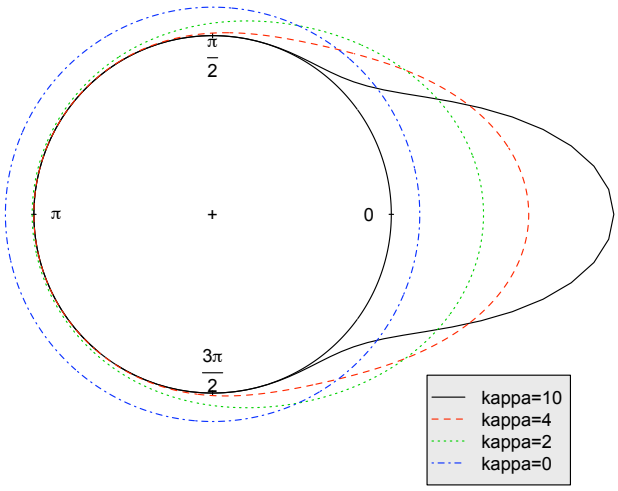


Figure 3.1: A circular plot of the von Mises distributions for different κ parameter values.

Figure 3.1 is a circular plot of the VM distributions with $\mu = 0$ and different values of κ . It shows the effect of the parameter κ; as κ increases, the density concentrates around $\mu = 0$.

3.3 Circular beta (CB) distribution

One way to generate a family of circular distributions is based on the tangent normal decomposition of a distribution on the unit circle (Saw, 1978) (Johnson and Wehrly 1978). Specifically, one can map the symmetric distributions on the unit circle to the distributions on the interval [-1,1] via a one to one relationship and vice versa (see Lai, 1994, for details). Let $h(t)$ be the distribution on the interval [-1,1], then the symmetric distribution $g(t)$ on the circle is

$$g(t) = \frac{1}{2}h(t)(1-t^2)^{\frac{1}{2}}. \tag{3.3}$$

The density of x on the circle is $f(x) = g(\mu' x)$, where x is a unit vector from the center O of a circle to points P on the circumference of the circle, and μ is the central vector.

The well known Beta distribution on the real line has the density

$$h(s) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} s^{\alpha-1}(1-s)^{\beta-1},\ s \in [0,1].$$

Let $T = 2S - 1$, then the r.v.'s T have the p.d.f

$$h(t) = \frac{1}{2^{\alpha+\beta-1}B(\alpha,\beta)}(1+t)^{\alpha-1}(1-t)^{\beta-1},\ t \in [-1,1],$$

where $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ is the Beta function with $\alpha > 0$, $\beta > 0$. Following (3.3),

$$g(t) = \frac{1}{2^{\alpha+\beta}B(\alpha,\beta)}(1+t)^{\alpha-1/2}(1-t)^{\beta-1/2}.$$

Hence, the density of the directional beta distribution on the circle is

$$f(x) = \frac{1}{2^{\alpha+\beta}B(\alpha,\beta)}(1+\mu' x)^{\alpha-1/2}(1-\mu' x)^{\beta-1/2}. \tag{3.4}$$

Equivalently, using the polar coordinate system, θ is the angle made by the vector x on the positive x-axis, and the location parameter η is the angle made by the central vector μ on the positive x-axis, then $\mu' x = \cos(\theta - \eta)$. The equation (3.4) can be written as

$$f(\theta;\alpha,\beta) = \frac{1}{2^{\alpha+\beta}B(\alpha,\beta)}\{1+\cos(\theta-\eta)\}^{\alpha-1/2}\{1-\cos(\theta-\eta)\}^{\beta-1/2},\ \theta \in [0,2\pi). \tag{3.5}$$

When both α and β are equal to 0.5, the CB distribution becomes a uniform distribution; when both α and β are less than 0.5, the density goes to ∞ at 0° and 180°; when both $\alpha = 0.5$ and $\beta < 0.5$, the mode is at 0°; when both α and β are greater than 1, the density becomes bimodal (see figure 3.2).

Method of simulation:

An algorithm for generating CB pseudo r.v.'s is available in Lai (p.150, 1994). The command "rbeta" in program R was used to generate the Beta pseudo r.v.'s X with parameters α and β. To obtain CB r.v.'s Θ, let $Y = 2X - 1$, and generate Θ as

$$\Theta = \begin{cases} \cos^{-1}(Y), & \text{with probability} = \frac{1}{2} \\ 2\pi - \cos^{-1}(Y), & \text{with probability} = \frac{1}{2}, \end{cases}$$

and $\Theta + \mu$ if the location parameter is presented.

Figure 3.2 is a circular plot of the CB distributions for different values of α and different values of β. It verifies the discussion above about the shapes of the CB density when α and β equal different values.

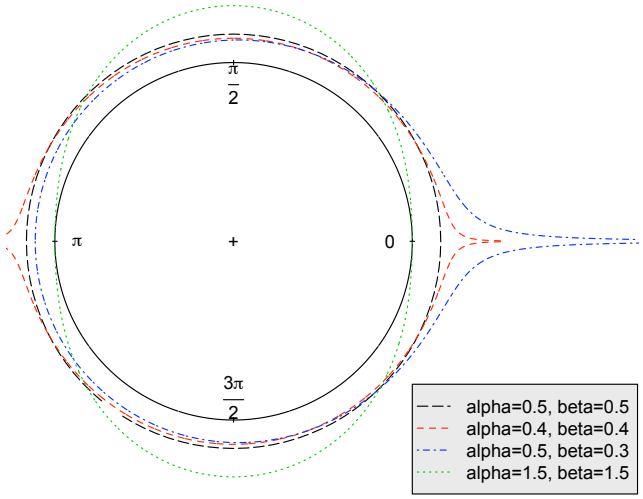


Figure 3.2: A circular plot of the Circular Beta distributions for different values of α and different values of β.

3.4 Wrapped distributions

Circular distributions can be obtained by wrapping distributions on the real line around a unit circle. Such an idea has a long history (i.e., Schmidt (1917) fitted a wrapped normal distribution to some geological data). In general, if X is any r.v.'s on the real line with p.d.f. $g(x)$, and c.d.f. $G(x)$, one can obtain circular r.v.'s Θ by defining:

$\Theta \equiv X (\text{mod } \ 2\pi)$.

The p.d.f. of Θ, $f(\theta)$, is obtained by wrapping $g(x)$ around the circumference of a unit circle and summing up the overlapping points:

$$f(\theta) = \sum_{k=-\infty}^{\infty} g(\theta + 2\pi k), \ \theta \in [0, 2\pi). \tag{3.6}$$

The c.d.f. is

$$F(\theta) = \sum_{k=-\infty}^{\infty} \{G(\theta + 2\pi k) - G(2\pi k)\}. \tag{3.7}$$

Stephens (1963) used a Fourier series expansion to investigate the wrapped normal distribution, discussed below. This technique has been adapted for other wrapped distributions and leads to the following results.

Let $f(\theta)$ denote a p.d.f. of variable Θ, which has a period of 2π. Then $f(\theta)$ can be written as an infinite sum of *sine* and *cosine* functions on the interval $[0, 2\pi)$.

$$f(\theta) = \frac{1}{2\pi}\left[1 + 2\sum_{p=1}^{\infty}\{\alpha_p \cos(p\theta) + \beta_p \sin(p\theta)\}\right], \tag{3.8}$$

where the Fourier coefficients α_p and β_p are determined as

$$\alpha_p = \int_0^{2\pi} \cos(p\theta)\, \mathrm{d}F(\theta),$$

and

$$\beta_p = \int_0^{2\pi} \sin(p\theta)\, \mathrm{d}F(\theta).$$

Note that $\phi_p = \alpha_p + i\beta_p$ constitutes the pth trigonometric moment of Θ introduced in (3.1) and (3.2).

The following proposition is from Mardia (1972). The result provides a way to obtain the characteristic function for wrapped distributions once the functions for the distribution on the real line is known.

Proposition 3.4.1. *The trigonometric moment of order p for a wrapped circular distribution corresponds to the value of the characteristic function of the unwrapped random variable X, say, $\Phi_X(t)$ at the integer value p, i.e., $\phi_p = \Phi_X(p)$.*

The Fourier expansion (3.8) will be used throughout the thesis to represent the

likelihood functions of wrapped distributions.

3.4.1 Wrapped Normal (WN) Distribution

The WN distribution is a symmetric unimodal distribution which is obtained by wrapping a normal distribution with mean μ and variance σ^2 around the circle. It arises as the distribution of the location after a fixed time of a particle following a random walk or Brownian motion on the circle (Stephens, 1963).

From (3.6), the p.d.f of the WN distribution is

$$f(\theta;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\sum_{p=-\infty}^{\infty}\exp\left\{\frac{-(\theta-\mu-2\pi p)^2}{2\sigma^2}\right\},\ \theta\in[0,2\pi). \tag{3.9}$$

Let X be $N(\mu,\sigma^2)$ r.v.'s defined on the real line. The char.f. is $\phi(t) = E\{\exp(itX)\} = \exp(i\mu t - \frac{t^2\sigma^2}{2})$. It then follows from Proposition 3.4.1 that the char.f. of circular r.v.'s $\Theta = X(\text{mod } 2\pi)$ is

$$\phi(p) = \exp\left(i\mu p - \frac{p^2\sigma^2}{2}\right),\ p = 0,\pm1,\pm2,...,$$

so that

$$E(e^{ip\Theta}) = \exp\left(-\frac{p^2\sigma^2}{2}\right)e^{i\mu p} = \alpha_p + i\beta_p,$$

where

$$\alpha_p = \exp\left(-\frac{p^2\sigma^2}{2}\right)\cos(p\mu)$$

and

$$\beta_p = \exp\left(-\frac{p^2\sigma^2}{2}\right)\sin(p\mu).$$

After substituting α_p and β_p into (3.8), and using the difference formula for the cosine function, an alternative form of the density is

$$f(\theta;\mu,\rho) = \frac{1}{2\pi}\left\{1 + 2\sum_{p=1}^{\infty}\rho^{p^2}\cos p(\theta-\mu)\right\},\ \rho \in [0,1],\ \theta \in [0,2\pi), \tag{3.10}$$

where $\rho = \exp(-\frac{\sigma^2}{2})$ is the mean resultant length and μ is the mean direction in $[0, 2\pi)$. Like the normal distribution on the line, the WN distribution has the additive property (i.e., the sum of two independent WN r.v.s, $\Theta_1 \backsim WN(\mu_1, \rho_1)$, $\Theta_2 \backsim WN(\mu_2, \rho_2)$, is a WN r.v. $\Theta_1 + \Theta_2 \backsim WN(\mu_1 + \mu_2, \rho_1\rho_2)$).

To estimate μ and σ^2 numerically, one can use the sample quantities $\bar{\theta}$ and $-2\log(\bar{R})$ as starting values.

Stephens (1963) showed that the WN distribution can be closely approximated by the VM distribution (see Stephens, 1963, table 1 for the relationship between distribution parameters at best fit). The approximate relationship is $\exp(-\sigma^2/2) = I_1(\kappa)/I_0(\kappa)$, where $I_0(\kappa)$ and $I_1(\kappa)$ are the modified Bessel functions of order zero and one, as discussed in section 3.2. In practice, when $\sigma^2 > 2\pi$, using the first 3 terms of the infinite sum in (3.10) gives adequate approximation of the WN density; when $\sigma^2 \leq 2\pi$, the term with $p = 0$ of (3.9) gives a reasonable approximation of the WN density (Mardia and Jupp, 1999, p. 50). In this thesis, all calculations with the WN used the first 20 terms of (3.10). As $\rho \to 0$, the distribution approaches the uniform distribution; as $\rho \to 1$, the distribution concentrates in the direction of μ (see Figure 3.3).

Figure 3.3 is a circular plot of the WN distributions for different values of standard deviation (s.d.). It shows the effect of the parameter σ; as σ decreases ($\rho \to 1$) , the

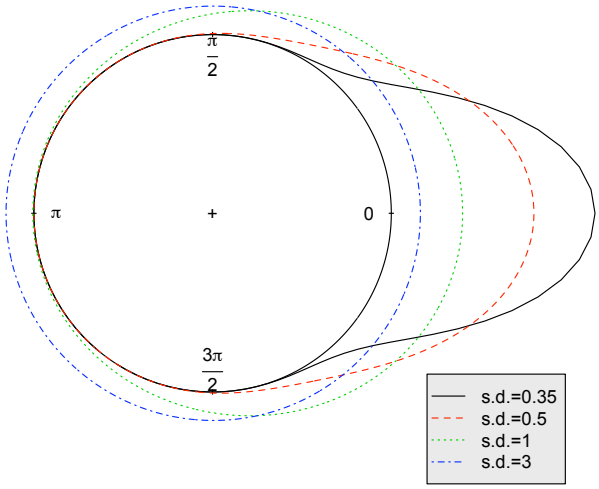


Figure 3.3: A circular plot of the wrapped Normal distributions for different values of the parameter σ (s.d.).

density concentrates around $\mu = 0$.

Method of simulation:

The WN pseudo r.v.'s can be obtained by generating the normal pseudo r.v.'s on the real line, and then considering them modulo 2π. The command "rwrappednormal" in *R* package "circular" written by Agostinelli, Lund, and Southworth (2007) was used to generate the WN pseudo r.v.'s.

3.4.2 Wrapped Cauchy (WC) distribution

The Cauchy distribution was introduced by Lévy (1939). A wrapped Cauchy distribution is symmetric and unimodal and it is obtained by wrapping a Cauchy distribution

around a unit circle. Let X be Cauchy r.v.s on the real line which has the p.d.f.

$$f(x;\mu,\rho) = \frac{\rho}{\pi\{\rho^2 + (x-\mu)^2\}},\ x \in (-\infty,\infty),\ \rho \in [0,\infty)$$

with a char. f. of $\phi(t) = E(e^{itX}) = \exp(i\mu t - \rho|t|)$. From Proposition 3.4.1, it follows that a WC distribution has a char. f.

$$\phi(p) = \exp(i\mu p - \rho|p|),\ p = 0, \pm 1, \pm 2, ...$$

So that the pth trigonometric moment of a WC distribution is

$$E(e^{ip\Theta}) = e^{-\rho|p|}e^{i\mu p} = \alpha_p + i\beta_p,$$

where

$$\alpha_p = e^{-\rho|p|}\cos(p\mu)$$

and

$$\beta_p = e^{-\rho|p|}\sin(p\mu).$$

After substituting α_p and β_p into (3.8), and using the difference formula for the cosine function, one obtains the p.d.f. of the WC distribution as

$$f(\theta;\mu,\rho) = \frac{1}{2\pi}\left\{1 + 2\sum_{p=1}^{\infty} e^{-\rho p}\cos p(\theta-\mu)\right\},\ \theta \in [0,2\pi),\ \rho \in [0,1], \tag{3.11}$$

where μ is the mean direction in $[0,2\pi)$ and $e^{-\rho}$ is the mean resultant length. By summing the real part of the geometric series

$$\sum_{p=1}^{\infty} \rho^p e^{-ip(\theta-\mu)},$$

(3.11) can be shown to have a closed form

$$f(\theta;\mu,\rho) = \frac{1-\rho^2}{2\pi\{1+\rho^2-2\rho\cos(\theta-\mu)\}},$$

and the c.d.f. of the WC distribution is

$$f(\theta;\mu,\rho) = \frac{1}{2\pi}\cos^{-1}\left\{\frac{(1+\rho^2)\cos(\theta-\mu)-2\rho}{1+\rho^2-2\rho\cos(\theta-\mu)}\right\},\ \theta\in[0,2\pi).$$

As $\rho\to 0$, the distribution approaches the uniform distribution; as $\rho\to 1$, the distribution becomes the point distribution concentrated in the direction of μ (see Figure 3.4). For an appropriate choice of ρ, the WC distributions are quite similar to the von Mises distributions. The WC distributions can be generalized to wrapped t and wrapped stable distributions, which will be discussed later.

To estimate μ and ρ numerically, one can use sample quantities $\bar{\theta}$ and $-\log(\bar{R})$ as starting values.

Figure 3.4 is a circular plot of the WC distributions for different values of parameter ρ. It shows the effect of the parameter ρ; as ρ approaches to 1, the density concentrates around $\mu = 0$.

Method of simulation:

An algorithm for generating WC pseudo random variables was given by Fisher (1993, p.46). The command "rwrappedcauchy" in R package "circular" was used to generate wrapped Cauchy r.v.'s . Also, WC r.v.'s can be obtained by taking modulo 2π of the corresponding Cauchy r.v.'s on the real line.

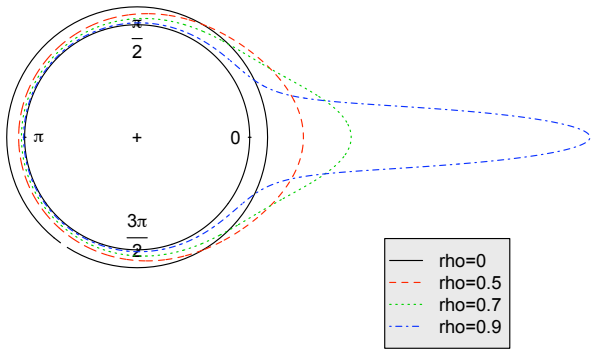


Figure 3.4: A circular plot of the wrapped Cauchy distributions for different values of ρ.

3.4.3 Wrapped t (WT)

The WC distributions discussed in the last section is a special case of WT distributions with the degrees of freedom (ν) equals 1. A natural extension is to consider the WT distribution where $\nu > 1$. First considered by Kato and Shimizu (2005) and then by Pewsey, Lewis and Jones (2007), the three parameter, symmetric, unimodal WT distributions are obtained by wrapping a shifted and scaled t distribution onto the unit circle. The family contains the WN and the WC distributions as special cases. The Student-t r.v.'s X on the real line has a p.d.f.

$$f(x;\nu) = c\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \ \nu \in (0,\infty), \tag{3.12}$$

where $c = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}}$. Let $Y = \mu + \lambda X$ be the shifted and scaled Student t r.v.'s. Then wrapping Y around the unit circle, one obtains, from (3.6), the p.d.f. of the WT distribution as

$$f(\theta;\mu,\lambda,\nu) = \frac{c}{\lambda}\sum_{p=-\infty}^{\infty}\left\{1+\frac{(\theta+2\pi p-\mu)^2}{\lambda^2\nu}\right\}^{-\frac{\nu+1}{2}},\ \nu\in(0,\infty),\ \theta\in[0,2\pi). \tag{3.13}$$

The char. f of X on the real line has been studied by Lebedev (1965, Chapter 5) and later by Hurst (1995). For $t>0$,

$$\phi_X(t) = E\{\exp(itX)\} = \frac{K_{\nu/2}(t\sqrt{\nu}) * (t\sqrt{\nu})^{(\nu/2)}}{\Gamma(\nu/2)2^{\nu/2-1}},$$

and it can be shown that the char. f of Y on the real line for $t>0$,

$$\phi_Y(t) = E\{\exp(itY)\} = \frac{e^{i\mu t}K_{\nu/2}(t\lambda\sqrt{\nu}) * (t\lambda\sqrt{\nu})^{(\nu/2)}}{\Gamma(\nu/2)2^{\nu/2-1}},$$

where

$$K_\omega(x) = \frac{1}{2}\int_0^\infty s^{\omega-1}\exp\left\{-\frac{x}{2}\left(s+\frac{1}{s}\right)\right\}ds.$$

It then follows from Proposition 3.4.1 that the char.f. of circular r.v. $\Theta = Y(\text{mod } 2\pi)$, (Pewsey, Lewis and Jones (2007) gave the char.f. of Θ when $\mu=0$) is

$$\phi_Y(p) = E\{\exp(ip\Theta)\} = e^{ip\mu}\frac{K_{\nu/2}(p\lambda\sqrt{\nu}) * (p\lambda\sqrt{\nu})^{(\nu/2)}}{\Gamma(\nu/2)2^{\nu/2-1}},\ p=0,\pm1,\pm2,...,$$

so that

$$\alpha_p = \cos(p\mu)\frac{K_{\nu/2}(p\lambda\sqrt{\nu}) * (p\lambda\sqrt{\nu})^{(\nu/2)}}{\Gamma(\nu/2)2^{\nu/2-1}}$$

and

$$\beta_p = \sin(p\mu)\frac{K_{\nu/2}(p\lambda\sqrt{\nu}) * (p\lambda\sqrt{\nu})^{(\nu/2)}}{\Gamma(\nu/2)2^{\nu/2-1}}.$$

After substituting α_p and β_p into (3.8), and using the difference formula for the cosine function, an alternative form of the density is

$$f(\theta;\nu,\lambda) = \frac{1}{2\pi}\left\{1 + 2\sum_{p=1}^{\infty}\frac{K_{\nu/2}(p\lambda\sqrt{\nu}) * (p\lambda\sqrt{\nu})^{(\nu/2)}}{\Gamma(\nu/2)2^{\nu/2-1}}\cos p(\theta-\mu)\right\},\ \nu \in (0,\infty),\ \theta \in [0,2\pi).$$

The parameters ν and λ determine the peakness and the concentration of a WT distribution. When $\nu = 1$, the WT distribution is a wrapped Cauchy (WC) distribution with $\rho = e^{-\lambda}$. It converges to a wrapped normal (WN) distribution as $\nu \longrightarrow \infty$. For $0 < \nu < 1$, the WT distribution has a higher peak and heavier tails than the WC distribution. By letting ν become big, a WT can be used to model data that have high density at the mode (see Figure 3.5). When $\nu \longrightarrow 0$, the WT distributions require a large number of the central terms in the infinite summation to achieve convergence (Pewsey, Lewis, and Jones 2007). The same situation happens when $\lambda \longrightarrow \infty$.

To estimate the parameters numerically, one can use the sample mean direction $\bar{\theta}$ as a starting value for μ and identify a suitable region of λ and ν from the contour plot of the profile likelihood as starting values for λ and ν(Pewsey, Lewis, and Jones 2007).

Method of simulation:

The algorithm for generating Student-t pseudo r.v.'s X is well developed (in program R, command "rt"). For such pseudo r.v.'s X, the corresponding wrapped t (WT) r.v.'s Θ can be obtained as $(\mu + \lambda X) \pmod{2\pi}$.

Figure 3.5 shows two circular plots of WT distributions for different values of ν and values of λ. The plot on the left shows the effect of the parameter ν; as ν

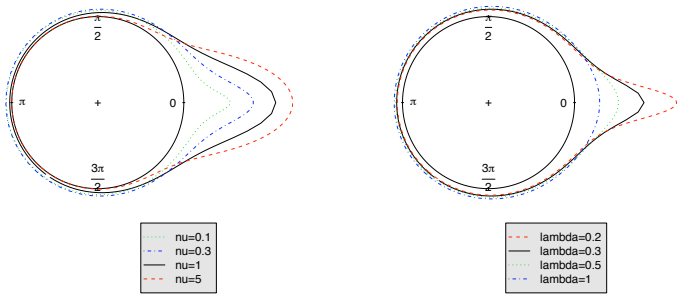


Figure 3.5: A circular plot of the wrapped t distributions for different values of ν and different values of λ.

increases, the density concentrates (narrow and high peak) around $\mu = 0$. The plot on the right shows the effect of the parameter λ; as λ decreases, the density concentrates around $\mu = 0$.

3.5 Skewed distributions

In early work on directional statistics, most attention was given to symmetric distributions, and many biological data sets were assumed to be symmetric. However, more recently, interest has increased for skewed distributions to handle data which are obviously not symmetric, see e.g. Figure 2.4 which shown a histogram of directions taken by 1827 birds recorded at an observational post near Stuttgart, Germany during the autumn migrating period of 1987 (Bruderer and Jenni (1990)).

A number of parametric skewed distributions which could be used for such data set are discussed in this section.

3.5.1 Batschelet's distribution

Batschelet introduced a family of skew distributions with two parameters. The Batschelet r.v. Θ has the p.d.f.

$$f(\theta;\kappa,\nu) = \frac{1}{2\pi} + \frac{\kappa}{2\pi}\sin(\theta + \nu\sin\theta),\ \kappa \in [-1,1],\ \nu \in [-1,1]. \tag{3.14}$$

By letting $\Psi = \Theta - \frac{\pi}{2}$, the p.d.f. of the r.v. Ψ is

$$f(\psi;\kappa,\nu) = \frac{1}{2\pi} + \frac{\kappa}{2\pi}\cos(\psi + \nu\cos\psi). \tag{3.15}$$

With fixed ν, a distribution with $\kappa = a$ is a mirror image (at 0 degree) of a distribution with $\kappa = -a$; with fixed κ, a distribution with $\nu = a$ is a mirror image (at 90 degree) of a distribution with $\nu = -a$. When $\kappa = 0$, both (3.14) and (3.15) degenerate to a uniform distribution (see Figure 3.6) and when $\nu = 0$, (3.15) becomes

$$f(\psi;\kappa,\nu) = \frac{1}{2\pi} + \frac{\kappa}{2\pi}\cos(\psi), \tag{3.16}$$

which is a *cosine* distribution (see p.44, Fisher (1993)) with mean angle $\mu = 0$. It is also called the cardioid distribution since its shape looks like a heart.

Method of simulation:

Acceptance-rejection (A.R.) sampling can be employed to generate r.v.'s from the Batschelet's model. The A.R. algorithm for continuous r.v.'s Y from distribution F, with p.d.f. $f(y)$ for which one wants to sample from the (3.14).

1. Generate a r.v. X from distribution G, with p.d.f. $g(x)$ for which we already have an efficient algorithm for generating from.

2. Determine $c = sup_x\{\frac{f(x)}{g(x)}\}$

3. Generate U from $U(0,1)$

4. If $U \leq \frac{f(X)}{cg(X)}$, then set $Y = X$ (accept); otherwise go back to 1 (reject).

In this thesis, $g(x)$ is taken to be the $U(0,2\pi)$ distribution and this is a very inefficient envelope function, which gives a low acceptance rate. A more efficient $g(x)$ should be considered in future work.

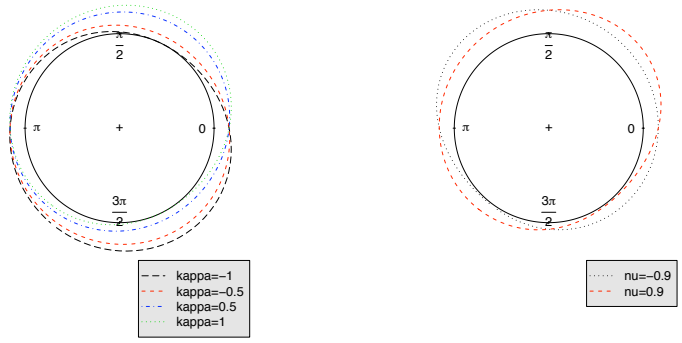


Figure 3.6: Circular plots of Batschelet's distributions: (a) $\nu = 0.5$, various values of κ, and (b) $\kappa = 0.9$, various values of ν.

Figure 3.6 shows two circular plots of Batschelet's distributions for different values of κ and different values of ν. Plot (a) shows the effect of the parameter κ; the plot (b) shows the effect of the parameter ν.

3.5.2 Wrapped exponential (WE)

The principle of wrapping distributions on the real line around the unit circle can be used to obtain skewed circular distributions. The shifted exponential distribution on the real line has two parameters with p.d.f.

$$f(x;\lambda,\mu_0) = \lambda e^{-\lambda(x-\mu_0)},\ x > \mu_0. \tag{3.17}$$

The char. f. of (3.17) is

$$\phi(t) = \frac{e^{i\mu_0 t}}{1 - it/\lambda}.$$

By following (3.6) and wrapping (3.17) around the unit circle yields the p.d.f. of the WE distribution

$$f(\theta; \lambda, \mu_0) = \sum_{k=-\infty}^{\infty} \lambda e^{-\lambda(\theta+2k\pi-\mu_0)} = \sum_{k=0}^{\infty} \lambda e^{-\lambda(\theta+2k\pi-\mu_0)} = \frac{\lambda e^{-\lambda(\theta-\mu_0)}}{1 - e^{-2\pi\lambda}}. \tag{3.18}$$

The second equality is due to the fact that the exponential distribution only takes positive values. The c.d.f. is obtained by integrating (3.18),

$$F(\theta; \lambda, \mu_0) = \int_0^{\theta} \frac{\lambda e^{-\lambda(\omega-\mu_0)}}{1 - e^{-2\pi\lambda}} d\omega = \frac{e^{\lambda\mu_0}(1 - e^{-\lambda\theta})}{1 - e^{-2\pi\lambda}}.$$

From Proposition 3.4.1 it follows that the char. f. of the WE distribution is

$$\phi(p) = \frac{e^{i\mu_0 p}}{1 - ip/\lambda}, \; p = 0, \pm 1, \pm 2, ...$$

So that

$$\alpha_p = \frac{\lambda\{\lambda\cos(\mu_0 p) - p\sin(\mu_0 p)\}}{\lambda^2 + p^2},$$

and

$$\beta_p = \frac{\lambda\{\lambda\sin(\mu_0 p) + p\cos(\mu_0 p)\}}{\lambda^2 + p^2}.$$

After substituting α_p and β_p into (3.8), and using the difference formula for the *cosine* function and the difference formula for the *sine* function, an alternative form of the

p.d.f. of the WE distribution is

$$f(\theta;\lambda,\mu_0) = \frac{1}{2\pi}\left[1 + 2\sum_{p=1}^{\infty}\frac{\lambda}{\lambda^2+p^2}\{\lambda\cos p(\mu_0-\theta) - p\sin p(\mu_0-\theta)\}\right].$$

Very often it is reasonable to take $\mu_0 = 0$. This special case was studied by Jammalamadaka and Kozubowski (2003). The WE distribution degenerates to the uniform distribution when $\lambda = 0$, and the WE distribution has a very sharp edge when θ is close to μ when $\lambda > 0$, see Figure 3.7; therefore, it is of limited use in practice. However, it is used later to demonstrate the procedure of obtaining an asymptotic distribution of the U^2 statistic for continuous data and to show that the asymptotic distribution of the U_d^2 for grouped data approximates U^2 for continuous data well, particularly at the upper tail. Figure 3.7 is a circular plot of WE distributions for

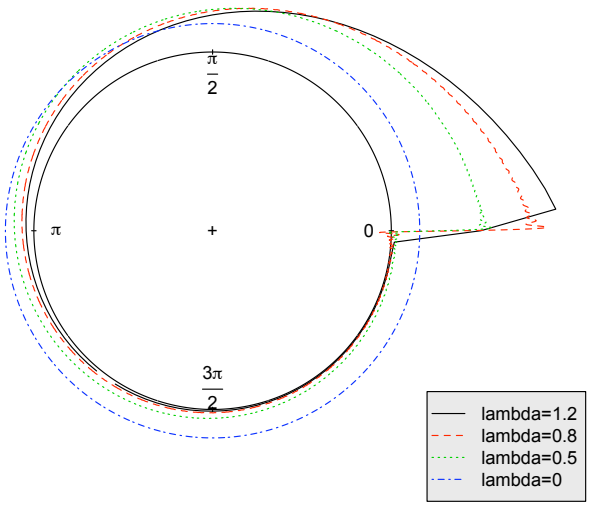


Figure 3.7: A circular plot of WE distributions for different values of λ

different values of λ. It shows that as λ increases, the height of the edge increases at

$\mu = 0$.

Method of simulation:

WE pseudo r.v.'s Θ can be obtained by taking module 2π of the corresponding exponential r.v.'s X on the real line ($X \bmod 2\pi$) .

To model skewed data, usually, distributions with more than 2 parameters are needed. A few more advanced wrapped skewed distributions are given next.

3.5.3 Wrapped skew-Laplace (WSL)

The skewed Laplace distribution was introduced by Hinkley and Revankar (1977). It has found applications in many fields (i.e., it is the one of the most common distributions used to describe the logarithm of particle sizes (Fieller et. al., 1992), it has been used to analyze bacterial sizes in axenic cultures (Vives-Rego, 2005), and it is also used in mathematical finance (Madan et. al., 1998 and Kotz et. al., 2001)). The Anderson Darling and Cramér-von Mises goodness-of-fit statistics are introduced by Puig and Stephens (2007) for the skew-Laplace distribution on the real line. The distribution is derived by taking the difference of two exponential distributions with unequal shape parameters. The distribution has the following density

$$f(x;\lambda_1,\lambda_2,\mu_0) = \begin{cases} \frac{\lambda_1\lambda_2}{\lambda_1+\lambda_2} e^{\lambda_1(x-\mu_0)}, & x \le \mu_0 \\ \frac{\lambda_1\lambda_2}{\lambda_1+\lambda_2} e^{\lambda_2(\mu_0-x)}, & x > \mu_0. \end{cases} \tag{3.19}$$

The distribution is asymmetric around the sharp peak at the mode. Puig and Stephens (2007) discussed the MLEs of all three parameters. Following (3.6), the p.d.f. of the WSL distribution is

$$f(\theta;\lambda_1,\lambda_2,\mu_0) = \frac{\lambda_1\lambda_2}{\lambda_1+\lambda_2}\left[\frac{e^{-\lambda_1(\theta-\mu_0)}}{1-e^{-2\pi\lambda_1}} + \frac{e^{-\lambda_2\{2\pi-(\theta-\mu_0)\}}}{1-e^{-2\pi\lambda_2}}\right]. \tag{3.20}$$

By integrating (3.20), the c.d.f. of the WSL distribution is

$$F(\theta, \lambda_1, \lambda_2, \mu_0) = \frac{\lambda_1}{\lambda_1 + \lambda_2} \frac{1 - e^{-\lambda_2(\theta - \mu_0)}}{1 - e^{-2\pi\lambda_2}} + \frac{\lambda_2}{\lambda_1 + \lambda_2} \frac{\{e^{-\lambda_1(\theta - \mu_0)} - 1\} e^{-2\pi\lambda_1}}{1 - e^{-2\pi\lambda_1}}. \tag{3.21}$$

The char. f. of (3.19) is

$$\phi(t) = \frac{e^{i\mu_0 t}}{(1 + it/\lambda_1)(1 - it/\lambda_2)}.$$

It then follows from Proposition 3.4.1 that the char. f. of $\Theta = X(\mathrm{mod} 2\pi)$ is

$$\phi(p) = \frac{e^{i\mu_0 p}}{(1 + ip/\lambda_1)(1 - ip/\lambda_2)}, \; p = 0, \pm 1, \pm 2, ...$$

So that

$$\alpha_p = \frac{\cos(\mu_0 p)\{1 + p^2/(\lambda_1\lambda_2)\} + \sin(\mu_0 p) p(1/\lambda_2 - 1/\lambda_1)}{(1 + p^2/\lambda_1^2)(1 + p^2/\lambda_2^2)},$$

and

$$\beta_p = \frac{\sin(\mu_0 p)\{1 + p^2/(\lambda_1\lambda_2)\} - \cos(\mu_0 p) p(1/\lambda_2 - 1/\lambda_1)}{(1 + p^2/\lambda_1^2)(1 + p^2/\lambda_2^2)}.$$

After substituting α_p and β_p into (3.8), and using the difference formula for the *cosine* function and the difference formula for the *sine* function, an alternative form of the p.d.f. is

$$f(\theta; \lambda_1, \lambda_2, \mu_0) =$$
$$\frac{1}{2\pi}\left[1 + 2\sum_{p=1}^{\infty} \frac{1}{(1 + p^2/\lambda_1^2)(1 + p^2/\lambda_2^2)} \left\{\left(1 + \frac{p^2}{\lambda_1\lambda_2}\right)\cos p(\theta - \mu_0) - p\left(\frac{1}{\lambda_1} - \frac{1}{\lambda_2}\right)\sin p(\theta - \mu_0)\right\}\right].$$

Very often it is reasonable to take $\mu_0 = 0$. This special case was studied by Jammalamadaka and Kozubowski (2004), who use a different parameterization ($\lambda = \sqrt{\lambda_1\lambda_2}, \kappa = \sqrt{\frac{\lambda_1}{\lambda_2}}$) and who also discuss estimation by the method of moments.

When $\lambda_1 = \lambda_2$, the distribution corresponds to a symmetric wrapped Laplace distribution, symmetric around μ. When $\lambda_1 < \lambda_2$, the distribution is skewed to the left; when $\lambda_1 > \lambda_2$, the distribution is skewed to the right. It can be shown using l'Hópital's rule that as $\lambda_1 \to 0$, $f(\theta) \to \frac{1}{2\pi}$, the uniform distribution. Similarly, as $\lambda_2 \to 0$, $f(\theta) \to \frac{1}{2\pi}$ (see Figure 3.8).

Method of simulation:

Using the result that a skew-Laplace r.v. X is a difference between two independent exponential pseudo r.v.'s: $E_1 \backsim \exp(\lambda_1)$ and $E_2 \backsim \exp(\lambda_2)$ ($X \stackrel{d}{=} E_1 - E_2$). The WSL pseudo r.v.'s can be generated by first generating two independent exponential r.v.'s (command "rexp" in *R*) E_1 and E_2 and then using $\Theta = X(\mathrm{mod}\ 2\pi)$.

Figure 3.8 is a circular plot of WSL distributions for different values of λ_1 and different values of λ_2. It shows the effect of the two parameters. The difference of λ_1 and λ_2 determines the skewness of the density. With increase in λ_1 and λ_2, the density becomes higher peaked.

3.5.4 Wrapped stable (WS)

The family of stable distributions defined on the real line was introduced by Lévy (1924). The normal, Cauchy, and Lévy distributions are special cases in the stable family, and these are the only cases known to have closed form in the stable family. The wrapped α stable distribution is capable of modeling both symmetric and skewed unimodal data around the unit circle. It is obtained by wrapping a four parameter stable distribution around the unit circle, and there are many different parametriza-

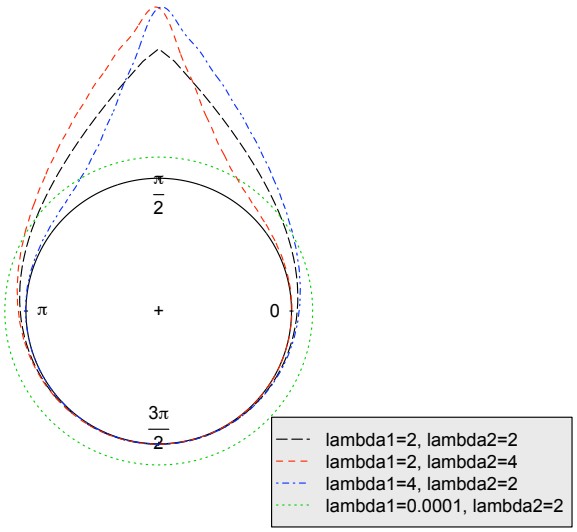


Figure 3.8: Circular plot of the WSL distributions for $\mu_0 = \frac{\pi}{2}$, and different values of λ_1 and different values of λ_2.

tions for its char. f.. A detailed discussion of WS distributions was given by Pewsey (2008). By using the parametrization recommended by Nolan (1998), the char. f. on the real line is

$$\phi_x(t) = \begin{cases} \exp\{-\gamma^\alpha |t|^\alpha [1 + i\beta \text{sign(t)} \tan(\frac{\alpha\pi}{2})\{(\gamma|t|)^{1-\alpha} - 1\}] + i\mu t\}, & \alpha \neq 1 \\ \exp[-\gamma|t|\{1 + i\beta \text{sign(t)} \frac{2}{\pi} \log(\gamma|t|)\} + i\mu t], & \alpha = 1 \end{cases}$$

and by following (3.6), the char. f. of the WS is

$$\phi(p) = \begin{cases} \exp\{-\gamma^\alpha p^\alpha [1 + i\beta \tan\frac{\alpha\pi}{2}\{(\gamma p)^{1-\alpha} - 1\}] + i\mu p\}, & \alpha \neq 1 \\ \exp[-\gamma\{1 + i\beta p \frac{2}{\pi} \log(\gamma p)\} + i\mu p], & \alpha = 1. \end{cases}$$

It follows that

$$\alpha_p = \begin{cases} \exp(-\gamma^\alpha p^\alpha)\cos p\{\mu + \beta\tan(\frac{\alpha\pi}{2}) * (\gamma^\alpha p^{\alpha-1} - \gamma)\}, & \alpha \neq 1 \\ \exp(-\gamma p)\cos p\{\mu - \gamma\beta\frac{2}{\pi}\log(\gamma p)\}, & \alpha = 1 \end{cases}$$

and

$$\beta_p = \begin{cases} \exp(-\gamma^\alpha p^\alpha)\sin p\{\mu + \beta\tan(\frac{\alpha\pi}{2}) * (\gamma^\alpha p^{\alpha-1} - \gamma)\}, & \alpha \neq 1 \\ \exp(-\gamma p)\sin p\{\mu - \gamma\beta\frac{2}{\pi}\log(\gamma p)\}, & \alpha = 1. \end{cases}$$

After substituting α_p and β_p into (3.8), and using the difference formula for *cosine*, an alternative form of the density is

$$f(\theta;\mu,\alpha,\beta,\gamma) = \frac{1}{2\pi} + \frac{1}{\pi}\sum_{p=1}^{\infty}\exp(-\gamma^\alpha p^\alpha)\begin{cases} \cos p\{\mu + \beta\tan(\frac{\alpha\pi}{2}) * (\gamma^\alpha p^{\alpha-1} - \gamma) - \theta\}, & \alpha \neq 1 \\ \cos p[\{\mu - \gamma\beta\frac{2}{\pi}\log(\gamma p)\} - \theta], & \alpha = 1. \end{cases} \tag{3.22}$$

In the density above, α controls the peakedness around the mode (see Figure 3.9), β determines the skewness (see Figure 3.10), and γ is the scale parameter (see Figure 3.11). The distribution is symmetric about μ when $\beta = 0$ and it is skewed to the right (left) when $0 < \beta < 1$ $(-1 < \beta < 0)$. The distribution degenerates to a wrapped Cauchy when $\alpha = 1$ and $\beta = 0$ and to a wrapped Normal when $\alpha = 2$ and $\beta = 0$. One can use the method of moments discussed in Pewsey (2008) to obtain initial values of the MLE's, or use the fact that the WC is a special case of the WS, i.e. let the ML estimates $\hat{\mu}$ and $\hat{\rho}$ from the WC fit be the initial values of μ and γ, and identify a suitable region of α and β as their starting values on the contour plot of the profile log-likelihood.

Figure 3.9 shows a circular plot of WS distributions for different values of α. With an increase in α, the density has a higher mode.

Method of simulation:

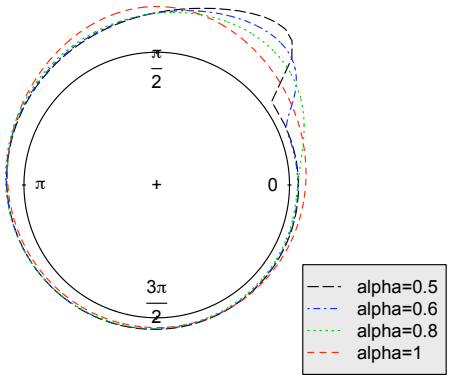


Figure 3.9: A circular plot of the WS distributions for $\beta = 1$, $\gamma = 1$, $\mu = \frac{\pi}{2}$, and different values of α.

The stable r.v.'s on the line can be simulated by the algorithm developed by Chambers et al (1976). In this thesis, the stable r.v.'s X on the line were generated from a stable distribution with parameters index (α), skewness (β) and scale = 1 by the command "rstable" in R package "circular". The more general shifted and scaled stable r.v.'s Y on the line can be obtained by taking $Y = \gamma X + \mu$ and then the WS r.v.'s Θ were obtained by using $\Theta = Y \pmod{2\pi}$.

Figure 3.10 shows a circular plot of the WS distributions for different values of β. Smaller negative values of β result in the distribution being skewed to the left; whereas bigger positive values of β result in the distribution being skewed to the right.

Figure 3.11 shows a circular plot of WS distributions for different values of γ. With an decrease in γ, the density concentrates around $\mu = \frac{\pi}{2}$.

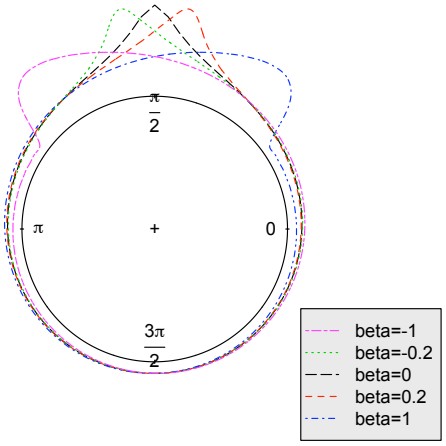


Figure 3.10: A circular plot of the WS distributions for $\alpha = 0.5$, $\gamma = 1$, $\mu = \frac{\pi}{2}$ and different values of β.

3.5.5 Wrapped skewed normal (WSN)

The family of skewed normal (SN) distributions is an extension of the normal family obtained by adding a shape parameter which regulates skewness. The skewed normal distribution on the real line was introduced by Azzalini (1985). Let X be skewed normal r.v.'s with skewness parameter λ. Then the standard skewed normal density with shape parameter λ, $-\infty < \lambda < \infty$ is

$$f(x;\lambda) = 2\phi(x)\Phi(\lambda x),\ -\infty < x < \infty,$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and distribution function. For the more general case with location parameter ξ and scale parameter η, let $Y = \xi + \eta X$

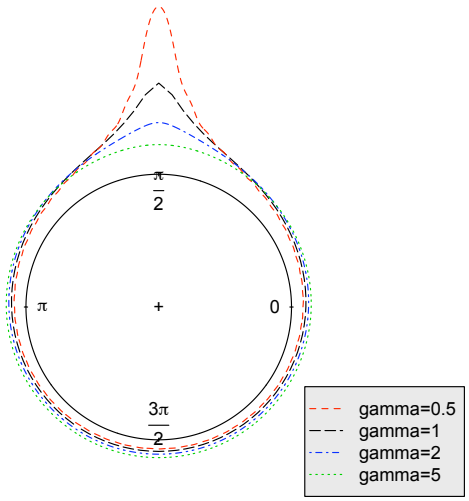


Figure 3.11: A circular plot of the WS distributions for $\alpha = 0.5$, $\beta = 1$, $\mu = \frac{\pi}{2}$ and different values of γ.

and the r.v. Y has the p.d.f.

$$f(y;\lambda,\xi,\eta) = \frac{2}{\eta}\phi(\frac{y-\xi}{\eta})\Phi\{\lambda(\frac{y-\xi}{\eta})\},$$

where $-\infty < y < \infty, -\infty < \xi < \infty, \eta > 0, -\infty < \lambda < \infty$. The char. f. of Y was given by Azzalini and Capitanio (1999). Pewsey studied the WSN distributions (2000, 2006), and established the char. f. for these distributions as

$$\phi_p = E(e^{ip\Theta}) = \exp(ip\xi - \frac{1}{2}\eta^2 p^2)\{1 + iQ(\delta\eta p)\},\ p = 0, 1, 2, \ldots,$$

where $Q(x) = \int_0^x \sqrt{2/\pi} e^{u^2/2} \mathrm{d}u$ and $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}} \in (-1, 1)$. It follows that

$$\alpha_p = \exp(-\frac{1}{2}\eta^2 p^2)\{\cos(p\xi) - Q(\delta\eta p)\sin(p\xi)\},$$

and

$$\beta_p = \exp(-\frac{1}{2}\eta^2 p^2)\{\sin(p\xi) + Q(\delta\eta p)\sin(p\xi)\}.$$

Again, after substituting α_p and β_p into (3.8) and using the difference formula for the *cosine* function, the density can be shown to be

$$f(\theta;\xi,\eta,\lambda) = \frac{1}{2\pi}\left[1 + 2\sum_{p=1}^{\infty} \exp(-\frac{1}{2}\eta^2 p^2)\{\cos\left(p(\theta-\xi)\right) + Q(\delta\eta p)\sin\left(p(\theta-\xi)\right)\}\right].$$

As η approaches 0, the distribution tends to a point distribution, as η approaches ∞, the distribution degenerates into the uniform distribution, and when $\lambda = 0$, the distribution becomes a WN distribution.

Method of simulation:

The skew-normal pseudo r.v.'s on the line can be simulated using the algorithm which was developed by Henze (1986). In this thesis, the SN pseudo r.v.'s X with location parameter δ, scale parameter η and shape parameter λ were generated using the command "rsn" in *R* package "sn". Having generated the SN pseudo r.v.'s X, the WSN pseudo r.v.'s were obtained as $\Theta = X \pmod{2\pi}$.

Figure 3.12 shows circular plots of WSN distributions for different values of λ and different values of η. The plot (left) shows the effect of the parameter λ. As λ increases, the mode increases. The plot (right) shows the effect of the parameter η. As η increases, the density concentrates more and more at the mode until it becomes a point distribution when $\eta = 0$.

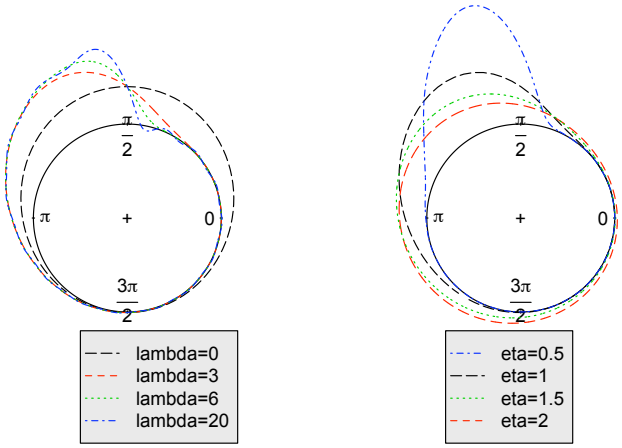


Figure 3.12: Circular plots of WSN distributions: (left) $\eta = 1$, $\xi = 90°$, and different values of λ and (right) $\lambda = 3$, $\xi = 90°$, and different values of η.

3.5.6 Wrapped normal Laplace (WNL)

The normal Laplace (NL) distribution was introduced by Reed and Jorgensen (2004). It is a special case of the generalized normal Laplace (GNL) (Reed, 2004). Reed and Jorgensen used the NL to model the distribution of incomes, particle sizes, oil-field sizes, and city sizes. The distribution arises as a convolution of Gaussian and skew-Laplace components and it occurs when the state of a Brownian motion (with the starting state normally distributed) is observed after an exponentially distributed time (Reed and Jorgensen 2004). Let Z be the normal pseudo r.v.'s with mean μ and variance σ^2, and W be the skew-Laplace pseudo r.v.'s (see Section 3.5.3) with the mode at 0, and with the shape parameters $a = \frac{1}{\lambda_1}$ and $b = \frac{1}{\lambda_2}$. Then the NL pseudo r.v.'s X are defined as

$$X \stackrel{d}{=} Z + W.$$

The char. f. of a NL is the product of the char. f. of a Normal and the char. f. of a skewed Laplace

$$\phi_t = \frac{\exp(i\eta t - \tau^2 t^2/2)}{(1 - iat)(1 + ibt)}.$$

In general, tails of NL distribution are fatter than normal for all $a, b > 0$. As $a \to \infty$, the lower tail is exponential and fatter than the normal distribution, while the upper tail is normal. As $b \to \infty$, the upper tail is exponential and the lower tail is normal. When $a = b$, the distribution is symmetric. Some special cases are

1. when $\tau = 0$, NL is a skewed Laplace distribution, WSL($\lambda_1 = \frac{1}{a}, \lambda_2 = \frac{1}{b}, \mu_0 = \eta$);
2. when $a = b = 0$, it is a normal distribution.

Following Proposition 3.4.1, the char. f. of WNL is (see Reed and Pewsey, 2008)

$$\phi_p = \frac{\exp(i\eta p - \tau^2 p^2/2)}{(1 - iap)(1 + ibp)}, p = 0, \pm 1, \pm 2, \ldots,$$

and

$$\alpha_p = \frac{e^{-\tau^2 p^2/2}}{(1 + a^2 p^2)(1 + b^2 p^2)}\{(1 + abp^2)\cos(\eta p) + (b - a)p\sin(\eta p)\} \tag{3.23}$$

and

$$\beta_p = \frac{e^{-\tau^2 p^2/2}}{(1 + a^2 p^2)(1 + b^2 p^2)}\{(1 + abp^2)\sin(\eta p) - (b - a)p\cos(\eta p)\}. \tag{3.24}$$

Again, after substituting α_p and β_p into (3.8), and using the difference formula for the *cosine* function and the difference formula for the sine functions, the p.d.f. is

$$f(\theta; a, b, \tau, \eta) = \frac{1}{2\pi}\left[1 + 2\sum_{p=1}^{\infty} \frac{\exp(\frac{-\tau^2 p^2}{2})}{(1 + a^2 p^2)(1 + b^2 p^2)}\{(1 + abp^2)\cos(p(\eta - \theta)) + (b - a)p\sin(p(\eta - \theta))\}\right].$$

It follows that the pth mean resultant length of the WNL is

$$\rho_p = \frac{e^{-\tau^2 p^2/2}}{(1 + a^2 p^2)(1 + b^2 p^2)}, \tag{3.25}$$

and the mean direction is

$$\mu = \left\{\eta + \tan^{-1}\left(\frac{a-b}{1+ab}\right)\right\}(mod\, 2\pi). \tag{3.26}$$

Numerical estimation startup values of the 4-parameter WNL can be obtained using the sample mean direction $\bar{\theta}$, and sample mean resultant length $\bar{R}$ for μ and ρ, and arbitrary values for a and b (Reed and Pewsey, 2008), then using (3.25) and (3.26) to solve for the starting values of τ and η.

Analogous to the NL distribution on the real line, the WNL becomes a wrapped normal (WN) when $a = b = 0$, $(\text{WNL}(\eta, \tau^2, 0, 0) \Leftrightarrow \text{WN}(\eta, \tau^2))$; and it becomes a wrapped (skew) Laplace when $\tau = 0$. Note that a and b are the shape parameters for the Laplace component of WNL. When $\tau^2 = 0$, WNL becomes a wrapped skewed-Laplace(WSL) with $\lambda_1 = \frac{1}{a}$ and $\lambda_2 = \frac{1}{b}$.

Method of simulation:

Generating the WNL r.v.'s will be discussed in the next section as a special case of generating the wrapped generalized normal Laplace random variables.

Figure 3.13 shows circular plots of WNL distributions for different values of τ and a. The plot (left) shows the effect of the parameter τ. With $\tau = 0$, the density is a wrapped laplace distribution. With an increase in τ, the flatness increases. The plot (right) shows the effect of the parameter a. With a decrease in a, the skewness increases.

3.5.7 Wrapped Generalized normal Laplace (WGNL)

The generalized normal Laplace (GNL) distribution (Reed, 2007) is obtained by taking the convolution of an independent normal distribution and an independent gener-

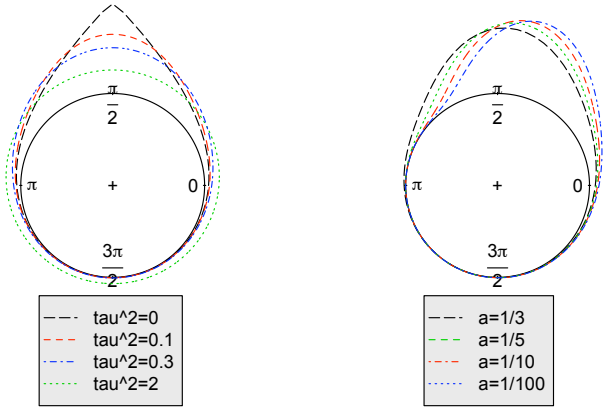


Figure 3.13: Circular plots of the WNL distributions: (left) $a = 0.5$, $b = 0.5$, and different values of τ^2, and (right) $b = 0.5$, $\tau^2 = 0.1$, and different values of a.

alized Laplace distribution (Kotz et al. 2001). The generalized Laplace distribution is obtained by taking the difference between two independent Gamma r.v.'s with the same shape parameter. Following Proposition 3.4.1, the char. f. of WGNL is (see Reed and Pewsey, 2008)

$$\phi_p = \phi_{WGNL}(p) = \{\phi_{WNL}(p)\}^{\zeta} = \left\{\frac{\exp(i\eta p - \tau^2 p^2/2)}{(1-iap)(1+ibp)}\right\}^{\zeta}, p = 0, \pm 1, \pm 2, \ldots,$$

$$\alpha_p = \left\{\frac{e^{-\tau^2 p^2}}{(1+a^2p^2)(1+b^2p^2)}\right\}^{\zeta/2} \cos[\zeta\{\eta p + w(a,b)\}]$$

and

$$\beta_p = \left\{\frac{e^{-\tau^2 p^2}}{(1+a^2p^2)(1+b^2p^2)}\right\}^{\zeta/2} \sin[\zeta\{\eta p + w(a,b)\}]$$

where $w(a,b) = \arctan\{\frac{(a-b)p}{1+abp^2}\}$. It follows that the pth mean resultant length of the WGNL is

$$\rho_p = \left\{\frac{e^{-\tau^2 p^2}}{(1+a^2p^2)(1+b^2p^2)}\right\}^{\zeta/2}, \tag{3.27}$$

and the mean direction is

$$\mu = [\zeta \{\eta + w(a,b)\}](mod\, 2\pi). \tag{3.28}$$

Reed and Pewsey (2008) presented stochastic models for the genesis of these distributions, which might be used to identify situations that are appropriate for these distributions. Both families contain the wrapped normal and the wrapped Laplace distributions as special cases. When $a = b = 0$, and $\zeta = 1$, WGNL becomes a normal distribution; whereas when $\tau = 0$, it becomes a generalized Laplace distribution. The WGNL has WNL, WSL, WN, and uniform as nested models. In Chapter 7, likelihood ratio tests are used to compare the fit of nested models. For numerical estimation, startup values of the 5-parameter WGNL can be obtained using the MLE's of the 4-parameter WNL and letting $\zeta = 1$.

Method of simulation:

The GNL r.v.'s X can be represented as $X \stackrel{d}{=} \eta\zeta + \tau\sqrt{\zeta}Z + aV_1 - bV_2$, where Z, V_1, and V_2 are independent, Z follows the standard normal distribution, and V_1 and V_2 both follow the Gamma distribution with shape parameter ζ and scale parameter 1 (Reed 2007). This can be used to generate pseudo r.v.'s following the GNL distribution. Then using these values modulo 2π yields the WGNL pseudo r.v.'s.

Figure 3.14 shows densities of WNL and WGNL distributions for various parameter values. In the middle row $\xi = 1$ (WNL), in the top row $\xi = 0.7$ and in the bottom row $\xi = 1.5$. In comparing the three rows note that the smaller value of ζ leads to taller peaks and thiner flanks, and also a move to the left of the mode. The left-hand panels show the effect of changing the parameter τ^2, with other parameters kept constant ($a = b = 0.5$ and $\eta = \pi$). For the three curves (moving downwards) $\tau^2 = 0, 0.25, 1$, respectively. The case $\tau^2 = 0$ corresponds to a wrapped laplace dis-

tribution (middle row) and wrapped generalized laplace distribution. Note that the bigger value of τ^2 leads to lower peaks and fatter tails. The right-hand panels show the effect of increasing a while fixing $b = 0$. The three curves (moving downwards) correspond to $a = 1, 3, 10$ (with $\eta = \pi$, $\tau^2 = 0.1$). Note that the skewness and flatness of the distributions increase with the increase with a.

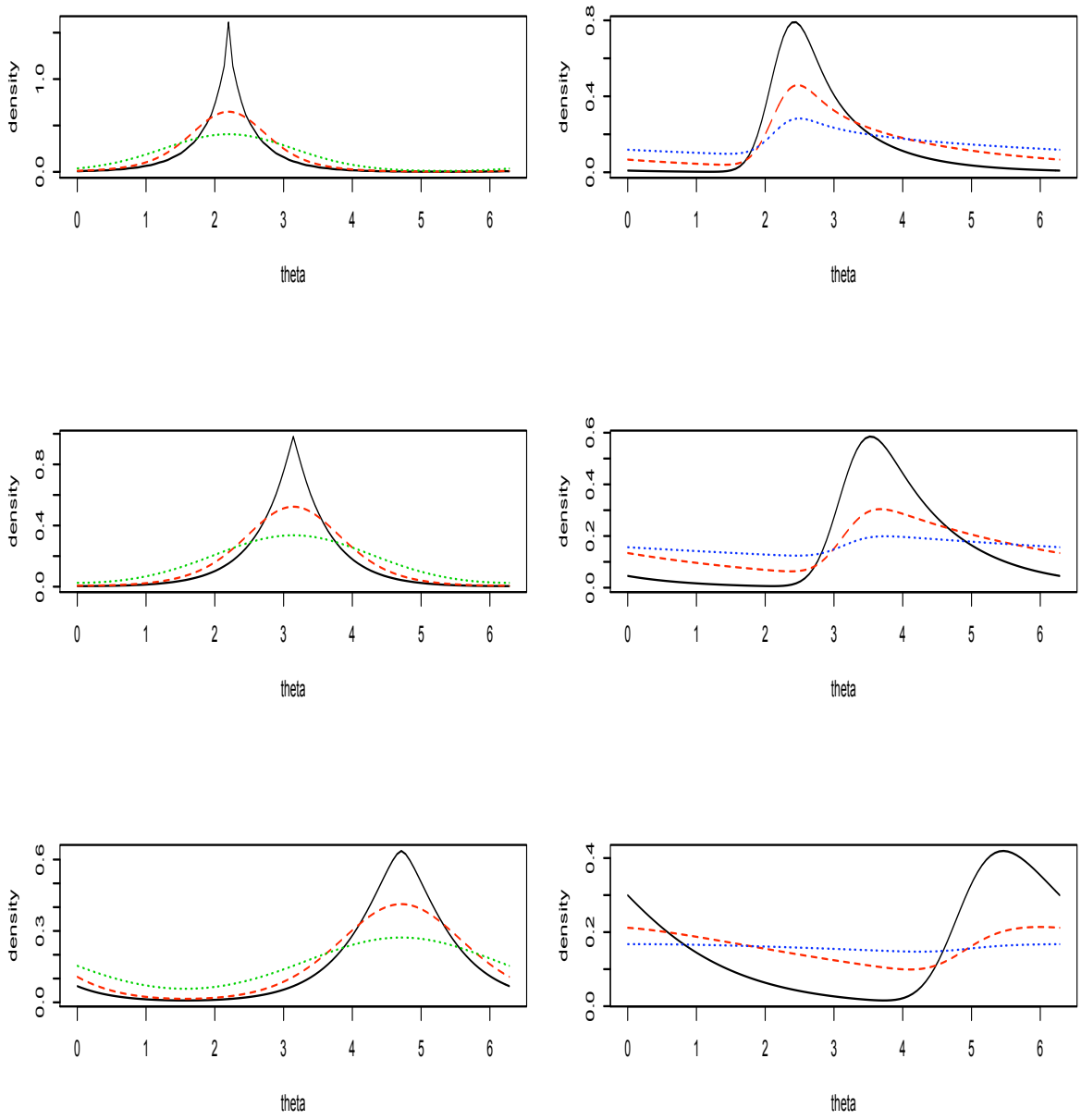


Figure 3.14: Plots of WGNL distributions: top row $\xi = 0.7$, middle row $\xi = 1$ (WNL), and bottom row $\xi = 1.5$. The left-hand panels show the three curves (moving downwards) $\tau^2 = 0, 0.25, 1$, with other parameters kept constant ($a = b = 0.5$ and $\eta = \pi$). The right-hand panels show the three curves (moving downwards) correspond to $a = 1, 3, 10$, with $b = 0$, $\eta = \pi$, and $\tau^2 = 0.1$.

Chapter 4

Maximum likelihood estimation for grouped and continuous data

The general theory and properties of ML Estimates have been discussed in many statistical text books. A good introduction to the sufficiency and consistency and other asymptotic properties (efficiency and asymptotic normality) was given by Stuart, Ord, and Arnold (1993). They also expanded the discussion into several parameters. In general, similar properties hold for several parameters as they do for single parameter case. A more in-depth discussion of the asymptotic properties of ML estimates for the multinomial distribution was given by Rao (1957, 1958).

4.1 Parameter estimation for grouped data

Suppose the n θ-values are grouped into k cells with boundaries $b_0, b_1, \cdots, b_k$ where $b_k = b_0$ since the cells are around a circle. Let $o_i, i = 1, 2, \ldots, k$ be the observed frequency in cell i. Suppose p_i is the probability of an observation falling into cell i with interval (b_{i-1}, b_i). The maximum likelihood estimators for grouped data can be obtained as follows: for any distribution $f(\theta; \vec{\psi})$, where $\vec{\psi}$ is a vector containing q

unknown parameters $\psi_1, \psi_2, \cdots, \psi_q$, and $p_i,\ i = 1, 2, \cdots, k$ is

$$p_i = \int_{b_{i-1}}^{b_i} f(\theta; \vec{\psi}) \mathrm{d}\theta, \tag{4.1}$$

the log-likelihood is (omitting constants)

$$l \propto \sum_{i=1}^{k} o_i \log p_i, \tag{4.2}$$

and this must be maximized with respect to the parameters (contained in p_i). One way to maximize (4.2) is by solving the equations

$$\frac{\partial l}{\partial \psi_j} = \sum_{i=1}^{k} \frac{o_i}{p_i} \frac{\partial p_i}{\partial \psi_j} = 0, \tag{4.3}$$

where $j = 1, 2, \cdots, q$, but it is usually easier to maximize (4.2) using a numerical maximization routine. One can use, for example, the simplex algorithm of Nelder and Mead (1965), or the quasi-Newton method (Broyden, Fletcher, Goldfarb and Shanno, 1970). Good starting values for ψ_j may be found as follows:

1. set the o_i unknown pseudo θ-values in cell i all equal to the midpoint value of the cell;

2. treat the whole n observations as continuous observations;

3. estimate the unknown parameters based on continuous MLE, as described in the following section, or based on the method of moments. Note however, neither method guarantees good starting values. It may be helpful to try different sets of starting values.

It can be shown that as $k \to \infty$, maximizing the grouped log-likelihood (l_g) results in the same MLEs as maximizing the continuous log-likelihood (l_c), see proof below:

Proof. Following (4.2)

$$\lim_{k\to\infty} l_g = \lim_{k\to\infty} \sum_{i=1}^{k} o_i \log p_i = \lim_{k\to\infty} \sum_{i=1}^{k} o_i \log\{\int_{b_{i-1}}^{b_i} f(\theta)d\theta\}.$$

As $k \to \infty$, there is a maximum of 1 observation falling into a cell and therefore $o_i = 0$ or 1, it then follows

$$\lim_{k\to\infty} l_g = \sum_{i=1}^{n} \log\{f(\theta_i)\Delta\theta_i\} = l_c + constant$$

since the $\log(\Delta\theta_i)$ does not contain any parameters. □

The asymptotic variance-covariance matrix of the estimates can be found, see section 6.2.

4.2 Parameter Estimation for continuous data

In many data sets, the observations (θs) are regarded as continuous, although they are grouped to the order of accuracy of the measuring device or method (e.g., the nearest degree in directional data, or the nearest day for the incidence of disease). In such data sets, suppose $\theta_1, \theta_2, \cdots, \theta_n$ are continuous independent observations then the log-likelihood is of the form

$$l = \sum_{i=1}^{n} \log f(\theta_i). \tag{4.4}$$

This is maximized over parameters to obtain maximum likelihood (ML) estimates. Details of ML estimation for each of the distributions in chapter 3 are given below:

1. *vonMises distribution*

   If the underlying distribution is von Mises, the MLE of μ is the angle of the resultant vector $\bar{\theta}$ as described in (2.1), and the estimate of the shape κ is given

by solving:

$$A(x) = I_1(\kappa)/I_0(\kappa) = \bar{R}, \tag{4.5}$$

where $\bar{R}$ is the length of the resultant vector defined in (2.2). With modern computers and statistical software, the inverse $A^{-1}(x)$ can to be found numerically, for example, using command "A1inv" in R package "circular". $I_1(\kappa)$ is a modified Bessel function of order one (the derivative of $I_0(\kappa)$). Abramowitz and Stegun (1970, p.378) gave polynomial approximations to $I_0(\kappa)$ and $I_1(\kappa)$. Best and Fisher (1981) gave a simple approximation to the MLE of κ.

$$\hat{\kappa} = \begin{cases} 2\bar{R} + \bar{R}^3 + 5\bar{R}^5/6, & \bar{R} < 0.53 \\ -0.4 + 1.39\bar{R} + 0.43/(1-\bar{R}), & 0.53 \le \bar{R} < 0.85 \\ 1/(\bar{R}^3 - 4\bar{R}^2 + 3\bar{R}), & \bar{R} \ge 0.85. \end{cases}$$

However, the estimate of κ is not reliable when $\bar{R}$ is small, say $\bar{R} < 0.7$.

2. *Circular beta distribution*

The MLE of α, β and η can be found by maximizing the log-likelihood function (4.4) using a numerical maximization routine, where $f(\theta)$ is from (3.5); or differentiating the log-likelihood function with respect to α, β and η and set the differentiated functions to 0. When the modal vector is set at 0 degree ($\eta = 0$), differentiating with respect to α and β (and set to 0) results:

$$\Psi(\alpha) - \Psi(\alpha+\beta) = \frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{\cos(\theta_i)+1}{2}\right)$$

and

$$\Psi(\beta) - \Psi(\alpha+\beta) = \frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{1-\cos(\theta_i)}{2}\right),$$

where $\Psi(x) = \dfrac{\partial \log \Gamma(x)}{\partial x}$ is the digamma function. The above equations need to be solved numerically; starting values are recommended in Lai (p.44, 1994).

3. *Batschelet's distribution*

   The MLEs of κ and ν can be found numerically by solving the equations:

$$\sum_{i=1}^{n} \frac{\sin(\theta_i + \nu \sin(\theta_i))}{1 + \kappa \sin(\theta_i + \nu \sin(\theta_i))} = 0$$

   and

$$\sum_{i=1}^{n} \frac{\kappa \cos(\theta_i + \nu \sin(\theta_i)) \sin(\theta_i)}{1 + \kappa \sin(\theta_i + \nu \sin(\theta_i))} = 0.$$

4. *Other wrapped distributions (WE, WSL, WN, WC, WS, WSN, WNL, WGNL)*

   In general, using the representation of the density in (3.8) in section 3.4, the log-likelihood of $\vec{\psi}$ is

$$l(\vec{\psi}, \theta) = -n \log(2\pi) + \sum_{i=1}^{n} \log \left\{ 1 + 2 \sum_{p=1}^{M} (\alpha_p \cos(p\theta) + \beta_p \sin(p\theta)) \right\}, \qquad (4.6)$$

   where M is the number of terms required to approximate the infinite sum, given some specified tolerance. M needs to be sufficiently large so that the additional contribution of any omitted term is less than some tolerance. In practice, with the tolerance level set to 1×10^{-12}, the number of terms in the finite series approximation seldom exceeds 100 for many wrapped distributions discussed, except for the WT distribution (see Pewsey, Lewis, and Jones (2007)). Analytic solutions cannot be found for the wrapped distributions discussed, therefore, numerical methods must be employed. Many algorithms, based on either Newton-Raphson type gradient methods or the simplex algorithm are available. We used function "optim" in the program R with the simplex algorithm

developed by Nelder and Mead (1965). The algorithm does not require the computation of partial derivatives (gradients) of the log-likelihood in (4.6) and this is appealing because the gradients for wrapped distributions can be hard to obtain. For WE, WSL, WN, WC, WS, WNL, and WGNL distributions, the MLEs are found by maximizing (4.6) with the corresponding α_p and β_p given in chapter 3. Although the density of some wrapped distributions (WC, WE, and WSL) have closed form, to avoid numerical problems, it is easier to maximize the Fourier representation of the density functions. Similarly, for the WT distribution, the MLEs can be found by maximizing the Fourier representation of the density functions.

5. *Wrapped t distribution*

The MLEs were found by maximizing the following log-likelihood function:

$$\begin{aligned} l(\mu,\lambda,\nu;\theta) &= -n[\log(\lambda)+\frac{1}{2}\log(\nu)+\log\{B(\frac{\nu}{2},\frac{1}{2})\}] \\ &+ \sum_{i=1}^{n}\log\left[\sum_{p=-\infty}^{\infty}\left\{1+\frac{(\theta_i+2\pi p-\mu)^2}{\lambda^2\nu}\right\}^{-\frac{\nu+1}{2}}\right], \end{aligned}$$

$$\text{where } B(\tfrac{\nu}{2},\tfrac{1}{2}) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi}}.$$

Asymptotic standard errors and correlation coefficients for the MLE's can be obtained from the observed information matrix, which can be calculated using the inverse of the numerical Hessian matrix evaluated at the maximum. In R, the Hessian matrix can be obtained by setting the option "hessian = TRUE" in calling the function "optim".

Many approaches are available which could be used to obtain initial values. Chapter 8 presents some approaches used to fit the examples. For a given data set, it is recommended to use of a variety of starting values for each unknown parameter to

accommodate the possibility of multiple local maxima.

Chapter 5

Goodness of fit: Watson's U^2

5.1 Overview

Suppose a random sample is obtained and a model for the data is proposed. A goodness-of-fit statistic is used to test whether the data are consistent with the model. Let $\theta_1, \theta_2, \ldots, \theta_n$ be a sample of directional observations. The null hypothesis for a goodness-of-fit test is

- H_0: the sample has been drawn from a population with a specified c.d.f. $F(\theta; \vec{\psi})$, where $\vec{\psi}$ is a vector of parameters, which may be known or unknown.

A well known goodness-of-fit statistic for circular data is Watson's U^2 (Watson 1961, 1962). It was originally introduced for continuous data, and later, was modified for grouped and discrete data by Choulakian, Lockhart, and Stephens (1994), (for completely specified distributions) and by Lockhart, Spinelli, and Stephens (2007) (for the case when parameters must be estimated from the sample). It has the following desirable properties:

1. the value of U^2 for a given sample does not depend on the origin used for θ, in contrast to other Cramér-von Mises statistics or the Kolmogorov statistic;

2. for grouped data, cells can be empty, unlike the Pearson's χ^2 statistic, which requires the expected counts in each group to be ≥ 5;

3. it has good power even for small sample sizes;

4. it works for any given distribution;

5. with some adjustments, the statistic does not depend on whether the data is measured clockwise or anti-clockwise.

5.2 The Watson's U^2 statistic for continuous data

The Empirical Distribution Function (EDF) of a sample $\theta_1, \theta_2, \ldots, \theta_n$ is defined as:

$$F_n(t) = \frac{(\#\theta_i \leq t)}{n} \text{ for } 0 \leq t \leq 2\pi.$$

The EDF statistics compare the EDF of a sample to the null distribution function $F(\theta; \vec{\psi})$, where $\vec{\psi}$ is a vector containing q known or unknown parameters. We can define

$$F(\theta; \vec{\psi}) = \int_0^{\theta} f(\phi; \vec{\psi}) d\phi, \tag{5.1}$$

abbreviated to $F(\theta)$.

The well known Cramér-von Mises statistic for data on the line is based on the discrepancy between $F_n(\theta)$ and $F(\theta)$:

$$W_n^2 = n \int_{-\infty}^{\infty} \{F_n(\theta) - F(\theta)\}^2 dF(\theta).$$

Watson (1961) adapted W^2 for circular data to obtain U^2 given by

$$U_n^2 = n \int_0^{2\pi} \left[F_n(\theta) - F(\theta) - \int_0^{2\pi} \{F_n(\phi) - F(\phi)\} dF(\phi) \right]^2 dF(\theta).$$

This change ensures that U^2 does not depend on the origin of θ around the circle. For simplicity, the dependence on n will be omitted in the notation U^2.

As outlined by Lockhart and Stephens (1985), a test of H_0 using the U^2 statistic is made as follows:

1. obtain MLEs ($\hat{\vec{\psi}}$) for unknown parameters as describe in chapter 4;
2. make the transformation $u_i = F(\theta_i; \hat{\vec{\psi}})$, $i = 1, 2, \cdots, n$, where unknown parameters are replaced by their MLEs if necessary;
3. put the u_i in ascending order: $u_{(1)}, u_{(2)}, \ldots u_{(n)}$;
4. calculate the U^2 statistic by computing the formula

$$U_n^2 = \sum_{i=1}^{n} \left(u_{(i)} - \frac{2i-1}{2n} \right)^2 - n(\bar{u} - 0.5)^2 + \frac{1}{12n}, \quad \bar{u} = \frac{\sum_{i=1}^{n} u_i}{n}. \tag{5.2}$$

5. compare U^2 with its null distribution (see chapter 7 for details).

If the parameters must be estimated, the distribution depends on $F(\theta; \hat{\vec{\psi}})$. The asymptotic ($n \longrightarrow \infty$) distribution can be found using theory given by Anderson and Darling (1952). For finite n, percentage points can be found by using Monte Carlo methods (i.e., parametric bootstrap methods). For many distributions, these points can be well approximated by the percentage points obtained from the asymptotic distribution (see D' Agostino and Stephens, 1986). Results will be shown for one parameter wrapped exponential distribution later.

5.3 The Watson's U_d^2 statistic for grouped data

Suppose the sample values $\theta_1, \theta_2, \ldots, \theta_n$ are grouped into k cells. Let p_i be the probability of falling into cell i with boundaries b_{i-1} and b_i, and define $t_i = \frac{p_i + p_{i+1}}{2}$, $i = 1, \ldots, k$,

where $p_{k+1} = p_1$. Using t_i instead of p_i is to ensure the statistic does not depend on the directions data are measured (clockwise or anti-clockwise). Assume first that the p_i are known. Given n independent observations, let o_i be the observed number of observations and $e_i = np_i$ be the expected number of observations in cell i. Define $S_j = \sum_{i=1}^{j} o_i$ and $T_j = \sum_{i=1}^{j} e_i$. Then S_j/n and $H_j = T_j/n$ are the cumulated relative frequencies of observed and expected values and correspond to the empirical distribution function $F_n(\theta)$ and the cumulative distribution function $F(\theta)$ for continuous distributions.

Let Z_j be $S_j - T_j, j = 1, 2, \ldots, k$. The weighted mean of Z_j is $\bar{Z} = \sum_{j=1}^{k} Z_j t_j$, with $t_{k+1} = t_1$. The Watson statistic U_d^2 for a discrete sample is defined as follows:

$$U_d^2 = n^{-1} \sum_{j=1}^{k} (Z_j - \bar{Z})^2 t_j. \tag{5.3}$$

When parameters must be estimated, $\hat{p}_i$ (the estimated probability for cell i) replaces p_i in the above calculations.

Similar to the case when data is continuous, to test the fit of the data to a model, U_d^2 should be calculated and compared to the null distribution of U_d^2 assuming the model is true. The asymptotic distribution theory for grouped U_d^2 will be given in the next chapter and used in later chapters to apply to specific circular distributions. For finite n, extensive Monte Carlo studies show that the null distribution converges quickly to the asymptotic distribution (see section 7.3 for details), so that asymptotic points can often be used.

Chapter 6

Asymptotic theory for U^2

6.1 U^2 for continuous data

The asymptotic theory for U^2 for the continuous distributions is well documented (see Anderson and Darling, 1952, Stephens, 1976a) and only a summary will be given here.

The asymptotic distributions depend on the behaviour of the empirical process for the u_i-values. Suppose that $F_n(u)$ is the EDF of the u_i. Then let

$$Z_n(u) = \sqrt{n}\{F_n(u) - u\}, \qquad 0 \le u \le 1$$

As $n \to \infty$, $Z_n(u)$ tends to a Gaussian process $Z(u)$, with mean zero and covariance function $\rho(s,t) = E\{Z(s)Z(t)\}$.

The covariance function for Case 0 (all parameters are known, the distribution is completely specified) is denoted by $\rho_0(s,t)$, where $\rho_0(s,t) = \min(s,t) - st + \frac{1}{2} - \frac{s-s^2}{2} - \frac{t-t^2}{2}$. When parameters are unknown, the covariance is found as follows: Suppose the distribution of x is defined by $F(x;\vec{\psi})$, where $\vec{\psi}$ is a q component vector containing q unknown parameters. Define $g_1(s) = \frac{\partial F(x;\vec{\psi})}{\partial \psi_1}$, ... $g_q(s) = \frac{\partial F(x;\vec{\psi})}{\partial \psi_q}$ evaluated at $x =$

$F^{-1}(s)$, and suppose V is the inverse of the Fisher information matrix. In general, for U^2,

$$\rho(s,t) = \rho_0(s,t) - G'(s)VG(t),$$

where $G(s)$ is the vector $\{g_1(s), g_2(s), ..., g_q(s)\}$ and $G'(s)$ is its transpose. Define

$$G'(s)\ VG(t) = \Omega'(s)\ \Omega(\mathrm{t}),$$

where $\Omega(s)$ is the vector $\{\omega_1(s), \omega_2(s), ..., \omega_q(s)\}$. The asymptotic distribution is that of

$$T = \sum_{i=1}^{\infty} \lambda_i v_i^2, \tag{6.1}$$

where v_i are independent standard normal random variables, and $\lambda_i, i = 1, 2, ..., \infty$ are the eigenvalues of the integral equation

$$\int_0^1 \rho(x,y) f(y) dy = \lambda f(x), \tag{6.2}$$

where $\rho(x,y)$ is the appropriate covariance.

The mean of T is given by $\sum_{i=1}^{\infty} \lambda_i = \int_0^1 \rho(t,t)dt$ and it is useful to evaluate the integral in order to calculate tables.

In order to determine T in (6.1), the eigenvalues λ_i must be found. This is done as follows. First consider λ_i^* and $f_i^*(t)$ the solutions of (6.2) for case 0. These λ_i^* will be called the standard eigenvalues. The standard eigenvalues are $\lambda_i^* = 4\pi^2 i^2$, with $i = 1, 2, ..., \infty$ and they occur twice. The corresponding eigenfunctions are $f_i(t) = \sqrt{2}\sin(2\pi it)$ and $f_i^*(t) = \sqrt{2}\cos(2\pi it)$.

Define $D_0(\lambda) = \Pi_i(\lambda - \lambda_i^*)$, and the coefficients $a_{i1} = \int_0^1 \omega_1(s) f_i^*(s) ds$, $i = 1, 2, ..., \infty$ and $a_{iq} = \int_0^1 \omega_q(s) f_i^*(s) ds$, $i = 1, 2, ..., \infty$. Let w_i be the vector $w_i' = (a_{i1}, a_{i2}, ..., a_{iq})$.

Then define

$$S(\lambda) = I + \sum_{i=1}^{\infty} \frac{w_i \; w_i' V}{\lambda - \lambda_i^*}. \tag{6.3}$$

The eigenvalues λ_i are given by the solutions of $D_0(\lambda)|S(\lambda)| = 0$, where $|.|$ indicates the determinant.

For illustration purpose, it is assumed that the distribution has two unknown parameters ($q = 2$). Let $S_1(\lambda) = 1 + \lambda \sum_{i=1}^{\infty} \frac{a_{i1}^2}{1-\lambda/\lambda_i^*}$, $S_2(\lambda) = 1 + \lambda \sum_{i=1}^{\infty} \frac{a_{i2}^2}{1-\lambda/\lambda_i^*}$, and $S_{12}(\lambda) = \lambda \sum_{i=1}^{\infty} \frac{a_{i1}a_{i2}}{1-\lambda\lambda_i^*}$. Define $T(\lambda) = S_1(\lambda)S_2(\lambda) - \{S_{12}(\lambda)\}^2$. Solutions of λs are given by $D_0(\lambda)T(\lambda) = 0$.

In summary,

1. As $n \to \infty$, the discrepancy $Z(\theta) = F_n(\theta) - F(\theta)$ tends to a Guassian process tied down at 0 and 1, with covariance $\rho(s,t)$ which must be found for each distribution tested.

2. The integral equation (6.2) must be solved for the values λ_j.

3. The asymptotic distribution is a sum of weighted χ^2 random variables with the weights λ_j.

Except for simple distributions, it is not easy to find $\rho(s,t)$ or, subsequently λ_j.

6.2 U_d^2 for grouped data

The asymptotic theory of $\hat{U}_d^2$ for the case when parameters are known is given by Choulakian, Lockhart, and Stephens (1994), and for the case when parameters are unknown and must be estimated efficiently is given by Lockhart, Spinelli, and Stephens

(2007). In this section, the steps to find the asymptotic distribution of U_d^2 when the tested distributions contains unknown parameters are given.

The definition of U_d^2 for grouped data is defined in Section 5.3. It is useful to put the definition into matrix notation. Let a prime, e. g. Z', denote the transpose of a vector or matrix. Let I be the $k \times k$ identity matrix, and let p' be the $1 \times k$ vector $(p_1, p_2, \ldots, p_k)$. Suppose o_i and e_i for $i = 1, \ldots, k$, are components of vectors $\vec{o}$ and $\vec{e}$, and let $\vec{d} = \vec{o} - \vec{e}$. Then $Z = A\vec{d}$, where A is the $k \times k$ partial-sum matrix

$$A = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ 1 & 1 & 0 & \ldots & 0 \\ 1 & 1 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \ldots & 1 \end{pmatrix}.$$

Suppose D is the $k \times k$ diagonal matrix whose j-th diagonal entry is $p_j,\ j = 1, \ldots, k$, and E is the $k \times k$ diagonal matrix with diagonal entries $t_j,\ j = 1, \ldots, k$. Recall from Section 5.3, $t_i = \frac{p_i + p_{i+1}}{2}$, where $i = 1, \ldots, k$ and where p_{k+1} is defined to be p_1, and $\vec{1}$ is a $k \times 1$ vector whose entry is 1. Then the matrix formulation of U_d^2 statistic is

$$S = Y'MY, \tag{6.4}$$

where $Y = Z/\sqrt{n} = A\vec{d}/\sqrt{n}$, and $M = (I - E\vec{1}\vec{1}')E(I - \vec{1}\vec{1}'E)$. This matrix formulation has been introduced for the asymptotic theory in Choulakian, Lockhart, and Stephens (1994), but it is also convenient to use it to calculate the U_d^2 statistic. When parameters must be estimated, the estimate $\hat{p}_i$ is the estimate of p_i evaluated at the MLEs. It is calculated from (4.1) using the parameter estimates. Then $\hat{t}_j, \hat{e}, \hat{T}_j,\ \hat{Z}_j$, and other statistics will be calculated using $\hat{p}_i$, and used in (5.3) or (6.4) to give $\hat{U}_d^2$.

Define a $k \times q$ matrix B with entries

$$B_{i,j} = \partial p_i / \partial \psi_j \tag{6.5}$$

for $i = 1, \cdots, k$ and $j = 1, 2, \cdots, q$. The matrix $B'D^{-1}B$ is the Fisher Information matrix for the estimates $\hat{\psi}_i$. Define $V = (B'D^{-1}B)^{-1}$. The asymptotic covariance of $\hat{\psi}$ is V/n.

The steps are

1. Calculate the inverse of Fisher's Information matrix of $\vec{\psi}$: $\hat{V} = (\hat{B}'\hat{D}^{-1}\hat{B})^{-1}$.
2. Calculate the covariance of $\hat{d}/\sqrt{n}$: $\hat{\Sigma}_d = \hat{D} - \hat{p}\hat{p}' - \hat{B}\hat{V}\hat{B}'$.
3. Calculate the covariance of $\hat{Z}/\sqrt{n}$: $\hat{\Sigma}_y = A\hat{\Sigma}_d A'$.
4. Find $\lambda_i, i = 1, 2, \cdots, k-1$, the $k-1$ non-zero eigenvalues of $\hat{M}\hat{\Sigma}_y$.
5. The final step is to find percentage points of the limiting distribution of $\hat{U}_d^2$, or the p-value of a given $\hat{U}_d^2$ (see section 6.2.1).

Comments on step 4: sometimes, it is more convenient to work with $X = \hat{M}^{1/2}\hat{Y}$; the covariance of X is $\Sigma_X = \hat{M}^{1/2}\hat{\Sigma}_y\hat{M}^{1/2}$, and the eigenvalues λ_i of Σ_X are the same as those of $\hat{M}\hat{\Sigma}_y$. Use of the symmetric matrix Σ_X will often make it easier to find eigenvalues when using some programs. The theory behind step 4 is as follows. As $n \to \infty$, the distribution of $\hat{Y}$ tends to the multivariate normal with mean 0 and covariance $\mathbf{\Sigma}_y$. Then $\hat{U}_d^2$ may be written

$$\hat{U}_d^2 = \hat{Y}'\hat{M}\hat{Y} = \sum_{i=1}^{k-1} \lambda_i s_i^2, \tag{6.6}$$

where $s_i = w_i'\hat{Y}$, and w_i are the $k-1$ eigenvectors corresponding to the eigenvalues λ_i, normalized so that $w_i'\hat{\mathbf{\Sigma}}_y w_j = \delta_{ij}$ where δ_{ij} is 1 if $i = j$ and 0 otherwise. The

distribution of a typical s_i tends to univariate normal with mean 0, variance 1, and in the limit, the s_i are independent. The limiting distribution of $\hat{U}_d^2$ is that of S_∞ where

$$S_\infty = \sum_{i=1}^{k-1} \lambda_i s_i^2 \tag{6.7}$$

which is a sum of independent weighted χ_1^2 random variables.

In summary, upon having the corresponding $\hat{B}$ matrix for the underlining distributions, the calculations of the asymptotic points of U_d^2 can be standardized into a computer program; therefore, computing the asymptotic points of U_d^2 becomes finding the $\hat{B}$ matrix for circular distributions tested.

6.2.1 Percentage Points and p-values

Percentage points of S_∞ can be well approximated in the upper tail, at levels $\alpha < 15\%$, by the percentage points of S_1, where S_1 has the distribution $a+b\chi_p^2$, and the a, b, p are chosen so that the first three cumulants $\kappa_1, \kappa_2, \kappa_3$ of S_1 match those of the asymptotic distribution of S_∞ in (6.7). The cumulants of the distribution in (6.7) are as follows (see Choulakian, Lockhart, and Stephens (1994)): the mean $\kappa_1 = \sum_{i=1}^{k-1} \lambda_i$, the variance $\kappa_2 = \sum_{i=1}^{k-1} 2\lambda_i^2$, and $\kappa_3 = 8\sum_{i=1}^{k-1} \lambda_i^3$. Then for the S_1 approximation, $b = \kappa_3/(4\kappa_2)$, $p = 8\kappa_2^3/\kappa_3^2$, and $a = \kappa_1 - bp$. However, the accuracy of this approximation gets worse at 25% and 50% levels. In such cases, the exact λ's are needed and they can be found to high accuracy by the method of Imhof (1961).

Table 6.1: Comparison of grouped (k=360 cells) and continuous asymptotic percentage points of U^2 for the von Mises distribution. The asymptotic percentage points for continuous data are taken from Table 1 of Lockhart and Stephens (1985). with $\kappa = 4$ and $\mu = \pi$

	$n\backslash\alpha$	0.50	0.25	0.10	0.05	0.025	0.01	0.005
Cont.	∞	0.047	0.067	0.093	0.113	0.132	0.158	0.178
Group	∞	0.046	0.068	0.094	0.114	0.133	0.158	0.177

6.3 Asymptotic distribution of U^2 for von Mises distributions and wrapped exponential distributions

6.3.1 U^2 for continuous data

Lockhart and Stephens (1985) provided percentage points for the asymptotic distribution of Watson's U^2 statistic for testing the hypothesis that a sample comes from the two-parameter (unknown) von Mises distribution when data is assumed continuous. Table 6.1 compares the asymptotic percentage points of U^2 when data is assumed both continuous and grouped (360 cells of width 1 degree) and shows that for finite k, asymptotic % points for grouped data are very close to those for continuous data.

The procedure of obtaining the asymptotic distribution of U^2 for continuous data is illustrated here when the null distribution is specified as a one parameter WE distribution. To avoid conflicts of notations, we define η to be the shape parameter of the WE and λ's to be the eigenvalues. The WE distribution has a c.d.f. defined as:

$$F(\theta, \eta) = \frac{\eta e^{-\eta\theta}}{1 - e^{-2\pi\eta}}. \tag{6.8}$$

Differentiating (6.8) with respect to η

$$g(s) = \frac{dF(\theta,\eta)}{d\eta} = \frac{\theta e^{-\eta\theta}(1-e^{-2\pi\eta}) - 2\pi e^{-2\pi\eta}(1-e^{-\eta\theta})}{(1-e^{-2\pi\eta})^2}, \tag{6.9}$$

where $\theta = -\frac{1}{\eta}\log[1 - s(1-e^{-2\pi\eta})]$.

Since there is only one parameter in this case, $\omega(s) = kg(s)$, where k is defined as

$$\frac{1}{k^2} = \int_0^{2\pi}\{\frac{d}{d\eta}\log f(\theta;\eta)\}^2 f(\theta;\eta)\mathrm{d}\theta = \int_0^{2\pi}\frac{\{(1-\eta\theta)(1-e^{-2\pi\eta}) - 2\pi\eta e^{-2\pi\eta}\}^2 e^{-\eta\theta}}{\eta(1-e^{-2\pi\eta})^3}\mathrm{d}\theta. \tag{6.10}$$

The integral can be solved numerically in R. Having obtained k, it follows that $a_i = \int_0^1 \omega(s)\sqrt{2}\sin(2\pi is)ds = 0$, $i = 1,2,...,\infty$ and $a_i^* = \int_0^1 \omega(s)\sqrt{2}\cos(2\pi is)ds$, $i = 1,2,...,\infty$. The above two integrals need to be solved numerically. The λ's are found by solving $S(\lambda) = 1 + \lambda\sum_{i=1}^{\infty}\frac{(a_i^*)^2}{1-\lambda/\lambda_i^*} = 0$. Note that such λ's always exists between any two consecutive standard eigenvalues ($\lambda^* = 4\pi^2 i^2$). Percentage points for U^2 are calculated from the λ's by using the methods in Subsection 6.2.1.

6.3.2 U_d^2 for grouped data

Calculating the asymptotic distribution for U_d^2 is computationally easier than for U^2. One needs to start from the B matrix. For the WE distribution, the B matrix is a $k \times 1$ matrix with entries

$$B_{i1} = \frac{dp_i}{d\eta} = \frac{d}{d\eta}\int_{b_{i-1}}^{b_i} f(\theta;\eta)\mathrm{d}\theta$$

when the c.d.f. is available, it is easier to use

$$\frac{dp_i}{d\eta} = \frac{d}{d\eta}\{F(b_i;\eta) - F(b_{i-1};\eta)\},$$

Table 6.2: Comparison of grouped (k=360 cells) and continuous asymptotic percentage points of U^2 for Wrapped Exponential distribution with $\eta = 2$.

	$n\backslash\alpha$	0.50	0.20	0.10	0.05	0.025	0.01	0.005
Cont.	∞	0.0582	0.0992	0.1296	0.1598	0.1899	0.2295	0.2593
Group	∞	0.0586	0.0999	0.1306	0.1611	0.1913	0.2312	0.2613

where $\frac{d}{d\eta}F(x;\eta)$ is shown in equation (6.9). The procedure is as follows: (1) having obtained the $\hat{B}matrix$, follow step 1 to 4 in Section 6.2 and calculate V, $\hat{\Sigma}_d$, $\hat{\Sigma}_y$, and $\hat{M}\hat{\Sigma}_y$ (or Σ_X), (2) find λ from $\hat{M}\hat{\Sigma}_y$ (or Σ_X) using command "eigen" in R, (3) compute percentage points for U_d^2 from the λ's by using the methods in Subsection 6.2.1.

Table 6.2 gives the asymptotic percentage points of U^2 assuming both continuous and grouped (360 cells of width 1 degree). It can be seen that for finite k, grouped asymptotic % points are very close to those for continuous data.

Chapter 7

Testing fit

To test the null hypothesis (H_0) that the random sample $\theta_1, \theta_2, \ldots, \theta_n$ comes from a given distribution, one first calculates U^2 or U_d^2 from the sample, as described in Chapter 5. The statistic must be compared with significance points of its distribution assuming H_0 is true. The H_0 is rejected at significance level α if the statistic used is greater than the αth percentage point. These significant points of the U^2 statistic can be obtained either using the parametric bootstrap (Monte Carlo) method (discussed in Section 7.1 and 7.2), or using the asymptotic distribution (discussed in Section 6.2 and 6.1). At present, the asymptotic distributions for continuous data are only available for a limited number of null distributions: the von Mises distribution (Lockhart and Stephens, 1985), and the wrapped exponential distribution (see Subsection 6.3.1). We recommend using the grouped asymptotic method proposed in the thesis, as we have shown by Monte Carlo studies that the finite-n sample distribution of U_d^2 is well approximated by the asymptotic distribution. This is for any sample size which might be used in testing fit, and avoids a lengthy Monte Carlo estimation of the the percentage points.

The asymptotic percentage points of the U_d^2 statistic for different underlying cir-

cular distributions discussed in Chapter 3 have been checked by taking 1000 or 5000 Monte Carlo samples of different sizes, from a range of parameter values. For each sample, the parameters were estimated and $\hat{U}^2$ was calculated as continuous data, before comparing the grouped asymptotic percentage points. Some of the results for upper-tail percentage points, are given in Tables 7.2,7.3,7.4,7.5, and 7.8.

Convergence of the distributions to their asymptotic distributions was rapid, and for assessing significance, the asymptotic points can be used for samples of sizes of $n > 20$.

7.1 Parametric bootstrap distributions of the U^2 statistic for grouped data

The parametric bootstrap distributions of the U^2 statistic, the percentage points, and p-values are obtained as follows.

1. Obtain ML estimates for the unknown parameters and calculate U^2 as described in Section 5.3 (Equation 5.3).

2. Let N_{mc} be the number of bootstrap samples. For each g from $1, 2, \ldots, N_{mc}$

   (a) Generate the g^{th} bootstrap sample with the same number of observations, say n, as the original sample, $\theta_{1g}, \theta_{2g}, \ldots, \theta_{ng}$, from the distribution specified by $F(\theta, \hat{\psi})$, where $\hat{\psi}$ is a vector of ML estimates from the original sample.

   (b) Put the g^{th} bootstrap sample $\theta_{1g}, \theta_{2g}, \ldots, \theta_{ng}$, into k cells with the same boundaries as the original sample, obtain m_g, the counts of g^{th} sample.

   (c) Obtain ML estimates for any unknown parameters from the g^{th} bootstrap sample.

   (d) Calculate $\hat{p}_{ig}$ using (4.1), and subsequently calculate $\hat{t}_{ig}, \hat{E}_g, \hat{e}_{ig}, \hat{d}_{ig} = \vec{d}$, and

$\hat{Z}_g$ for $i = 1, 2, \ldots, k$ as discussed in Section 6.2.

(e) The U_d^2 statistic for the g^{th} sample is calculated as

$$U_g^2 = \hat{Y}_g'(I - \hat{E}_g 11')\hat{E}_g(I - 11'\hat{E}_g)\hat{Y}_g$$

where $\hat{Y}_g = \hat{Z}_g/\sqrt{n} = A\vec{d}/\sqrt{n}$.

3. The distribution of $U_g^2, g = 1, 2, \ldots, N_{mc}$ is the parametric bootstrap distribution of the U_d^2 statistic from the original sample. Put $U_g^2, g = 1, 2, \ldots, N_{mc}$ into ascending order, the k^{th} order statistic $U_{(k)}^2$ is the $\frac{N_g - k}{N_g} \times 100\%$ point. Let N_g be the number of bootstrap samples with U_d^2 greater than the U_d^2 statistic from the original sample. An approximate p-value for the test specified by H_0 is $\frac{N_g}{N_{mc}}$. Note that the accuracy of the p-value increases as N_{mc} increases.

7.2 Parametric bootstrap distributions of the U^2 statistic for continuous data

1. Obtain ML estimates for the unknown parameters and calculated U^2 as described in Section 5.2 (Equation 5.2).

2. Let N_{mc} be the number of bootstrap samples. For each g from $1, 2, \ldots, N_{mc}$, do the following: (a) Generate the g^{th} bootstrap sample with the same number of observations, say n, as the original sample, $\theta_{1g}, \theta_{2g}, \ldots, \theta_{ng}$, from the distribution specified by $F(\theta, \hat{\psi})$, where $\hat{\psi}$ is a vector of ML estimates from the original sample.

   (b) Obtain ML estimates for any unknown parameters from the g^{th} bootstrap sample (see Section 4.2 for continuous data).

   (c) For each i in $1, 2, \ldots, n$, calculate $z_{ig} = F(\theta_{ig})$, where unknown parameters

are replaced by the ML estimates obtained in (b).

(d) Put the z_{ig} into ascending order.

(e) The U^2 statistic for the g^{th} sample is calculated as

$$U_g^2 = \sum_{i=1}^{n} \left\{ z_{(i)g} - \frac{(2i-1)}{2n} \right\}^2 - n(\bar{z}_g - 0.5)^2 + \frac{1}{12n}, \quad \bar{z}_g = \frac{\sum_{i=1}^{n} z_{ig}}{n}. \tag{7.1}$$

3. The distribution of $U_g^2, g = 1, 2, \ldots, N_{mc}$ is the parametric bootstrap distribution of the U^2 statistic from the original sample. Put $U_g^2, g = 1, 2, \ldots, N_{mc}$ into ascending order, the k^{th} order statistic $U_{(k)}^2$ is the $\frac{N_g - k}{N_g} \times 100\%$ point. Let N_g be the number of bootstrap samples with U^2 greater than the observed U^2 statistic from the original sample. An approximate p-value for the test specified by H_0 is $\frac{N_g}{N_{mc}}$. Note that the accuracy of the p-value increases as N_{mc} increases.

7.3 Tables of Monte Carlo and grouped asymptotic percentage points for some circular distributions

This section presents a number of tables comparing the U^2 statistic percentage points obtained by parametric bootstrap (Monte Carlo) methods along with asymptotic percentage points assuming data are grouped, when the null hypothesis is one of the distributions discussed in Chapter 3 and all parameters in the distributions need to be estimated. Tables 7.6, 7.7, and 7.8 show the behaviour of the asymptotic distribution of U_d^2 as $k \to \infty$.

It can be seen that for finite n, Monte Carlo percentage points for U^2 converge to asymptotic percentage points. For WE, WC, WN, WS, and WSL distributions, the Monte Carlo percentage points converge for $n = 100$ (Tables 7.1, 7.2, 7.3, 7.8, and 7.7).

Table 7.1: For the wrapped Exponential distributions ($\eta = 2$): the percentage points of the U^2 statistic for sample sizes $n = 20, 50, 100$, and 200; 1000 Monte Carlo (M.C.) samples were used for each case, comparing with points obtained using grouped (k=360 cells) asymptotic method (α is the level of significance).

	$n\backslash\alpha$	0.50	0.20	0.10	0.05	0.025	0.01	0.005
M.C.	20	0.0608	0.1052	0.1397	0.1622	0.2084	0.2405	0.2561
M.C.	50	0.0618	0.1015	0.1375	0.1720	0.2088	0.2406	0.2772
M.C.	100	0.0588	0.1007	0.1276	0.1524	0.1898	0.2227	0.2677
M.C.	200	0.0585	0.0998	0.1298	0.1638	0.1959	0.2430	0.2539
Group	∞	0.0586	0.0999	0.1306	0.1611	0.1913	0.2312	0.2613

Table 7.2: For the wrapped Cauchy distributions ($\mu = \pi$ and $\rho = -0.131$): the percentage points of the U^2 statistic for sample sizes $n = 50$, and 100; 1000 Monte Carlo (M.C.) samples were used for each case, comparing with points obtained using grouped (k=360 cells) asymptotic method (α is the level of significance).

	$n\backslash\alpha$	0.50	0.25	0.10	0.05	0.025	0.01	0.005
M.C.	50	0.0297	0.0409	0.0521	0.0606	0.0686	0.0791	0.0873
M.C.	100	0.0292	0.0393	0.0512	0.0613	0.0680	0.0780	0.0826
Group	∞	0.0293	0.0397	0.0521	0.0610	0.0696	0.0808	0.0891

For the WSN distribution, the Batschelet's distribution, and the CB distribution, the Monte Carlo percentage points converge slower, but they converge for $n = 500$ (see Tables 7.4, 7.5, and 7.6).

Tables 7.6, 7.7, and 7.8 also show that with an increase in the number of cells k, from 360 to 1800 or to 3600, the grouped asymptotic percentage points change very little. So asymptotic percentage points assuming $k = 360$ cells is a good approximation for asymptotic percentage points assuming continuous data.

Note Table 7.7 has 10000 M.C. samples and the upper tail M.C. percentage points are very accurate even at the $\alpha = 0.001$ significance level.

Table 7.3: For the wrapped Normal distributions (μ = π and σ = 1.3): the percentage points of the U^2 statistic for sample sizes n = 50, and 100; 1000 Monte Carlo (M.C.) samples were used for each case, comparing with points obtained using grouped (k=360 cells) asymptotic method (α is the level of significance).

	$n\backslash\alpha$	0.50	0.25	0.10	0.05	0.025	0.01	0.005
M.C.	50	0.0467	0.0657	0.0931	0.1073	0.1226	0.1428	0.1614
M.C.	100	0.0450	0.0643	0.0904	0.1117	0.1314	0.1551	0.1701
Group	∞	0.0451	0.0660	0.0921	0.1113	0.1302	0.1549	0.1734

Table 7.4: For the wrapped skewed Normal distributions (ξ = π, λ = 5, and η = 2): the percentage points of the U^2 statistic for sample sizes n = 50, 100, and 300; 1000 Monte Carlo (M.C.) samples were used for each case, comparing with points obtained using grouped (k=360 cells) asymptotic method (α is the level of significance).

	$n\backslash\alpha$	0.50	0.25	0.10	0.05	0.025	0.01	0.005
M.C.	50	0.0418	0.0647	0.0900	0.1086	0.1328	0.1588	0.1906
M.C.	100	0.0416	0.0606	0.0834	0.1054	0.1268	0.1492	0.1852
M.C.	300	0.0403	0.0581	0.0862	0.1031	0.1244	0.1455	0.1666
Group	∞	0.0392	0.0583	0.0834	0.1022	0.1211	0.1459	0.1647

Table 7.5: For the Batschelet's distributions (κ = 0.5 and ν = 0.2): the percentage points of the U^2 statistic for sample sizes n = 100, 300, and 500; 1000 Monte Carlo (M.C.) samples were used for each case, comparing with points obtained using grouped (k=360 cells) asymptotic method (α is the level of significance).

	$n\backslash\alpha$	0.50	0.25	0.10	0.05	0.025	0.01	0.005	0.001
M.C.	100	0.0469	0.0715	0.1060	0.1314	0.1631	0.1909	0.2310	0.3021
M.C.	300	0.0474	0.0725	0.1050	0.1299	0.1543	0.1870	0.2220	0.2968
M.C.	500	0.0459	0.0705	0.1035	0.1293	0.1538	0.1879	0.2176	0.3017
Group	∞	0.0437	0.0704	0.1069	0.1350	0.1632	0.2008	0.2293	0.2959

Table 7.6: For the Circular Beta distributions ($\alpha = 3$ and $\beta = 2$): the percentage points of the U^2 statistic for sample sizes $n = 100$, 300, and 1000; 1000 Monte Carlo (M.C.) samples were used for each case, comparing with points obtained using grouped ($k = 360$ and $k = 1800$ cells) asymptotic method (α is the level of significance).

	$n\backslash\alpha$	0.50	0.25	0.10	0.05	0.025	0.01	0.005
M.C.	100	0.0456	0.0679	0.1019	0.1274	0.1600	0.1956	0.2286
M.C.	300	0.0463	0.0695	0.1014	0.1249	0.1506	0.1803	0.2026
M.C.	1000	0.0438	0.0705	0.1055	0.1341	0.1666	0.2082	0.2247
Group	$k = 360$	0.0438	0.0693	0.1057	0.1342	0.1633	0.2023	0.2321
Group	$k = 1800$	0.0438	0.0692	0.1056	0.1342	0.1633	0.2023	0.2321

Table 7.7: For the wrapped skewed Laplace distributions ($\lambda_1 = 2$, $\lambda_2 = 1$, and $\mu_0 = 0$): the percentage points of the U^2 statistic for sample sizes $n = 50$, 100, and 500; 10000 Monte Carlo (M.C.) samples were used for each case, comparing with points obtained using grouped ($k = 360$ and $k = 3600$ cells) asymptotic method (α is the level of significance).

	$n\backslash\alpha$	0.50	0.10	0.05	0.025	0.01	0.005	0.001
M.C.	50	0.0452	0.0922	0.1130	0.1336	0.1628	0.1878	0.2366
M.C.	100	0.0449	0.0910	0.1119	0.1343	0.1574	0.1777	0.2214
M.C.	500	0.0447	0.0909	0.1098	0.1310	0.1634	0.1891	0.2363
Group	$k = 360$	0.0434	0.0920	0.1125	0.1328	0.1596	0.1797	0.2264
Group	$k = 3600$	0.0435	0.0922	0.1127	0.1330	0.1598	0.1800	0.2266

Table 7.8: For the wrapped Stable distributions ($\alpha = 1$, $\beta = -0.9$, $\gamma = 0.5$, and $\mu = \pi$): the percentage points of the U^2 statistic for sample sizes $n = 100$, and 300; 1000 Monte Carlo (M.C.) samples were used for each case, comparing with points obtained using grouped ($k = 36$, $k = 180$, $k = 360$, and $k = 1800$ cells) asymptotic method (α is the level of significance).

	$n\backslash\alpha$	0.50	0.25	0.10	0.05	0.025	0.01	0.005
M.C.	100	0.0294	0.0400	0.0555	0.0652	0.0740	0.0927	0.0986
M.C.	300	0.0288	0.0409	0.0545	0.0653	0.0766	0.0954	0.0997
Group	$k = 36$	0.0292	0.0425	0.0586	0.0702	0.0814	0.0961	0.1070
Group	$k = 180$	0.0296	0.0414	0.0557	0.0660	0.0762	0.0893	0.0991
Group	$k = 360$	0.0293	0.0407	0.0545	0.0646	0.0744	0.0871	0.0967
Group	$k = 1800$	0.0292	0.0405	0.0542	0.0641	0.0738	0.0864	0.0958

Chapter 8

Examples

8.1 Example 1: ants data (grouped MLE, grouped U_d^2)

The directions chosen by 100 ants in response to a stimulus was analyzed; (see Figure 2.1, the black target was placed opposite the origin) and a description of the data is in Chapter 2. For these data, the mean resultant $\bar{\theta}$ = 3.196 (radians) and the mean resultant length $\bar{R}$ = 0.610. Assuming these data were continuous, Rayleigh's test for uniformity gave a p-value < 0.000. Fisher (1995) examined these data by fitting a von Mises distribution and the von Mises Q-Q plot suggested systematic departure from a straight line. Jammalamadaka and Kozubowski (2003) fitted a symmetric Laplace model with known μ_0 and obtained the moment estimators $\hat{\lambda}_1 = \hat{\lambda}_2 = 1.3$ and examined the fit with a linear histogram of these data along with the fitted wrapped laplace density. Bentley (2006) studied these data by using a mixture von Mises and uniform distribution and the P-P plot appeared to fit well along the line $y = x$. As these data were measured in the nearest 10 degrees (see Table 8.1), they were really grouped data and should be analyzed by grouped data techniques.

Table 8.1: Table of the ant data grouped in the nearest 10 degrees: the first cell boundaries are 355° and 5°

1	1	0	2	1	1	2	1	1	1	0	2	2	1	3	5	7	6
17	8	13	8	5	2	0	1	1	1	2	1	2	0	0	1	0	1

A number of circular distributions discussed in Chapter 3 were fitted to these data set using the following steps:

1. Find the ML estimates for unknown parameters, as described in Section 4.1 from Chapter 4.

2. Calculate the Akaike information criterion (AIC), the Bayesian information criterion (BIC), and the likelihood ratio statistic (LRS) from the estimated log-likelihood value, if the distribution is nested under the WGNL distribution.

3. Compute U_d^2 (see Section 5.3 in Chapter 5) and compute its asymptotic percentage points as described in Section 6.2 in Chapter 6 using the ML estimates obtained in step 1.

4. Calculate p-values for the U_d^2 statistic and the appropriate likelihood ratio statistics (LRS).

8.1.1 von Mises fit

When the VM distribution was fitted to the data, using $\bar{\theta}$ = 3.196 and 1.550 as starting values for μ and κ, the ML estimates were $\hat{\kappa}$ = 1.559, and $\hat{\mu}$ = 3.196. The Watson's U_d^2 test suggested that the VM was not an adequate model (Table 8.2).

8.1.2 Wrapped Normal fit

When the WN distribution was fitted to the data, using $\bar{\theta}$ = 3.196 and 0.988 as starting values for μ and σ^2, the ML estimates (and standard deviation) were $\hat{\mu}$ = 3.133(0.132), and $\hat{\sigma}^2$ = 1.224(0.179), giving an estimate of the mean resultant length ρ = 0.542 and the asymptotic correlation between the two parameters was 0.030. The Watson's U_d^2 test suggested that the WN was not an adequate model (Table 8.2).

8.1.3 Wrapped Cauchy fit

When the WC distribution was fitted to the data, using $\bar{\theta}$ = 3.196 and 0.494 as starting values for μ and ρ, the ML estimates (and standard deviation) were $\hat{\mu}$ = 3.242(0.063), and $\hat{\rho}$ = 0.427(0.063), giving an estimate of the mean resultant length $e^{-\rho}$ = 0.652 and the asymptotic correlation between the two parameters was -0.025. The Watson's U_d^2 test suggested that the WC was not an adequate model (Table 8.2).

8.1.4 Wrapped t fit

When the 3-parameter WT distribution was fitted to the data, using $\bar{\theta}$ = 3.196, 1, and 0.4 (from the contour plot of the profile log-likelihood) as starting values for μ, λ, and ν, the ML estimates (and standard deviation) were $\hat{\mu}$ = 3.244(0.061), $\hat{\nu}$ = 0.802(0.215), and $\hat{\lambda}$ = 0.388(0.076). The asymptotic correlation between $\hat{\mu}$ and the other two estimates were negative and were quite small. The correlation between $\hat{\nu}$ and $\hat{\lambda}$ was 0.588. Since ν (d.f.) was less than 1, the distribution had a heavier tailer than the WC distribution. The log-likelihood value of the WT fit was slightly bigger than the WC fit. It suggested a similar fit with the WC, but the WC had 1 less parameter. The AIC values for the WC and the WT models were: 616.29 and 617.71, while the BIC values were 621.50 and 625.52. These both favoured the WC

over the WT models. However, the Watson's U_d^2 test suggested that the WC was not an adequate model (Table 8.2).

8.1.5 Wrapped Stable fit

When the 4-parameter WS distribution was fitted to the data, using the estimates $\hat{\mu}$ = 3.242, $\hat{\rho}$ = 0.427 (from the WC estimates), 0.9, and -0.4 (from the contour plot of the profile log-likelihood) as starting values for μ, γ, α, and β respectively, the ML estimates (and standard deviation) were $\hat{\mu}$ = 3.269(0.282), $\hat{\gamma}$ = 0.426(0.780), $\hat{\alpha}$ = 0.897(0.153), and $\hat{\beta}$ = −0.421(0.067). The largest asymptotic correlation among the ML estimates was 0.245, between $\hat{\mu}$ and $\hat{\gamma}$. The log-likelihood value of the WS fit was −305.36, slightly bigger than the value of the WC fit. It suggested a similar fit with the WC, but the WC distribution has 2 less parameters. The AIC values for the WC and the WT models were: 616.29 and 618.72, while the BIC values were 621.50 and 629.14. These both favoured the WC over the WS models. The Watson's U_d^2 test suggested that the WS was not an adequate model (Table 8.2).

8.1.6 Wrapped skew-Laplace fit

When the 3-parameter WSL distribution was fitted to the data, the ML estimates (and standard deviation) were $\hat{\lambda}_1$ = 1.890(0.512), $\hat{\lambda}_2$ = 1.0002(0.180), and $\hat{\mu}_0$ = 3.414(0.120). The asymptotic correlation between $\hat{\lambda}_1$ and $\hat{\lambda}_2$ was -0.488, between $\hat{\lambda}_1$ and $\hat{\mu}_0$ was 0.732, and between $\hat{\lambda}_2$ and $\hat{\mu}_0$ was -0.648. The Watson's U_d^2 test suggested that the WSN was not an adequate model (Table 8.2).

8.1.7 Wrapped normal Laplace fit

When the 4 parameter WNL distributions was fitted to the data, using $\bar{\theta}$ = 3.196 and $\bar{R}$ = 0.610 as the starting values for μ and ρ, then using (3.25) and (3.26) to solve for

starting values of τ and η, using inverse of the ML estimates from the WSL $\frac{1}{\lambda_1}$ and $\frac{1}{\lambda_2}$ as the starting values for a and b, the ML estimates (and standard deviation) were $\hat{a} = 0.529(0.118)$, $\hat{b} = 0.999(0.148)$, $\hat{\eta}_0 = 3.414(0.067)$, and $\hat{\tau}^2 = 0.000$.

The log-likelihood value from the WNL fit was −310.40, which was very similar to the log-likelihood value from the WSL fit. This was because $\hat{\tau}^2$ was very small, recall that when $\hat{\tau}^2 = 0$, the WNL distribution becomes a WSL with $\lambda_1 = \frac{1}{a}$ ($\frac{1}{0.529} = 1.889$) and $\lambda_2 = \frac{1}{b}$ ($\frac{1}{0.999} = 1.001$). The AIC values for the WNL and the WSL models were 628.79, and 626.8, while the BIC values were 639.21, and 634.62. These both favoured the WSL over the WNL. However, the Watson's U_d^2 test suggested that neither model was adequate (Table 8.2).

8.1.8 Wrapped Generalized normal Laplace fit

When the 5-parameter WGNL distributions was fitted to the data, using the ML estimates from WNL and $\zeta = 1$ as starting values, the ML estimates (and standard deviation) were $\hat{a} = 0.0076(0.542)$, $\hat{b} = 1.194(0.149)$, $\hat{\eta} = 3.773(0.546)$, $\hat{\tau}^2 = 0.025(0.022)$, and $\hat{\zeta} = 1.068(0.028)$. The largest asymptotic covariance among the ML estimates was that between $\hat{a}$ and $\hat{\eta}$ (-0.294). The Watson's U_d^2 test suggested that the WGNL was not an adequate model (Table 8.2).

8.1.9 Circular Beta fit

When the 3-parameter CB distributions was fitted to the data, the ML estimates (and standard deviation) were $\hat{\alpha} = 1.096(0.090)$, $\hat{\beta} = 0.970(0.103)$ and $\hat{\eta} = 4.633(0.172)$. The asymptotic correlation between $\hat{\alpha}$ and $\hat{\beta}$ was -0.671, between $\hat{\alpha}$ and $\hat{\eta}$ was -0.109 and between $\hat{\beta}$ and $\hat{\eta}$ was -0.100. The Watson's U_d^2 test suggested that the CB was not an adequate model (Table 8.2).

Table 8.2: Watson's U_d^2 statistics, percentage points, U_d^2 p-value, the log-likelihood value (l), AIC, and BIC under different null distributions in Example 8.1.

Distn	U_d^2	10%	5%	1%	p-value	l	AIC	BIC
VM	0.283	0.079	0.095	0.132	<0.005	-316.72	637.44	642.65
WN	0.596	0.093	0.113	0.157	<0.001	-323.69	651.38	656.59
WC	0.081	0.054	0.064	0.087	0.016	-306.15	616.29	621.50
WT	0.080	0.044	0.053	0.073	0.006	-305.85	617.71	625.52
WS	0.074	0.040	0.048	0.066	0.005	-305.36	618.72	629.14
WSL	0.119	0.069	0.085	0.120	0.010	-310.40	626.80	634.62
WNL	0.163	0.060	0.074	0.105	<0.001	-310.40	628.79	639.21
WGNL	0.086	0.056	0.069	0.092	0.025	-306.95	623.90	636.93
CB	1.687	0.104	0.133	0.201	0.000	-343.52	693.05	700.86
WE	0.334	0.115	0.143	0.206	<0.001	-323.93	651.87	657.08
WSN	0.518	0.075	0.092	0.129	<0.000	-321.14	648.28	656.09

8.1.10 Wrapped exponential fit

When the WE distributions was fitted to the data, the ML estimates were $\hat{\mu} = 2.488$, and $\hat{\lambda} = 0.725$. The Watson's U_d^2 test suggested that the WE was not an adequate model (Table 8.2).

8.1.11 Wrapped skew-normal fit

When the 3-parameter WSN distributions was fitted to the data, letting $\hat{\mu} = 3.133, \hat{\sigma} = 1.106$ the ML estimates from WN and 0 be the starting values for ξ, η, and λ, the ML estimates (and standard deviation) were $\hat{\xi} = 4.129(0.512)$, $\hat{\eta} = 1.628(0.180)$, and $\hat{\lambda} = -2.394(0.120)$. Again, the Watson's U_d^2 test suggested that the WSN was not an adequate model (Table 8.2).

8.1.12 Likelihood Ratio Tests

Between the WGNL fit and the WNL fit, the likelihood ratio statistic was 6.9 with 1 d.f.; between the WNL fit and the WSL fit, the LRS was less than 0.001 with 1 d.f.; and between the WGNL fit and the WN fit, the LRS was 33.48 with 3 d.f.. This suggested the WGNL model was a better fit than all its nested models and the WNL did not fit better than its nested model, the WSL.

In conclusion, both AIC and BIC suggested the WC (616.29 and 621.50) provided the best fit for the ants data among all the circular distributions fitted, while the WT (617.71 and 625.52) and the WS (618.72 and 629.14) provided slightly worse fit. However, according to Watson's test, none of the models suggested adequate fit since the p-values were all less than 0.05. Bentley (2006) fitted a mixture of the von Mises and the uniform distributions to these data assuming the observations were continuous by adding a random component to the observations and obtained $U^2 = 0.019$ with p-value$= 0.87$. This mixture however has 5 parameters.

Table 8.3 gives rankings based on AIC and BIC and on the p-value of the U_d^2 statistics. The AIC and BIC gave similar rankings while the U_d^2 p-value gave a very different ranking. However, the approximations $(a + b\chi_p^2)$ of the U_d^2 asymptotic distributions (particularly the upper tail) might not be accurate enough to correctly rank the fitted distributions, since the p-values were less than 3%.

The analysis of this example illustrates how the use of model selection criteria alone did not necessarily lead to an adequate model. While the WC was suggested as the best model by both AIC and BIC, the Watson's goodness-of-fit test indicated that it did not really fit these data well.

Table 8.3: Model rankings based on: U_d^2 p-value, AIC, BIC, and the log-likelihood value.

	WGNL	WC	WSL	WT	WS	VM	WNL	WE	WN	WSN	CB
p-value	1	2	3	4	5	6	7	8	9	10	11
AIC	4	1	5	2	3	7	6	10	9	8	11
BIC	5	1	4	2	3	7	6	10	9	8	11
l	4	3	6	2	1	9	5	10	9	8	11

8.2 Example 2: Fish reacting to artificial sunlight (grouped MLE, grouped U_d^2)

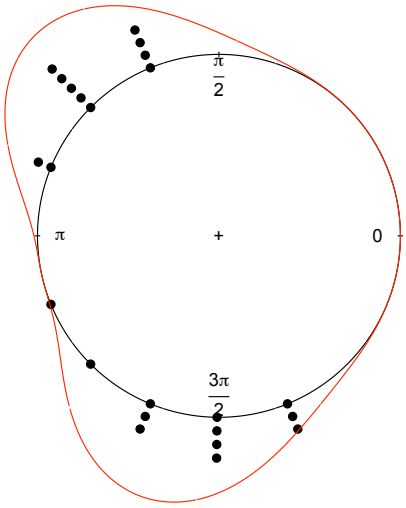
The data set consists of directions chosen by 23 fish reacting to an articial light, the data are described descriptions in Chapter 2. These data appear to have two modes that are not opposite to each other; therefore, it makes sense to only fit the circular beta (CB) distribution to these data because all other circular distributions mentioned are unimodal.

When fitting the CB distribution to these data, the ML estimates (and standard deviation) were $\hat{\alpha}$ = 4.669(0.059), $\hat{\beta}$ = 1.97(0.142) and $\hat{\mu}$ = 3.570(0.552) and the maximized log-likelihood value was −49.086. The Watson's statistic was U_d^2 = 0.052, which when compared with its asymptotic values at 50% = 0.038, at 90% = 0.1049, and at 95% = 0.1348, yielded a p-value of 0.352. The U_d^2 statistic suggested good fit. The plot of the fitted CB model is in Figure 8.1. Note that this example was a good illustration of U_d^2 used to test fit, whereas Pearson's X^2 could not be used because of low cell counts.

8.3 Example 3: Sandstone data (continuous MLEs, grouped U_d^2)

This data set consists of 104 cross-bed measurements from Himalayan molasse in Pakistan (see Figure 2.3). these data were measured to the degree, so, strictly speaking, they were grouped data with cell width of 1°. However, for illustrative proposes, these data were assumed to be continuous. Fisher (1995) suggested fitting a von Mises distribution and using a von Mises Q-Q plot to examine the fit. Pewsey (2007) compared the fit of a wrapped Cauchy and a wrapped t distribution using the likelihood ratio

Figure 8.1: Fish react to artificial sunlight fitted to a Circular Beta distribution.



test and presented a wrapped t Q-Q plot. The Q-Q plots showed that both VM and WT models provide adequate fit.

As the number of cells gets large ($k \to \infty$), it has been shown that for the same underlying distribution, the ML estimation for continuous data and the ML estimation for grouped data were essentially maximizing over the same log-likelihood (see proof in section 4.1), and the ML estimates and subsequently U^2 are very similar from both methods. The following section presents three possible method to calculate U^2 (U_d^2) and its p-value for the VM fit.

8.3.1 von Mises fit

- Method 1: ML estimates are calculated assuming observations are continuous, and U^2 and its asymptotic percentage points are calculated assuming observa-

tions are continuous as well.

- Method 2: ML estimates are calculated assuming observations are grouped into 360 cells, and U^2 and its asymptotic percentage points are calculated assuming observations are grouped as well.

- Method 3: ML estimates are calculated assuming observations are continuous, but U_d^2 and its asymptotic percentage points are calculated assuming observations are grouped.

Table 8.4 shows that all three methods gave similar results, but Method 3 was the easiest computationally and it has been adapted for all other distributions considered in this example.

The parametric bootstrap method, discussed in subsection 7.2, was used as an alternative method to find percentage points for the distribution of U^2. Table 8.5 shows that (1) the Monte Carlo values converged rapidly to the asymptotic values, so the latter could be used to obtained the p-value, and this avoided a lengthy Monte Carlo study, (2) when the number of cells k was large, the asymptotic points given by grouped data matched those given by continuous data. As suggested in the introduction, the grouped data analysis could be used even for apparently continuous data since the grouped analysis was usually easier to preform than the continuous analysis.

8.3.2 Wrapped Normal fit

When the WN distribution was fitted to these data, the ML estimates (and standard deviation) were $\hat{\mu}$ = 2.178(0.209), $\hat{\sigma^2}$ = 2.05(0.137) and gave an estimate of the mean resultant vector ρ = 0.358. The Watson statistic was U_d^2 = 0.137, which when compared with its asymptotic values at 10% = 0.080, and at 1% = 0.133, yielded a

Table 8.4: The ML estimates, Watson's U^2 statistics, percentage points, and p-value from the three different methods for tests under the von Mises distributions in Example 8.3.

	μ	κ	U^2	25%	15%	10%	p-value
Method 1	2.134	0.9112	0.052	0.048	0.057	0.048	≈ 0.22
Method 2	2.154	0.932	0.053	0.0479	0.058	0.065	0.232
Method 3	2.134	0.9112	0.050	0.0479	0.058	0.065	0.229

Table 8.5: Percentage points of the U^2 statistic from the von Mises distributions with $\kappa = 0.911$ and $\mu = 2.134$.

	n/α	0.50	0.25	0.10	0.05	0.025	0.01	0.005
M.C.	20	0.0362	0.0512	0.0724	0.0861	0.0989	0.1181	0.1440
M.C.	50	0.0349	0.0483	0.0636	0.0791	0.0972	0.1122	0.1166
M.C.	100	0.0359	0.0493	0.0652	0.0793	0.0936	0.1112	0.1185
Grouped	k=360	0.0340	0.0479	0.0650	0.0775	0.0898	0.1057	0.1176

p-value of 0.008. The WN fit was thus rejected. Note that the parameters for the VM distribution and for the WN distribution followed the relationship $\kappa = \frac{-\sigma^2}{2}$ given by Stephens (1963).

8.3.3 Wrapped Cauchy fit

When the WC distribution was fitted to these data, using $\bar{\theta} = 2.134$ and $-\log(0.415)$ as starting values for μ and ρ, the ML estimates (and standard deviation) were $\hat{\mu} = 2.113(0.127)$, $\hat{\rho} = 0.828(0.130)$ and gave an estimate of the mean resultant vector $e^{-\rho} = 0.437$. The asymptotic correlation between the two parameters was 0.024. The Watson statistic was $U_d^2 = 0.014$, which when compared with its asymptotic values at 50% = 0.029, at 10% = 0.052, and at 1% = 0.081, yielded a p-value of 0.994. The WC distribution appeared to be a good fit.

8.3.4 Wrapped t fit

When the 3 parameter WT distribution was fitted to these data, using $\bar{\theta} = 2.134$, 0.9, and 0.8 (from the contour plot of the profile log-likelihood) as starting values for μ, λ, and ν, the ML estimates (and standard deviation) were $\hat{\mu} = 2.111(0.124)$, $\hat{\nu} = 0.862(0.335)$, and $\hat{\lambda} = 0.789(0.165)$, and the asymptotic correlation between $\hat{\mu}$ and the other two estimates were positive and were quite small. The asymptotic correlation between $\hat{\nu}$ and $\hat{\lambda}$ was 0.6151. The Watson statistic was $U_d^2 = 0.951$, which when compared with its asymptotic values at 50% = 0.022, at 10% = 0.042, and 1% = 0.081, yielded a p-value of 0.951. The WT distribution thus fit these data well. Since ν (d.f.) was less than 1, the distribution has a heavier tailer than the WC distribution. The log-likelihood value of the WT fit was −170.02, slightly bigger than the value of the WC fit(−170.038). It suggested a similar fit with the WC distribution, but the WC distribution has 1 less parameter. The AIC values for the WC and the

WT distributions were: 344.08 and 346.035, while the BIC values were 349.37 and 353.968. These both favoured the WC fit over the WT fit.

8.3.5 Wrapped stable fit

When the 4 parameter WS distribution was fitted to these data, using the estimates $\hat{\mu}$ = 2.113, $\hat{\rho}$ = 0.828 (from the WC estimates), 0.9, and 1.5 (from the contour plot of the profile log-likelihood) as starting values for μ, γ, α, and β respectively, the ML estimates (standard deviation) were $\hat{\mu}$ = 2.124(0.131), $\hat{\gamma}$ = 0.845(0.157), $\hat{\alpha}$ = 0.857(0.221), and $\hat{\beta}$ = 0.139(0.353). The largest asymptotic correlation among the ML estiamtes was that between $\hat{\mu}$ and $\hat{\beta}$ (0.3892). The Watson statistic was U_d^2 = 0.012, which when compared with its asymptotic values at 50% = 0.018, at 10% = 0.030, and at 5% = 0.034, yielded a p-value of 0.923. The WS distribution thus fit these data well. The log-likelihood value of WS fit was −169.80, slightly larger than the value of the WC fit (−170.038) and the WT fit (-170.02). It suggested a similar fit with the WC distribution and the WT distribution, but the WC distribution has 2 less parameter. The AIC values for the WC and the WS distributions were: 344.08 and 347.599, while the BIC values were 349.37 and 358.18. These both favoured the WC distribution over the WS distribution.

8.3.6 Wrapped exponential fit

When the 2 parameter WE distribution was fitted to these data, the ML estimates(and standard deviation) were $\hat{\lambda}$ = 0.335(0.0617) and $\hat{\mu}_0$ = 1.075(0.0286) and the asymptotic correlation was small. The Watson Statistic was U_d^2 = 0.1326, which when compared with its asymptotic values, yielded a p-value of 0.0756. The WE fit was not rejected at 5%.

8.3.7 Batschelet fit

When the 2 parameter Batschelet's distribution was fitted to these data, the ML estimates (and standard deviation) were $\hat{\kappa} = 0.719(0.111)$ and $\hat{\nu} = -0.750(0.256)$, and the asymptotic correlation between $\hat{\kappa}$ and $\hat{\nu}$ was -0.2710. The Watson statistic was $U_d^2 = 0.123$, which when compared with asymptotic values at 50% = 0.049, at 10% = 0.112, and at 5% = 0.140, yielded a p-value of 0.076. The Batschelet's distribution was not rejected at 5%.

8.3.8 Wrapped skew-normal fit

When the 3 parameter WSN distribution was fitted to these data, the ML estimates were $\hat{\xi} = 2.189$, $\hat{\eta} = 1.433$, and $\hat{\lambda} = -0.0099$. The Watson statistic was $U_d^2 = 0.134$, which when compared with asymptotic values at 50% = 0.031, at 10% = 0.062, and at 5% = 0.0753, yielded a p-value of 0.002. The WSN fit was rejected. Note that the log-likelihood value from the WSN fit was the same as the WN fit. This was not surprising because λ was very close to 0 and the WSN distribution degenerated to a WN distribution($\xi = 2.189$ compared to $\mu = 2.179$ and $\eta^2 = 2.053$ compared to $\sigma^2 = 2.053$).

8.3.9 Wrapped skew-Laplace fit

When the 3 parameter WSL distribution was fitted to these data, the ML estimates (and standard deviation) were $\hat{\lambda}_1 = 0.744(0.179)$, $\hat{\lambda}_2 = 0.987(0.277)$, and $\hat{\mu}_0 = 2.030(0.121)$. The asymptotic correlation between $\hat{\lambda}_1$ and $\hat{\lambda}_2$ was -0.5244, between $\hat{\lambda}_1$ and $\hat{\mu}_0$ was 0.5079 and between $\hat{\lambda_2}$ and $\hat{\mu_0}$ was -0.5669. The Watson statistic was $U_d^2 = 0.020$, which when compared with asymptotic values at 50% = 0.028, and at 10% = 0.052, yielded a p-value of 0.772. The WSL distribution thus fit these data

quite well.

8.3.10 Wrapped normal Laplace fit

When the 4 parameter WNL distribution was fitted to these data, the ML estimates (and standard deviation) were $\hat{a}$ = 1.345(0.314), $\hat{b}$ = 1.011(0.276), $\hat{\eta}$ = 2.028(0.109), and $\hat{\tau^2}$ = 0.000. We used $\bar{\theta}$ = 2.134 and $\bar{R}$ = 0.415 as starting values for μ and ρ and used (3.25) and (3.26) to solve for starting values of τ and η and used the inverse of MLE's from WSL $\frac{1}{\lambda_1}$ and $\frac{1}{\lambda_2}$ as starting values for a and b.

The Watson's statistic was U_d^2 = 0.020, which when compared with asymptotic values at 50% = 0.0275, and at 10% = 0.051, yielded a p-value of 0.754. The WNL distribution thus fit these data quite well.

The log-likelihood value from the WNL fit was −170.46, which was very similar to the log-likelihood value from the WSL fit (-170.47). The AIC values for the WSL and the WNL distributions were: 346.94 and 348.92, while the BIC values were: 354.87 and 359.50. These both favoured the WSL fit over the WNL fit.

This was because the WNL distribution has one more parameter than the WSL distribution, but the ML estimate of this parameter, $\hat{\tau}^2$, was very small, recall that when $\hat{\tau}^2$ = 0, the WNL distribution becomes a WSL distribution with $\lambda_1 = \frac{1}{a}$ ($\frac{1}{1.345}$ = 0.743) and $\lambda_2 = \frac{1}{b}$ ($\frac{1}{1.011}$ = 0.989).

Note that $\hat{\tau}^2$ happened to be on the boundary of its parameter space, therefore it may not have the desired asymptotic properties of ML estimates anymore (e.g., $\hat{\tau}^2$ may not be a sufficient estimate). Furthermore, if the estimates were not sufficient estimates, the asymptotic theory for U^2 will not hold. Therefore, the asymptotic standard deviations of $\hat{\tau}^2$ were not calculated.

8.3.11 Wrapped Generalized normal Laplace fit

When the 5 parameter WGNL distribution was fitted to these data, using the MLE's from the WNL and ζ = 1 as the starting values, the parameter estimates were $\hat{a}$ = 1.828, $\hat{b}$ = 0.294, $\hat{\eta}$ = 1.829, $\hat{\tau}^2$ = 0.000, and $\hat{\zeta}$ = 1.128. The Watson's statistic was U_d^2 = 0.027, which when compared with asymptotic values at 10% = 0.03, at 5% = 0.035 and at 1% = 0.047, yielded a p-value of 0.149. The WGNL distribution fit these data reasonably well. Like the WNL fit, $\hat{\tau}^2$ was very close to 0, the boundary of its parameter space, and the asymptotic standard deviations of the ML estimates of the WGNL distribution were thus not calculated.

Table 8.6: Comparison of the log-likelihood values (l), the U_d^2 p-value, AIC, and BIC for some circular distributions.

	VM	WN	WC	WT	WSN	Bat
l	-172.39	-175.26	-170.04	-170.02	-175.26	-176.71
U_d^2 p-value	0.235	0.008	0.994	0.951	0.002	0.076
AIC	348.78	354.52	344.08	346.03	356.52	357.41
BIC	354.07	359.81	349.37	353.97	364.44	362.70
	WE	WS	WSL	WNL	WGNL	
l	-174.54	-169.80	-170.47	-170.46	-169.97	
U_d^2 p-value	0.133	0.923	0.772	0.754	0.149	
AIC	353.08	347.60	346.94	348.92	349.95	
BIC	358.37	358.18	354.87	359.50	363.17	

8.3.12 Likelihood ratio tests

Between the WGNL fit and the WNL fit, the likelihood ratio statistic (LRS) was 0.49 with 1 d.f. (p-value=0.322); between the WGNL fit and the WSL fit, the LRS was 1 with 1 d.f. (p-value=0.317); and between the WGNL fit and the WN fit, it was 10.58 with 3 d.f. (p-value=0.0011). The WGNL, the WNL and the WSL distributions fit these data almost equally well. In conclusion, all the distributions fitted except

Table 8.7: Model rankings: on the log-likelihood values (l), U_d^2 p-value, AIC, and BIC.

	WC	WT	WS	WSL	WNL	VM	WGNL	WE	Bat	WN	WSN
p-value	1	2	3	4	5	6	7	8	9	10	11
AIC	1	2	4	3	6	5	7	8	11	9	10
BIC	1	2	5	4	7	3	10	6	9	8	11

the WN distribution, and the WSN distribution were not rejected by U_d^2 at 5% level (Table 8.6). Among the distributions, the WC fit provided the best AIC, BIC, and U_d^2 p-value while the WS fit had the largest maximum likelihood value. The Watson's test suggested both the VM fit and the WT fit these data well and this was consistent with the Q-Q plots (Pewsey 2007) . In this example, the WC distribution seemed unambiguously the best model: it ranked highest on AIC and BIC and the Watson's test suggested a very good fit (Table 8.7).

8.4 Example 4: Bird data (continuous ML estimates, grouped U_d^2

The directions of 1827 flights of autumnal migratory birds were analyzed. Like the sandstone data in Example 3, they are grouped data with cell width 1 degree. However, it was analyzed as an example of continuous observations. A histogram of the data is shown in Figure 2.4. A number of circular distributions discussed in Chapter 3 were fitted to the dataset as following steps:

1. Find the ML estiamtes for unknown parameters, as described in Section 4.2.

2. Calculate AIC, BIC, and the likelihood ratio statistic (LRS) from the estimated log-likelihood value if a distribution is nested under WGNL distribution.

3. Treat these data as grouped with 360 cells, and compute U_d^2 (Section 5.3) and its asymptotic percentage points (Section 6.2) using the ML estimates obtained from step 1.

4. Calculate p-values for U_d^2 and for the LRS.

8.4.1 Wrapped normal Laplace fit

When the WNL distribution was fitted to the bird data, the ML estimates (and standard deviation) were $\hat{a}$ = 0.369(0.032), $\hat{b}$ = 0.696(0.033), $\hat{\eta}$ = 4.167(0.039), and $\hat{\tau}^2$ = 0.097(0.021) and the asymptotic correlations between the parameters were small. The maximized log-likelihood value was -2137.08 (AIC=4282.16 and BIC=4304.202). The Watson's statistic was U_d^2 = 0.113, which when compared with asymptotic values in Table 8.8, yielded a p-value of 0.0006. The WNL fit was thus rejected.

From 1000 Monte Carlo samples, the first moment (and bootstrap standard deviation) of $\hat{\alpha}$ was 0.365(0.034), of $\hat{\beta}$ was 0.694(0.030), of $\hat{\eta}$ was 4.170(0.037), and of $\hat{\tau}^2$ was 0.098(0.025). They matched the ML estimates and the asymptotic standard deviations of the ML estimates very well.

Table 8.9 shows the 95% bootstrap confidence intervals of the estimates and of goodness-of-fit measures l, AIC, and BIC. The estimated l, AIC, and BIC were within the 95% bootstrap confidence intervals, therefore the three statistics were not significant, while U_d^2 rejected the WNL fit (see Table 8.8).

Table 8.8: Grouped (k=360 cells) asymptotic and Monte Carlo (M.C.) percentage points of the U_d^2 statistic from the wrapped Normal Laplace fit with $\hat{a}$ = 0.365, $\hat{b}$ = 0.694, $\hat{\eta}$ = 4.167, and $\hat{\tau}^2$ = 0.097 (α is the level of significance).

α	0.50	0.10	0.05	0.025	0.01
Group (k = 360)	0.0271	0.0494	0.0583	0.0670	0.0783
M.C. (mc = 1000)	0.0280	0.0488	0.0595	0.0665	0.0821

8.4.2 Wrapped generalized normal Laplace fit

When the WGNL distribution was fitted to the bird data, the ML estimates were $\hat{a}$ = 0.000, $\hat{b}$ = 3.429, $\hat{\eta}$ = −19.390, $\hat{\xi}$ = 0.1142, and $\hat{\tau}^2$ = 2.852 and the maximized log-likelihood value was -2129.943 (AIC=4269.886 and BIC=4297.438). The Watson's

Table 8.9: The 95% bootstrap confidence intervals of the ML estimates, maximized log-likelihood (l), AIC, and BIC from the wrapped Normal Laplace fit.

	a	b	η	τ^2	l	AIC	BIC
Upper Limit.	0.432	0.752	4.242	0.148	-2063.57	4418.00	4440.04
Lower Limit	0.299	0.636	4.101	0.049	-2205.72	4133.30	4155.34

statistic was U_d^2 = 0.112, which when compared with asymptotic values, yielded a p-value of 0.023. The WGNL fit was thus rejected.

Note that $\hat{a}$ happened to be on the boundary of its parameter space, and it may not be a sufficient estimate. Consequently the U_d^2 asymptotic theory may not be valid. Therefore, the asymptotic standard deviations for the ML estimates were not calculated.

Table 8.10 and 8.11 show that the effects on the U_d^2 statistic and its p-value while fixing parameters and jittering a or b around $\hat{a}$ or $\hat{b}$. Table 8.11 shows that the ML estimate for $\hat{b}$, gave the best fit based on U_d^2, because $\hat{b}$ was not on the boundary of the parameter space. However, Table 8.11 shows that the $\hat{a}$=0.3 gave the better fit based on U_d^2, since the MLE, $\hat{a}$, lay on the boundary of its parameter space; This motivates the idea of minimum U_d^2 estimation: find a set of values that minimize the U_d^2 statistic. Table 8.10 shows that with $\hat{a}$ = 0.3 and other estimates fixed, the Watson's statistic was U_d^2 = 0.057, which when compared with asymptotic values, yielded a p-value 0.117. The WGNL fit was not significant at 10% level. Further investigations are needed to understand the robustness of U_d^2, see Chapter 9.

8.4.3 Wrapped stable fit

When the WS distribution was fitted to the bird data, the ML estimates (and standard deviations) were $\hat{\alpha}$ = 1.542(0.042), $\hat{\beta}$ = −0.641(0.092), $\hat{\gamma}$ = 0.461(0.011), and $\hat{\mu}$ = 4.010(0.020) and the maximized log-likelihood was -2127.733 (AIC=4263.466 and

Table 8.10: For the wrapped generalized Normal Laplace distribution: the value of U_d^2 statistic and its p-value with different $\hat{a}$ while fixing $\hat{b}$ = 3.429, $\hat{\eta}$ = −19.390 and $\hat{\xi}$ = 0.1142 τ^2 = 2.852.

$\hat{a}$	0.0000	0.1	0.2	0.3	0.4	0.5
l	-2129.943	-2130.201	-2130.973	-2132.195	-2133.828	-2135.856
U_d^2	0.112	0.077	0.058	0.057	0.071	0.097
p-value	0.023	0.030	0.107	0.117	0.048	0.009

Table 8.11: For the wrapped generalized Normal Laplace distribution: the value of the U_d^2 statistic and its p-value with different $\hat{b}$ while fixing $\hat{a}$ = 0.0000, $\hat{\eta}$ = −19.390 and $\hat{\xi}$ = 0.1142 τ^2 = 2.852.

$\hat{b}$	3.429	3.6	3.5	3.4	3.3
l	-2129.943	-2130.030	-2129.959	-2129.946	2129.999
U_d^2	0.1125	0.1123	0.1120	0.1128	0.1146
p-value	0.0232	0.0090	0.0070	0.0043	0.0077

BIC=4285.508). The Watson's statistic was U_d^2 = 0.061, which when compared with asymptotic values in Table 8.13, yielded a p-value of 0.079. The WS fit was thus not significant at 5% level.

From 1000 Monte Carlo samples, the first moment (and bootstrap standard deviation) of $\hat{\alpha}$ was 1.541(0.045), of $\hat{\beta}$ was -0.640 (0.093) , of $\hat{\gamma}$ was 0.459 (0.011) and of $\hat{\mu}$ was 4.269 (0.020); they matched the ML estimates well.

Table 8.12 shows the 95% bootstrap confidence intervals of the estimates and of the goodness-of-fit measures l, AIC and BIC. The estimated l, AIC and BIC were within the 95% bootstrap confidence intervals, therefore the three statistics were not significant.

Table 8.12: The 95% bootstrap confidence intervals of the ML estimates, maximized log-likelihood (l), AIC, and BIC from the wrapped Stable fit.

	α	β	μ	γ	l	AIC	BIC
Upper Limit.	1.626	-0.455	4.307	0.482	-2039.78	4375.98	4398.02
Lower Limit	1.451	-0.813	4.229	0.440	-2185.04	4087.33	4109.38

Table 8.13: Grouped (k=360 cells) asymptotic and Monte Carlo (M.C.) percentage points of U_d^2 from the wrapped Stable fit with $\hat{\alpha}$ = 1.542, $\hat{\beta}$ = −0.641, $\hat{\gamma}$ = 0.461, and $\hat{\mu}$ = 4.010 (α is the level of significance).

α	0.50	0.10	0.05	0.025	0.01
Group (k=360)	0.0303	0.0571	0.0678	0.0783	0.0919
M.C. (1000)	0.0305	0.0559	0.0695	0.0766	0.0931

8.4.4 Wrapped skew normal fit

When the WSN distribution was fitted to the bird data, the ML estimates (and standard deviations) were $\hat{\eta}$ = 1.210(0.034), $\hat{\lambda}$ = −2.216(0.186), and $\hat{\xi}$ = 4.704(0.032) and the estimated log-likelihood was -2202.056 (AIC=4410.112 and BIC=4426.643). The Watson's statistic was U_d^2 = 1.170, which when compared with asymptotic values in Table 8.15, yielded a p-value of 0.000. The WSN fit was thus rejected.

From 1000 Monte Carlo samples, the first moment (the bootstrap standard deviation) of $\hat{\eta}$ was 1.210(0.036), of $\hat{\lambda}$ was -2.226(0.203), and of $\hat{\xi}$ was 4.704(0.039); they matched the ML estimates well.

Table 8.14: The 95% bootstrap confidence intervals of the ML estimates, maximized log-likelihood (l), AIC, and BIC from the wrapped skewed Normal fit.

	η	λ	ξ	l	AIC	BIC
Upper Limit.	1.284	-1.645	4.879	-2153.53	4550.04	4566.57
Lower Limit	1.141	-2.768	4.674	-2272.28	4312.91	4329.44

Table 8.15: Grouped (k=360 cells) asymptotic and Monte Carlo (M.C.) percentage points of U_d^2 from the wrapped skewed Normal fit with $\hat{\eta}$ = 1.210, $\hat{\lambda}$ = −2.216, and $\hat{\xi}$ = 4.704 (α is the level of significance).

α	0.50	0.10	0.05	0.025	0.01
Group (k=360)	0.0390	0.0783	0.0945	0.1106	0.1317
M.C. (1000)	0.0396	0.0772	0.0936	0.1096	0.1372

8.4.5 Wrapped exponential fit

When the WE distribution was fitted to the bird data, the ML estimates (and standard deviations) were $\hat{\lambda}$ = 0.0494(0.0023) and $\hat{\mu}_0$ = 2.863(0.0054). The Watson's statistic yielded a p-value of 0.000, when compared with asymptotic values. The WE fit was thus rejected.

8.4.6 Wrapped skewed Laplace fit

When the WSL distribution was fitted to the bird data, the ML estimates (and standard deviations) were $\hat{\lambda}_1$ = 2.206(0.014), $\hat{\lambda}_2$ = 1.305(0.0026), and $\hat{\mu}_0$ = 4.158(0.033). The Watson's statistic was U_d^2 = 0.464, which when compared with asymptotic values, yielded a p-value of 0.000. The WSL fit was thus rejected.

8.4.7 Batchelet fit

When the Batschelet's distribution was fitted to the bird data, the ML estimates (and standard deviations) were $\hat{\kappa}$ = −0.909(0.015) and $\hat{\nu}$ = −0.524(0.057). The Watson's statistic was U_d^2 = 1.610, which when compared with asymptotic values, yielded a p-value of 0.000. The Batchelet fit was thus rejected.

8.4.8 Circular beta fit

When the CB distribution was fitted to the bird data, the ML estimates (and standard deviations) were $\hat{\alpha}$ = 1.813(0.073), $\hat{\beta}$ = 0.339(0.009) and $\hat{\eta}$ = 3.875(0.0002) and the asymptotic correlation between $\hat{\alpha}$ and $\hat{\beta}$ was 0.516, and the others were very small .

The Watson's statistics was U_d^2 = 2.556, which when compared with asymptotic values at 0.094(10%), at 0.117(5%) and at 0.171(1%), yielded a p-value of 0.000. The CB fit was thus rejected.

8.4.9 Wrapped cauchy fit

When the WC distribution was fitted to the bird data, the ML estimates (and standard deviations) were $\hat{\mu}$ = 3.978(0.017) and $\hat{\sigma}$ = 0.450(0.014). The Watson's statistic was U_d^2 = 1.008, which when compared with asymptotic values, yielded a p-value of 0.000. The WC fit was thus rejected.

8.4.10 Wrapped normal fit

When the WN distribution was fitted to the bird data, the ML estimates (and standard deviations) were $\hat{\mu}$ = 3.8898(0.020), and $\hat{\sigma}^2$ = 0.688(0.023). The Watson's statistic was U_d^2 = 2.124, which when compared with asymptotic values, yielded a p-value of 0.000. The WN fit was thus rejected.

8.4.11 Wrapped t fit

When the WT distribution was fitted to the bird data, the ML estimates (and standard deviations) were $\hat{\mu}$ = 3.943(0.017), $\hat{\epsilon}$ = 3.307(0.358) ,and $\hat{\lambda}$ = 0.5945(0.019). The Watson's statistic was U_d^2 = 0.325, which when compared with asymptotic values, yielded a p-value of 0.000. The WT fit was thus rejected.

8.4.12 vonMise fit

When the VM distribution was fitted to the bird data, the ML estimates (and standard deviations) were $\hat{\mu}$ = 3.8898(0.020), and $\hat{\kappa}$ = 0.688(0.023). The Watson's statistic was U_d^2 = 0.550, which when compared with asymptotic values, yielded a p-value of 0.000. The VM fit was thus rejected.

Table 8.16: Comparison of the log-likelihood(l), U_d^2 p-value, AIC, and, BIC for some circular distributions fit to the bird data.

	VM	WE	WN	WC	WT	BAT	CB
l	-2162.57	-2688.44	-2243.12	-2233.21	-2150.373	-2795.48	-2258.35
U_d^2 p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AIC	4329.14	5380.88	4490.24	4470.42	4306.75	5594.96	4522.69
BIC	4340.17	5391.90	4501.26	4481.441	4323.28	5605.98	4539.22
	WSN	WSL	WS	WNL	WGNL	WGNL*	
l	-2202.06	-2154.50	-2127.73	-2137.08	-2129.94	-2130.97	
U_d^2 p-value	0.0000	0.000	0.079	0.001	0.023	0.107	
AIC	4410.11	4314.99	4263.47	4282.16	4269.89	4271.94	
BIC	4426.64	4331.53	4285.51	4304.20	4297.44	4299.49	

Table 8.17: Model rankings: log-likelihood, U_d^2 p-value, AIC, and BIC.

	WS	WGNL	WGNL*	WNL	WT	WSL	VM
AIC	1	2	3	4	5	6	7
BIC	1	2	3	4	5	6	7
p-value	2	3	1	4	**	**	**
	WSN	WC	WN	CB	WE	BAT	
AIC	8	9	10	11	12	13	
BIC	8	9	10	11	12	13	
U_d^2 p-value	**	**	**	**	**	**	

8.4.13 Likelihood ratio tests

Between the WGNL fit and the WNL fit, the likelihood ratio statistic was 14.28 with 1 d.f.; between the WNL fit and the WSL fit, the LRS was 34 with 1 d.f.; and between the WGNL fit and the WN fit, the LRS was 226.36 with 3 d.f.. The WGNL fit was clearly better than all its nested models and the WNL fit was better than the WSL fit.

Note that in Table 8.16, WGNL* is the model where $\hat{a}$ = 0.2 and WGNL is the model where $\hat{a}$ = 0.0.

In conclusion, for the bird data, all the distributions considered except the wrapped stable distribution were rejected by Watson's statistic at the 5% significance level. This was not surprising because the sample size was large (n=1827). As the sample size gets sufficiently large, any goodness-of-fit test will tend to reject the null hypothesis. For these data it seemed that the WS distribution unambiguously provided best fit. Not only did it have the smallest AIC and BIC values, it was the only distribution for which the Watson's statistic was not significant. Of the symmetric distributions, the WT distribution provided best fit.

By jittering a around the MLE $\hat{a}$ results in a WGNL fit that was not significant at the 10% level.

Chapter 9

Conclusion

We studied the use of Watson's U^2 statistic for assessing the goodness-of-fit of parametric models with several parameters for both grouped and ungrouped (continuous) circular data, while the parameters are estimated by using maximum likelihood estimation. Further, we compared the use of U^2 with the results of model selection criteria (AIC, BIC) on a number of real data sets. For grouped data, the distribution of U_d^2 statistic for several circular parametric models was obtained via one of two methods: (1) U_d^2 asymptotic theory for grouped data; (2) the parametric bootstrap technique for grouped data. For ungrouped data, the distributions of the statistic were obtained via one of the three methods: (1) U^2 asymptotic theory for ungrouped data,(2) the parametric bootstrap technique for ungrouped data, (3) U_d^2 asymptotic theory for grouped data while assuming the data are grouped into sufficient large number of cells. Due to the fact that the U^2 asymptotic distributions for many circular distributions are difficult to obtain when data is ungrouped, and to the fact that the parametric bootstrap requires extensive computing time, we suggest that for both grouped and ungrouped data, one should always use asymptotic theory for grouped data to obtain the distribution of U_d^2. When the data were treated as grouped, asymp-

totic percentage points for U^2 match well with the percentage points when the data are treated as continuous for von Mises (Table 6.1) and wrapped exponential (Table 6.2) distributions, and they also match well in all cases with the percentage points obtained by parametric bootstrapping the percentage points.

We studied four examples. Example 1 (ants data) provides a good illustration of how model selection criteria alone do not necessarily provide an adequate model, because even the best fitting model, the wrapped Cauchy, is rejected by U^2 at the 5% significant level. Example 2 shows that for bimodal data, a Circular Beta model can be fitted to analyze the data. While the data sets are grouped in both Example 1 and 2, the data sets are continuous in Example 3 and 4. Example 3 and 4 use asymptotic theory for grouped data to approximate the distribution of U^2 for ungrouped data. We compared the rankings for the models fitted to Example 3 and 4 based on different criteria (AIC, BIC, maximized log-likelihood and U^2 p-value), and showed that they can provide somewhat different rankings.

9.1 Future works

9.1.1 Power studies

The power of the tests for different circular distributions discussed in Chapter 3 have not been studied. Specifically, the U^2 statistic and the likelihood ratio statistic.

9.1.2 Estimation by Minimum U_d^2

Estimation based on minimising U_d^2 can be used as an alternative to MLE. For many cases, when sample sizes are large, the minimum U_d^2 and MLE methods are asymptotically equivalent, as the minimum X^2 estimation method is equivalent to the ML estimation method for continuous data (Berkson 1980). In some cases, the minimum

U_d^2 estimates might give a better fit. For example, when one cell contains a very high number of values, the ML estimates may then give a low expected number; then statistics like Pearsons X^2 or U_d^2 will become large because of this one cell value of $o_i - e_i$. Further, in the cases when the ML estimate lies on the boundary of the parameter space, (hence the MLE loses the desired properties and as a result, the U_d^2 statistic is not computed from sufficient estimators) the minimum U_d^2 estimates can be used to compute U_d^2 instead.

9.1.3 Mixture distributions

U_d^2 asymptotic percentage points for the mixture circular distributions have not been studied yet, for example, the mixture of the Uniform and the wrapped skewed Normal distributions (Pewsey, 2006) and the mixture of the Uniform and the von Mises distributions (Bentley, 2006), the mixture of two vonMises distributions. It would be interesting to compare the fit of these mixture models with the circular models discussed on some data sets.

9.2 Software availability

Functions in R to do all calculations are available from the author.

Appendix A

R code

We attach the R code for the WGNL distributions. One can do the similar calculations by replacing "a.p.fun" and "b.p.fun" functions with corresponding functions from other circular distributions.

```
make.cell<- function(k)
{
#making k cells with equal length around a circle.
xx <- c(seq(2*pi/k,(k-1)*2*pi/k, by=2*pi/k),2*pi)
cells<- c(0,xx)-pi/k
return(cells)
}


count.data<- function(data,cells)
{
k<-length(cells)-1
xxx1<- c(cells,2*pi)
m<- table(cut(data,xxx1))
```

```
m[1]<- m[1]+m[k+1]
m<- m[-(k+1)]
return(m)
}

################################################
# ML estimates for WGNL distributions when data is grouped.
fr_wgnl <- function(est,cells,m)
{
#cells= cell boundaries generated from "makecell" function.
#number of counts in each cell generated from "count.data" function.

a.p.fun<- function(p, a, b, eta, sigma2,xi)
{
tem<- atan2((a-b)*p, 1+a*b*p^2)%%(2*pi)
(exp((-sigma2*p^2)/2)/((1+a^2*p^2)*(1+b^2*p^2))^0.5)^xi*
(cos(xi*(eta*p+tem)))
}

b.p.fun<- function(p, a, b, eta, sigma2,xi)
{
tem<- atan2((a-b)*p, 1+a*b*p^2)%%(2*pi)
(exp(-sigma2*p^2/2)/((1+a^2*p^2)*(1+b^2*p^2))^0.5)^xi*
(sin(xi*(eta*p+tem)))
}
```

```
pdf.wgnl<- function(theta)
{
temp<-
a.p.fun(1,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*cos(1*theta)+
b.p.fun(1,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*sin(1*theta)+
a.p.fun(2,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*cos(2*theta)+
b.p.fun(2,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*sin(2*theta)+
a.p.fun(3,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*cos(3*theta)+
b.p.fun(3,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*sin(3*theta)+
a.p.fun(4,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*cos(4*theta)+
b.p.fun(4,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*sin(4*theta)+
a.p.fun(5,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*cos(5*theta)+
b.p.fun(5,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*sin(5*theta)+
a.p.fun(6,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*cos(6*theta)+
b.p.fun(6,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*sin(6*theta)+
a.p.fun(7,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*cos(7*theta)+
b.p.fun(7,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*sin(7*theta)+
a.p.fun(8,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*cos(8*theta)+
b.p.fun(8,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*sin(8*theta)+
a.p.fun(9,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*cos(9*theta)+
b.p.fun(9,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*sin(9*theta)+
a.p.fun(10,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*cos(10*theta)+
b.p.fun(10,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*sin(10*theta)+
a.p.fun(11,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*cos(11*theta)+
b.p.fun(11,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*sin(11*theta)+
```

```
a.p.fun(12,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*cos(12*theta)+
b.p.fun(12,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*sin(12*theta)+
a.p.fun(13,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*cos(13*theta)+
b.p.fun(13,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*sin(13*theta)+
a.p.fun(14,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*cos(14*theta)+
b.p.fun(14,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*sin(14*theta)+
a.p.fun(15,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*cos(15*theta)+
b.p.fun(15,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*sin(15*theta)+
a.p.fun(16,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*cos(16*theta)+
b.p.fun(16,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*sin(16*theta)+
a.p.fun(17,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*cos(17*theta)+
b.p.fun(17,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*sin(17*theta)+
a.p.fun(18,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*cos(18*theta)+
b.p.fun(18,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*sin(18*theta)+
a.p.fun(19,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*cos(19*theta)+
b.p.fun(19,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*sin(19*theta)+
a.p.fun(20,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*cos(20*theta)+
b.p.fun(20,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*sin(20*theta)
#Note, users could add in more terms in the following format if higher
#accuracy required:
#a.p.fun(p,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*cos(p*theta)+
#b.p.fun(p,exp(est[1]),exp(est[2]),est[3],exp(est[4]),est[5])*sin(p*theta),
#where p is pth term.

     (1+2*temp)/(2*pi)
}
```

```
p<- rep(0,k)

for (i in 1:k)
{
p[i]<- integrate(pdf.wgnl, lower=xxx[i], upper=xxx[i+1],stop.on.error=FALSE)$
value
}
res<- -sum(m*log(p))
return(res)
}


out.wgnl<- optim(c(log(start.a),  log(start.b ),   start.eta,  start.sigma2,
start.xi), fn= fr_wgnl,  xxx=xxx,m=m, hessian=TRUE,
control = list(trace=TRUE))
#paper suggested an efficient way  to obtain the starting values.


##########################################
##########################################
#Compute grouped asymptotic points for Watson's U square statistic after
#fitting the WGNL distributions.

a.p.fun<- function(p, a, b, eta, sigma2,xi)
{
tem<- atan2((a-b)*p, 1+a*b*p^2)%%(2*pi)
```

```
(exp((-sigma2*p^2)/2)/((1+a^2*p^2)*(1+b^2*p^2))^0.5)^xi*(cos(xi*(eta*p+tem)))
}

b.p.fun<- function(p, a, b, eta, sigma2,xi)
{
tem<- atan2((a-b)*p, 1+a*b*p^2)%%(2*pi)
(exp(-sigma2*p^2/2)/((1+a^2*p^2)*(1+b^2*p^2))^0.5)^xi*(sin(xi*(eta*p+tem)))
}

pdf.wgnl<- function(theta)
{
temp<-
a.p.fun(1,a,b,eta,sigma2,xi)*cos(1*theta)+
b.p.fun(1,a,b,eta,sigma2,xi)*sin(1*theta)+
a.p.fun(2,a,b,eta,sigma2,xi)*cos(2*theta)+
b.p.fun(2,a,b,eta,sigma2,xi)*sin(2*theta)+
a.p.fun(3,a,b,eta,sigma2,xi)*cos(3*theta)+
b.p.fun(3,a,b,eta,sigma2,xi)*sin(3*theta)+
a.p.fun(4,a,b,eta,sigma2,xi)*cos(4*theta)+
b.p.fun(4,a,b,eta,sigma2,xi)*sin(4*theta)+
a.p.fun(5,a,b,eta,sigma2,xi)*cos(5*theta)+
b.p.fun(5,a,b,eta,sigma2,xi)*sin(5*theta)+
a.p.fun(6,a,b,eta,sigma2,xi)*cos(6*theta)+
b.p.fun(6,a,b,eta,sigma2,xi)*sin(6*theta)+
a.p.fun(7,a,b,eta,sigma2,xi)*cos(7*theta)+
b.p.fun(7,a,b,eta,sigma2,xi)*sin(7*theta)+
```

```
a.p.fun(8,a,b,eta,sigma2,xi)*cos(8*theta)+
b.p.fun(8,a,b,eta,sigma2,xi)*sin(8*theta)+
a.p.fun(9,a,b,eta,sigma2,xi)*cos(9*theta)+
b.p.fun(9,a,b,eta,sigma2,xi)*sin(9*theta)+
a.p.fun(10,a,b,eta,sigma2,xi)*cos(10*theta)+
b.p.fun(10,a,b,eta,sigma2,xi)*sin(10*theta)+
a.p.fun(11,a,b,eta,sigma2,xi)*cos(11*theta)+
b.p.fun(11,a,b,eta,sigma2,xi)*sin(11*theta)+
a.p.fun(12,a,b,eta,sigma2,xi)*cos(12*theta)+
b.p.fun(12,a,b,eta,sigma2,xi)*sin(12*theta)+
a.p.fun(13,a,b,eta,sigma2,xi)*cos(13*theta)+
b.p.fun(13,a,b,eta,sigma2,xi)*sin(13*theta)+
a.p.fun(14,a,b,eta,sigma2,xi)*cos(14*theta)+
b.p.fun(14,a,b,eta,sigma2,xi)*sin(14*theta)+
a.p.fun(15,a,b,eta,sigma2,xi)*cos(15*theta)+
b.p.fun(15,a,b,eta,sigma2,xi)*sin(15*theta)+
a.p.fun(16,a,b,eta,sigma2,xi)*cos(16*theta)+
b.p.fun(16,a,b,eta,sigma2,xi)*sin(16*theta)+
a.p.fun(17,a,b,eta,sigma2,xi)*cos(17*theta)+
b.p.fun(17,a,b,eta,sigma2,xi)*sin(17*theta)+
a.p.fun(18,a,b,eta,sigma2,xi)*cos(18*theta)+
b.p.fun(18,a,b,eta,sigma2,xi)*sin(18*theta)+
a.p.fun(19,a,b,eta,sigma2,xi)*cos(19*theta)+
b.p.fun(19,a,b,eta,sigma2,xi)*sin(19*theta)+
a.p.fun(20,a,b,eta,sigma2,xi)*cos(20*theta)+
b.p.fun(20,a,b,eta,sigma2,xi)*sin(20*theta)
```

```
(1+2*temp)/(2*pi)
}

derv.a.a.p.fun<- function(p, a, b, eta, sigma2, xi)
{
derv.a.t1(p,a,b,sigma2,xi)*cos(xi*(eta*p-w.fun(p,a,b)))+
t1.fun(p,a,b,sigma2,xi)*sin(xi*(eta*p-w.fun(p,a,b)))*xi*derv.a.w(p,a,b)
}

derv.a.b.p.fun<- function(p, a, b, eta, sigma2, xi)
{
derv.a.t1(p,a,b,sigma2,xi)*sin(xi*(eta*p-w.fun(p,a,b)))-
t1.fun(p,a,b,sigma2,xi)*cos(xi*(eta*p-w.fun(p,a,b)))*xi*derv.a.w(p,a,b)
}


derv.b.a.p.fun<- function(p, a, b, eta, sigma2, xi)
{
derv.b.t1(p,a,b,sigma2,xi)*cos(xi*(eta*p-w.fun(p,a,b)))+
t1.fun(p,a,b,sigma2,xi)*sin(xi*(eta*p-w.fun(p,a,b)))*xi*derv.b.w(p,a,b)
}

derv.b.b.p.fun<- function(p, a, b, eta, sigma2, xi)
{
derv.b.t1(p,a,b,sigma2,xi)*sin(xi*(eta*p-w.fun(p,a,b)))-
```

```
t1.fun(p,a,b,sigma2,xi)*cos(xi*(eta*p-w.fun(p,a,b)))*xi*derv.b.w(p,a,b)
}

derv.sigma2.a.p.fun<- function(p, a, b, eta, sigma2, xi)
{
derv.sigma2.t1(p,a,b,sigma2,xi)*cos(xi*(eta*p-w.fun(p,a,b)))
}

derv.sigma2.b.p.fun<- function(p, a, b, eta, sigma2, xi)
{
derv.sigma2.t1(p,a,b,sigma2,xi)*sin(xi*(eta*p-w.fun(p,a,b)))
}

derv.eta.a.p.fun<- function(p, a, b, eta, sigma2, xi)
{
 -t1.fun(p,a,b,sigma2,xi)*sin(xi*(eta*p-w.fun(p,a,b)))*xi*p
}

derv.eta.b.p.fun<- function(p, a, b, eta, sigma2, xi)
{
 t1.fun(p,a,b,sigma2,xi)*cos(xi*(eta*p-w.fun(p,a,b)))*xi*p
}

derv.xi.a.p.fun<- function(p, a, b, eta, sigma2, xi)
{
derv.xi.t1(p,a,b,sigma2,xi)*cos(xi*(eta*p-w.fun(p,a,b)))-
```

```
t1.fun(p,a,b,sigma2,xi)*sin(xi*(eta*p-w.fun(p,a,b)))*(eta*p-w.fun(p,a,b))
}

derv.xi.b.p.fun<- function(p, a, b, eta, sigma2, xi)
{
 derv.xi.t1(p,a,b,sigma2,xi)*sin(xi*(eta*p-w.fun(p,a,b)))+
 t1.fun(p,a,b,sigma2,xi)*cos(xi*(eta*p-w.fun(p,a,b)))*(eta*p-w.fun(p,a,b))
}

derin.a<- function(theta)
{

temp<-
       derv.a.a.p.fun(1,a,b,eta, sigma2, xi)*cos(1*theta)+
       derv.a.b.p.fun(1,a,b,eta, sigma2, xi)*sin(1*theta)+
       derv.a.a.p.fun(2,a,b,eta, sigma2, xi)*cos(2*theta)+
       derv.a.b.p.fun(2,a,b,eta, sigma2, xi)*sin(2*theta)+
       derv.a.a.p.fun(3,a,b,eta, sigma2, xi)*cos(3*theta)+
       derv.a.b.p.fun(3,a,b,eta, sigma2, xi)*sin(3*theta)+
       derv.a.a.p.fun(4,a,b,eta, sigma2, xi)*cos(4*theta)+
       derv.a.b.p.fun(4,a,b,eta, sigma2, xi)*sin(4*theta)+
       derv.a.a.p.fun(5,a,b,eta, sigma2, xi)*cos(5*theta)+
       derv.a.b.p.fun(5,a,b,eta, sigma2, xi)*sin(5*theta)+
       derv.a.a.p.fun(6,a,b,eta, sigma2, xi)*cos(6*theta)+
       derv.a.b.p.fun(6,a,b,eta, sigma2, xi)*sin(6*theta)+
       derv.a.a.p.fun(7,a,b,eta, sigma2, xi)*cos(7*theta)+
```

```
       derv.a.b.p.fun(7,a,b,eta, sigma2, xi)*sin(7*theta)+
       derv.a.a.p.fun(8,a,b,eta, sigma2, xi)*cos(8*theta)+
       derv.a.b.p.fun(8,a,b,eta, sigma2, xi)*sin(8*theta)+
       derv.a.a.p.fun(9,a,b,eta, sigma2, xi)*cos(9*theta)+
       derv.a.b.p.fun(9,a,b,eta, sigma2, xi)*sin(9*theta)+
       derv.a.a.p.fun(10,a,b,eta, sigma2, xi)*cos(10*theta)+
       derv.a.b.p.fun(10,a,b,eta, sigma2, xi)*sin(10*theta)+
       derv.a.a.p.fun(11,a,b,eta, sigma2, xi)*cos(11*theta)+
       derv.a.b.p.fun(11,a,b,eta, sigma2, xi)*sin(11*theta)+
       derv.a.a.p.fun(12,a,b,eta, sigma2, xi)*cos(12*theta)+
       derv.a.b.p.fun(12,a,b,eta, sigma2, xi)*sin(12*theta)+
       derv.a.a.p.fun(13,a,b,eta, sigma2, xi)*cos(13*theta)+
       derv.a.b.p.fun(13,a,b,eta, sigma2, xi)*sin(13*theta)+
       derv.a.a.p.fun(14,a,b,eta, sigma2, xi)*cos(14*theta)+
       derv.a.b.p.fun(14,a,b,eta, sigma2, xi)*sin(14*theta)+
       derv.a.a.p.fun(15,a,b,eta, sigma2, xi)*cos(15*theta)+
       derv.a.b.p.fun(15,a,b,eta, sigma2, xi)*sin(15*theta)

temp/pi
}



derin.b<- function(theta)
{
temp<-
      derv.b.a.p.fun(1,a,b,eta, sigma2, xi)*cos(1*theta)+
```

```
derv.b.b.p.fun(1,a,b,eta, sigma2, xi)*sin(1*theta)+
derv.b.a.p.fun(2,a,b,eta, sigma2, xi)*cos(2*theta)+
derv.b.b.p.fun(2,a,b,eta, sigma2, xi)*sin(2*theta)+
derv.b.a.p.fun(3,a,b,eta, sigma2, xi)*cos(3*theta)+
derv.b.b.p.fun(3,a,b,eta, sigma2, xi)*sin(3*theta)+
derv.b.a.p.fun(4,a,b,eta, sigma2, xi)*cos(4*theta)+
derv.b.b.p.fun(4,a,b,eta, sigma2, xi)*sin(4*theta)+
derv.b.a.p.fun(5,a,b,eta, sigma2, xi)*cos(5*theta)+
derv.b.b.p.fun(5,a,b,eta, sigma2, xi)*sin(5*theta)+
derv.b.a.p.fun(6,a,b,eta, sigma2, xi)*cos(6*theta)+
derv.b.b.p.fun(6,a,b,eta, sigma2, xi)*sin(6*theta)+
derv.b.a.p.fun(7,a,b,eta, sigma2, xi)*cos(7*theta)+
derv.b.b.p.fun(7,a,b,eta, sigma2, xi)*sin(7*theta)+
derv.b.a.p.fun(8,a,b,eta, sigma2, xi)*cos(8*theta)+
derv.b.b.p.fun(8,a,b,eta, sigma2, xi)*sin(8*theta)+
derv.b.a.p.fun(9,a,b,eta, sigma2, xi)*cos(9*theta)+
derv.b.b.p.fun(9,a,b,eta, sigma2, xi)*sin(9*theta)+
derv.b.a.p.fun(10,a,b,eta, sigma2, xi)*cos(10*theta)+
derv.b.b.p.fun(10,a,b,eta, sigma2, xi)*sin(10*theta)+
derv.b.a.p.fun(11,a,b,eta, sigma2, xi)*cos(11*theta)+
derv.b.b.p.fun(11,a,b,eta, sigma2, xi)*sin(11*theta)+
derv.b.a.p.fun(12,a,b,eta, sigma2, xi)*cos(12*theta)+
derv.b.b.p.fun(12,a,b,eta, sigma2, xi)*sin(12*theta)+
derv.b.a.p.fun(13,a,b,eta, sigma2, xi)*cos(13*theta)+
derv.b.b.p.fun(13,a,b,eta, sigma2, xi)*sin(13*theta)+
derv.b.a.p.fun(14,a,b,eta, sigma2, xi)*cos(14*theta)+
```

```
       derv.b.b.p.fun(14,a,b,eta, sigma2, xi)*sin(14*theta)+
       derv.b.a.p.fun(15,a,b,eta, sigma2, xi)*cos(15*theta)+
       derv.b.b.p.fun(15,a,b,eta, sigma2, xi)*sin(15*theta)

     temp/pi
}



derin.eta<- function(theta)
{

temp<-
     derv.eta.a.p.fun(1,a,b,eta, sigma2, xi)*cos(1*theta)+
      derv.eta.b.p.fun(1,a,b,eta, sigma2, xi)*sin(1*theta)+
      derv.eta.a.p.fun(2,a,b,eta, sigma2, xi)*cos(2*theta)+
      derv.eta.b.p.fun(2,a,b,eta, sigma2, xi)*sin(2*theta)+
      derv.eta.a.p.fun(3,a,b,eta, sigma2, xi)*cos(3*theta)+
      derv.eta.b.p.fun(3,a,b,eta, sigma2, xi)*sin(3*theta)+
      derv.eta.a.p.fun(4,a,b,eta, sigma2, xi)*cos(4*theta)+
      derv.eta.b.p.fun(4,a,b,eta, sigma2, xi)*sin(4*theta)+
      derv.eta.a.p.fun(5,a,b,eta, sigma2, xi)*cos(5*theta)+
      derv.eta.b.p.fun(5,a,b,eta, sigma2, xi)*sin(5*theta)+
      derv.eta.a.p.fun(6,a,b,eta, sigma2, xi)*cos(6*theta)+
      derv.eta.b.p.fun(6,a,b,eta, sigma2, xi)*sin(6*theta)+
      derv.eta.a.p.fun(7,a,b,eta, sigma2, xi)*cos(7*theta)+
      derv.eta.b.p.fun(7,a,b,eta, sigma2, xi)*sin(7*theta)+
```

```
      derv.eta.a.p.fun(8,a,b,eta, sigma2, xi)*cos(8*theta)+
      derv.eta.b.p.fun(8,a,b,eta, sigma2, xi)*sin(8*theta)+
      derv.eta.a.p.fun(9,a,b,eta, sigma2, xi)*cos(9*theta)+
      derv.eta.b.p.fun(9,a,b,eta, sigma2, xi)*sin(9*theta)+
      derv.eta.a.p.fun(10,a,b,eta, sigma2, xi)*cos(10*theta)+
      derv.eta.b.p.fun(10,a,b,eta, sigma2, xi)*sin(10*theta)+
      derv.eta.a.p.fun(11,a,b,eta, sigma2, xi)*cos(11*theta)+
      derv.eta.b.p.fun(11,a,b,eta, sigma2, xi)*sin(11*theta)+
      derv.eta.a.p.fun(12,a,b,eta, sigma2, xi)*cos(12*theta)+
      derv.eta.b.p.fun(12,a,b,eta, sigma2, xi)*sin(12*theta)+
      derv.eta.a.p.fun(13,a,b,eta, sigma2, xi)*cos(13*theta)+
      derv.eta.b.p.fun(13,a,b,eta, sigma2, xi)*sin(13*theta)+
      derv.eta.a.p.fun(14,a,b,eta, sigma2, xi)*cos(14*theta)+
      derv.eta.b.p.fun(14,a,b,eta, sigma2, xi)*sin(14*theta)+
      derv.eta.a.p.fun(15,a,b,eta, sigma2, xi)*cos(15*theta)+
      derv.eta.b.p.fun(15,a,b,eta, sigma2, xi)*sin(15*theta)

      temp/pi
}



derin.sigma2<- function(theta)
{

temp<- derv.sigma2.a.p.fun(1,a,b,eta, sigma2, xi)*cos(1*theta)+
derv.sigma2.b.p.fun(1,a,b,eta, sigma2, xi)*sin(1*theta)+
```

```
derv.sigma2.a.p.fun(2,a,b,eta, sigma2, xi)*cos(2*theta)+
derv.sigma2.b.p.fun(2,a,b,eta, sigma2, xi)*sin(2*theta)+
derv.sigma2.a.p.fun(3,a,b,eta, sigma2, xi)*cos(3*theta)+
derv.sigma2.b.p.fun(3,a,b,eta, sigma2, xi)*sin(3*theta)+
derv.sigma2.a.p.fun(4,a,b,eta, sigma2, xi)*cos(4*theta)+
derv.sigma2.b.p.fun(4,a,b,eta, sigma2, xi)*sin(4*theta)+
derv.sigma2.a.p.fun(5,a,b,eta, sigma2, xi)*cos(5*theta)+
derv.sigma2.b.p.fun(5,a,b,eta, sigma2, xi)*sin(5*theta)+
derv.sigma2.a.p.fun(6,a,b,eta, sigma2, xi)*cos(6*theta)+
derv.sigma2.b.p.fun(6,a,b,eta, sigma2, xi)*sin(6*theta)+
derv.sigma2.a.p.fun(7,a,b,eta, sigma2, xi)*cos(7*theta)+
derv.sigma2.b.p.fun(7,a,b,eta, sigma2, xi)*sin(7*theta)+
derv.sigma2.a.p.fun(8,a,b,eta, sigma2, xi)*cos(8*theta)+
derv.sigma2.b.p.fun(8,a,b,eta, sigma2, xi)*sin(8*theta)+
derv.sigma2.a.p.fun(9,a,b,eta, sigma2, xi)*cos(9*theta)+
derv.sigma2.b.p.fun(9,a,b,eta, sigma2, xi)*sin(9*theta)+
derv.sigma2.a.p.fun(10,a,b,eta, sigma2, xi)*cos(10*theta)+
derv.sigma2.b.p.fun(10,a,b,eta, sigma2, xi)*sin(10*theta)+
derv.sigma2.a.p.fun(11,a,b,eta, sigma2, xi)*cos(11*theta)+
derv.sigma2.b.p.fun(11,a,b,eta, sigma2, xi)*sin(11*theta)+
derv.sigma2.a.p.fun(12,a,b,eta, sigma2, xi)*cos(12*theta)+
derv.sigma2.b.p.fun(12,a,b,eta, sigma2, xi)*sin(12*theta)+
derv.sigma2.a.p.fun(13,a,b,eta, sigma2, xi)*cos(13*theta)+
derv.sigma2.b.p.fun(13,a,b,eta, sigma2, xi)*sin(13*theta)+
derv.sigma2.a.p.fun(14,a,b,eta, sigma2, xi)*cos(14*theta)+
derv.sigma2.b.p.fun(14,a,b,eta, sigma2, xi)*sin(14*theta)+
```

```
      derv.sigma2.a.p.fun(15,a,b,eta, sigma2, xi)*cos(15*theta)+
      derv.sigma2.b.p.fun(15,a,b,eta, sigma2, xi)*sin(15*theta)

temp/pi
}


derin.xi<- function(theta)
{
temp<- derv.xi.a.p.fun(1,a,b,eta, sigma2, xi)*cos(1*theta)+
derv.xi.b.p.fun(1,a,b,eta, sigma2, xi)*sin(1*theta)+
      derv.xi.a.p.fun(2,a,b,eta, sigma2, xi)*cos(2*theta)+
      derv.xi.b.p.fun(2,a,b,eta, sigma2, xi)*sin(2*theta)+
      derv.xi.a.p.fun(3,a,b,eta, sigma2, xi)*cos(3*theta)+
      derv.xi.b.p.fun(3,a,b,eta, sigma2, xi)*sin(3*theta)+
      derv.xi.a.p.fun(4,a,b,eta, sigma2, xi)*cos(4*theta)+
      derv.xi.b.p.fun(4,a,b,eta, sigma2, xi)*sin(4*theta)+
      derv.xi.a.p.fun(5,a,b,eta, sigma2, xi)*cos(5*theta)+
      derv.xi.b.p.fun(5,a,b,eta, sigma2, xi)*sin(5*theta)+
      derv.xi.a.p.fun(6,a,b,eta, sigma2, xi)*cos(6*theta)+
      derv.xi.b.p.fun(6,a,b,eta, sigma2, xi)*sin(6*theta)+
      derv.xi.a.p.fun(7,a,b,eta, sigma2, xi)*cos(7*theta)+
      derv.xi.b.p.fun(7,a,b,eta, sigma2, xi)*sin(7*theta)+
      derv.xi.a.p.fun(8,a,b,eta, sigma2, xi)*cos(8*theta)+
      derv.xi.b.p.fun(8,a,b,eta, sigma2, xi)*sin(8*theta)+
      derv.xi.a.p.fun(9,a,b,eta, sigma2, xi)*cos(9*theta)+
```

```
      derv.xi.b.p.fun(9,a,b,eta, sigma2, xi)*sin(9*theta)+
      derv.xi.a.p.fun(10,a,b,eta, sigma2, xi)*cos(10*theta)+
      derv.xi.b.p.fun(10,a,b,eta, sigma2, xi)*sin(10*theta)+
      derv.xi.a.p.fun(11,a,b,eta, sigma2, xi)*cos(11*theta)+
      derv.xi.b.p.fun(11,a,b,eta, sigma2, xi)*sin(11*theta)+
      derv.xi.a.p.fun(12,a,b,eta, sigma2, xi)*cos(12*theta)+
      derv.xi.b.p.fun(12,a,b,eta, sigma2, xi)*sin(12*theta)+
      derv.xi.a.p.fun(13,a,b,eta, sigma2, xi)*cos(13*theta)+
      derv.xi.b.p.fun(13,a,b,eta, sigma2, xi)*sin(13*theta)+
      derv.xi.a.p.fun(14,a,b,eta, sigma2, xi)*cos(14*theta)+
      derv.xi.b.p.fun(14,a,b,eta, sigma2, xi)*sin(14*theta)+
      derv.xi.a.p.fun(15,a,b,eta, sigma2, xi)*cos(15*theta)+
      derv.xi.b.p.fun(15,a,b,eta, sigma2, xi)*sin(15*theta)

     temp/pi
}



g.asym.points<- function(k, xxx, m)
{
phat1<- numeric(k)
for (i in 1:k)
{
phat1[i]<- integrate(pdf.wgnl, lower=xxx[i], upper=xxx[i+1],
stop.on.error=FALSE)$value
}
```

```
#che<- sum(phat1)

#che

phat<- phat1


#asymptotics percentage points.

phat_a<-c(phat[-1],phat[1])

t_hat<-  (phat+phat_a)/2

I <- diag(k)

n<- sum(m)

exphat<- n*t_hat


#create a partial sum matrix.

A<- matrix(rep(1,k^2),k)

for (i in 1:k)

{ for (j in 1:k)

  { if (j>i)

    {A[i,j]<- 0}

 }

}

di<- m-exphat

Z<- A%*%di

# D is the k by k diagonal matrix whose jth diagonal entry is pj.

#D<- diag(phat)

E<- diag(t_hat)

derp.alpha<- numeric(k)

derp.beta<- numeric(k)
```

```
derp.eta<- numeric(k)
derp.sigma<- numeric(k)
derp.xi<- numeric(k)
for (i in 1:k)
{
derp.alpha[i]<- integrate(derin.a, lower=xxx[i], upper=xxx[i+1],
stop.on.error=FALSE)$value
derp.beta[i]<- integrate(derin.b, lower=xxx[i], upper=xxx[i+1],
stop.on.error=FALSE)$value
derp.eta[i]<- integrate(derin.eta, lower=xxx[i], upper=xxx[i+1],
stop.on.error=FALSE)$value
derp.sigma[i]<- integrate(derin.sigma2, lower=xxx[i], upper=xxx[i+1],
stop.on.error=FALSE)$value
derp.xi[i]<- integrate(derin.xi, lower=xxx[i], upper=xxx[i+1],
stop.on.error=FALSE)$value
}
bhat<- cbind(derp.alpha,derp.beta, derp.eta, derp.sigma, derp.xi)
#bhat<- cbind(derp.mu, derp.sigma,derp.alpha,derp.beta)
#vhat<- solve(t(bhat)%*%solve(D)%*%bhat)
vhat<- chol2inv(chol(t(bhat)%*%solve(E)%*%bhat))
one<- rep(1,k)
Usq1<-
t(Z)%*%(I-E%*%one%*%t(one))%*%E%*%(I-one%*%t(one)%*%E)%*%Z/n
Usq1
cat("U square statistic=", Usq1)
```

```
#Calculating the asymptotic percentage points:
#SigO<- D-phat%*%t(phat)

SigO<- E-t_hat%*%t(t_hat)
SigD<- SigO-bhat%*%vhat%*%t(bhat)
SigU<- A%*%SigD%*%t(A)

#for p, using the following code:
#MsigU<-
(I-D%*%one%*%t(one))%*%D%*%(I-one%*%t(one)%*%D)%*%SigU
MsigU<-
 (I-E%*%one%*%t(one))%*%E%*%(I-one%*%t(one)%*%E)%*%SigU
eigU<- eigen(MsigU)
eigUval<- Re(eigU$values)

k1u<- sum(eigUval)
k1u
#chk.
k2u<- 2*sum((eigUval)^2)
k2u
#chk.
k3u<- 8*sum((eigUval)^3)
k3u

bappu<- k3u/(4*k2u)
```

```
pappu<- 8*(k2u^3)/(k3u^2)

aappu<- k1u-bappu*pappu
aappu

Usq50<- aappu+bappu*qchisq(0.5,pappu)
Usq90<- aappu+bappu*qchisq(0.9,pappu)
Usq95<- aappu+bappu*qchisq(0.95,pappu)
Usq975<- aappu+bappu*qchisq(0.975,pappu)
Usq99<- aappu+bappu*qchisq(0.99,pappu)
Usq995<- aappu+bappu*qchisq(0.995,pappu)
Usq999<- aappu+bappu*qchisq(0.999,pappu)
Uasymout<- c(Usq50, Usq90,Usq95,Usq975,Usq99,Usq995,Usq999)
cat("U square asymptotic points:
at 0.5, 0.9, 0.95, 0.975, 0.995 and 0.999 levels", Uasymout)
p_val<- 1-pchisq((Usq1-aappu)/bappu,pappu)
cat("U square p value=", p_val)
}



########################################
########################################
#Obtaining ML estimates for WGNL distributions
#when data is continuous.
#Calculating the value of continuous U square statistic.
```

```
#Monte carlo programs using continuous data
#checking grouped asymptotic
#points for Watson's U square statistic after
#fitting the WGNL distributions.

modulo<-function(x,k)
{
# finds x mod(k)
 ifelse(x>0,k*(x/k -floor(x/k)),k*(x/k -ceiling(x/k)))
}

#A function generates WGNL pseudo random variables.
#Be careful! alpha=1/a, beta=1/b in the following generator.
WGNL.sim<-function(n,alpha,beta,mu,sigma2,xi)
{
temp<-mu*xi+ sqrt(sigma2*xi)*rnorm(n,0,1) +
rexp(n,1)/alpha -rexp(n,1)/beta
ifelse(temp>0,modulo(temp,2*pi),2*pi+modulo(temp,2*pi))
}

###################
mu.fun<-function(p, a, b, mu)
{
tem1<- (1+a*b*p^2)*cos(mu*p)+(b-a)*p*sin(mu*p)
tem2<- (1+a*b*p^2)*sin(mu*p)-(b-a)*p*cos(mu*p)
```

```
atan2(tem2, tem1)
}

a.p.fun<- function(p, a, b, mu, sigma2, xi)
{
(exp(-sigma2*p^2)/((1+a^2*p^2)*(1+b^2*p^2)))^(xi/2)*
cos(xi*mu.fun(p, a, b, mu))
}

b.p.fun<- function(p, a, b, mu, sigma2, xi)
{
(exp(-sigma2*p^2)/((1+a^2*p^2)*(1+b^2*p^2)))^(xi/2)*
sin(xi*mu.fun(p, a, b, mu))
}


pdf.gnl<- function(theta)
{
temp<- a.p.fun(1,a,b,mu,sigma2,xi)*cos(1*theta)+
b.p.fun(1,a,b,mu,sigma2,xi)*sin(1*theta)+
a.p.fun(2,a,b,mu,sigma2,xi)*cos(2*theta)+
b.p.fun(2,a,b,mu,sigma2,xi)*sin(2*theta)+
a.p.fun(3,a,b,mu,sigma2,xi)*cos(3*theta)+
b.p.fun(3,a,b,mu,sigma2,xi)*sin(3*theta)+
a.p.fun(4,a,b,mu,sigma2,xi)*cos(4*theta)+
b.p.fun(4,a,b,mu,sigma2,xi)*sin(4*theta)+
```

```
a.p.fun(5,a,b,mu,sigma2,xi)*cos(5*theta)+
b.p.fun(5,a,b,mu,sigma2,xi)*sin(5*theta)+
a.p.fun(6,a,b,mu,sigma2,xi)*cos(6*theta)+
b.p.fun(6,a,b,mu,sigma2,xi)*sin(6*theta)+
a.p.fun(7,a,b,mu,sigma2,xi)*cos(7*theta)+
b.p.fun(7,a,b,mu,sigma2,xi)*sin(7*theta)+
a.p.fun(8,a,b,mu,sigma2,xi)*cos(8*theta)+
b.p.fun(8,a,b,mu,sigma2,xi)*sin(8*theta)+
a.p.fun(9,a,b,mu,sigma2,xi)*cos(9*theta)+
b.p.fun(9,a,b,mu,sigma2,xi)*sin(9*theta)+
a.p.fun(10,a,b,mu,sigma2,xi)*cos(10*theta)+
b.p.fun(10,a,b,mu,sigma2,xi)*sin(10*theta)+
a.p.fun(11,a,b,mu,sigma2,xi)*cos(11*theta)+
b.p.fun(11,a,b,mu,sigma2,xi)*sin(11*theta)+
a.p.fun(12,a,b,mu,sigma2,xi)*cos(12*theta)+
b.p.fun(12,a,b,mu,sigma2,xi)*sin(12*theta)+
a.p.fun(13,a,b,mu,sigma2,xi)*cos(13*theta)+
b.p.fun(13,a,b,mu,sigma2,xi)*sin(13*theta)+
a.p.fun(14,a,b,mu,sigma2,xi)*cos(14*theta)+
b.p.fun(14,a,b,mu,sigma2,xi)*sin(14*theta)+
a.p.fun(15,a,b,mu,sigma2,xi)*cos(15*theta)+
b.p.fun(15,a,b,mu,sigma2,xi)*sin(15*theta)+
a.p.fun(16,a,b,mu,sigma2,xi)*cos(16*theta)+
b.p.fun(16,a,b,mu,sigma2,xi)*sin(16*theta)+
a.p.fun(17,a,b,mu,sigma2,xi)*cos(17*theta)+
b.p.fun(17,a,b,mu,sigma2,xi)*sin(17*theta)+
```

```
a.p.fun(18,a,b,mu,sigma2,xi)*cos(18*theta)+
b.p.fun(18,a,b,mu,sigma2,xi)*sin(18*theta)+
a.p.fun(19,a,b,mu,sigma2,xi)*cos(19*theta)+
b.p.fun(19,a,b,mu,sigma2,xi)*sin(19*theta)+
a.p.fun(20,a,b,mu,sigma2,xi)*cos(20*theta)+
b.p.fun(20,a,b,mu,sigma2,xi)*sin(20*theta)
     (1+2*temp)/(2*pi)
}

obj.pdf.gnl<- function(theta, a, b, mu, sigma2,xi)
{
temp<- a.p.fun(1,a,b,mu,sigma2,xi)*cos(1*theta)+
b.p.fun(1,a,b,mu,sigma2,xi)*sin(1*theta)+
a.p.fun(2,a,b,mu,sigma2,xi)*cos(2*theta)+
b.p.fun(2,a,b,mu,sigma2,xi)*sin(2*theta)+
a.p.fun(3,a,b,mu,sigma2,xi)*cos(3*theta)+
b.p.fun(3,a,b,mu,sigma2,xi)*sin(3*theta)+
a.p.fun(4,a,b,mu,sigma2,xi)*cos(4*theta)+
b.p.fun(4,a,b,mu,sigma2,xi)*sin(4*theta)+
a.p.fun(5,a,b,mu,sigma2,xi)*cos(5*theta)+
b.p.fun(5,a,b,mu,sigma2,xi)*sin(5*theta)+
a.p.fun(6,a,b,mu,sigma2,xi)*cos(6*theta)+
b.p.fun(6,a,b,mu,sigma2,xi)*sin(6*theta)+
a.p.fun(7,a,b,mu,sigma2,xi)*cos(7*theta)+
b.p.fun(7,a,b,mu,sigma2,xi)*sin(7*theta)+
a.p.fun(8,a,b,mu,sigma2,xi)*cos(8*theta)+
```

```
b.p.fun(8,a,b,mu,sigma2,xi)*sin(8*theta)+
a.p.fun(9,a,b,mu,sigma2,xi)*cos(9*theta)+
b.p.fun(9,a,b,mu,sigma2,xi)*sin(9*theta)+
a.p.fun(10,a,b,mu,sigma2,xi)*cos(10*theta)+
b.p.fun(10,a,b,mu,sigma2,xi)*sin(10*theta)+
a.p.fun(11,a,b,mu,sigma2,xi)*cos(11*theta)+
b.p.fun(11,a,b,mu,sigma2,xi)*sin(11*theta)+
a.p.fun(12,a,b,mu,sigma2,xi)*cos(12*theta)+
b.p.fun(12,a,b,mu,sigma2,xi)*sin(12*theta)+
a.p.fun(13,a,b,mu,sigma2,xi)*cos(13*theta)+
b.p.fun(13,a,b,mu,sigma2,xi)*sin(13*theta)+
a.p.fun(14,a,b,mu,sigma2,xi)*cos(14*theta)+
b.p.fun(14,a,b,mu,sigma2,xi)*sin(14*theta)+
a.p.fun(15,a,b,mu,sigma2,xi)*cos(15*theta)+
b.p.fun(15,a,b,mu,sigma2,xi)*sin(15*theta)+
a.p.fun(16,a,b,mu,sigma2,xi)*cos(16*theta)+
b.p.fun(16,a,b,mu,sigma2,xi)*sin(16*theta)+
a.p.fun(17,a,b,mu,sigma2,xi)*cos(17*theta)+
b.p.fun(17,a,b,mu,sigma2,xi)*sin(17*theta)+
a.p.fun(18,a,b,mu,sigma2,xi)*cos(18*theta)+
b.p.fun(18,a,b,mu,sigma2,xi)*sin(18*theta)+
a.p.fun(19,a,b,mu,sigma2,xi)*cos(19*theta)+
b.p.fun(19,a,b,mu,sigma2,xi)*sin(19*theta)+
a.p.fun(20,a,b,mu,sigma2,xi)*cos(20*theta)+
b.p.fun(20,a,b,mu,sigma2,xi)*sin(20*theta)
     (1+2*temp)/(2*pi)
```

```
}
obj.GNL<- function(par, data)
{
-sum(log(obj.pdf.gnl(data,exp(par[1]),
exp(par[2]),par[3],exp(par[4]), par[5])))
}
#mc is the number of Monte carlo samples that
#the user is running.
alphahat<- numeric(mc)
betahat<- numeric(mc)
muhat<- numeric(mc)
sigma2hat<- numeric(mc)
xihat<- numeric(mc)
Usq.gnl<- numeric(mc)
lik.gnl<- numeric(mc)
for(re in 1: mc)
{
y<- WGNL.sim(n,MLE.alpha, MLE.beta,
MLE.mu, MLE.sigma2, MLE.xi)
#generating n WGNL observations
# based on the ML estimates from the data.
out.gnl<- optim(par=c(start.alpha, start.beta,
start.mu, start.sigma2, start.xi),
fn=obj.GNL, data=y, control=list(maxit=10000))
#starting values here can be the ML estimates above.
#or the method suggested in the thesis.
```

```
alphahat[re]<- exp(out.gnl$par[1])
betahat[re]<-exp(out.gnl$par[2])
muhat[re]<- out.gnl$par[3]
sigma2hat[re]<-exp(out.gnl$par[4])
xihat[re]<-out.gnl$par[5]
lik.gnl[re]<- -out.gnl$value
a<- exp(out.gnl$par[1])
b<-exp(out.gnl$par[2])
mu<- out.gnl$par[3]
sigma2<-exp(out.gnl$par[4])
xi<-out.gnl$par[5]
Zi<- numeric(n)
for (i in 1:n)
{
Zi[i]<- integrate(pdf.gnl, lower=0, upper= y[i],
stop.on.error=FALSE)$value
}
z_ord<- sort(Zi)
z_bar<- sum(Zi)/n
get.usq<- numeric(n)
for(i in 1:n)
{
get.usq[i]<- z_ord[i]-(2*i-1)/(2*n)
}
Usq.gnl[re]<- sum(get.usq^2)-n*(z_bar-0.5)^2+1/(12*n)
cat("ML estimates=",c(a,b,mu,sigma2,xi))
```

```
cat("likelihood value=", -out.gnl$value)
}
sort.U<- sort(Usq.gnl)
sort.alpha<- sort(alphahat)
sort.beta<- sort(betahat)
sort.muhat<- sort(muhat)
sort.sigma2hat<- sort(sigma2hat)
sort.xihat<- sort(xihat)
sort.lik<- sort(lik.gnl)
AIC.gnl<- AICzs(lik.gnl,5)
BIC.gnl<- BICzs(lik.gnl,5, 1827)
sort.AIC<- sort(AIC.gnl)
sort.BIC<- sort(BIC.gnl)
#90%, 95%, 99%
#c(sort.U[4500],sort.U[4750],sort.U[4950])
#50%, 75%, 90%, 95%, 97.5%, 99%, 99.5%, 99.9%
#c(sort.U[2500],sort.U[3750], sort.U[4500],sort.U[4750],sort.U[4875],
sort.U[4950], sort.U[4975], sort.U[4995])
#90%, 95%, 99%
c(sort.U[900],sort.U[950],sort.U[990])
#50%, 75%, 90%, 95%, 97.5%, 99%, 99.5%, 99.9%
c(sort.U[500],sort.U[750], sort.U[900],sort.U[950],sort.U[975],
sort.U[990], sort.U[995], sort.U[999])
#estimates percentiles (also called Bootstrap CIs in the thesis):
#0.5%, 2.5%, 5%, 10%, 25%, 50%,
#75%, 90%, 95%,97.5%, 99%, 99.5%
```

```
c(sort.alpha[5],sort.alpha[25],sort.alpha[50],sort.alpha[100],
sort.alpha[250],sort.alpha[500],sort.alpha[750], sort.alpha[900],
sort.alpha[950], sort.alpha[975], sort.alpha[990], sort.alpha[995])
#using the similar code, we can
#obtain Bootstrap CIs for other estimates.
#checking moments for ML estimates.
mon1.alpha<- sum(alphahat)/mc
mon2.alpha<- sum(alphahat^2)/mc
mon3.alpha<-sum(alphahat^3)/mc
mon4.alpha<-sum(alphahat^4)/mc
mon1.alpha
mon2.alpha
mon3.alpha
mon4.alpha
mon1.beta<- sum(betahat)/mc
mon2.beta<-sum(betahat^2)/mc
mon3.beta<-sum(betahat^3)/mc
mon4.beta<- sum(betahat^4)/mc
mon1.beta
mon2.beta
mon3.beta
mon4.beta
mon1.mu<- sum(muhat)/mc
mon2.mu<-sum(muhat^2)/mc
mon3.mu<-sum(muhat^3)/mc
mon4.mu<- sum(muhat^4)/mc
```

```
mon1.mu

mon2.mu

mon3.mu

mon4.mu

mon1.sigma2<- sum(sigma2hat)/mc

mon2.sigma2<-sum(sigma2hat^2)/mc

mon3.sigma2<-sum(sigma2hat^3)/mc

mon4.sigma2<- sum(sigma2hat^4)/mc

mon1.sigma2

mon2.sigma2

mon3.sigma2

mon4.sigma2

mon1.xi<- sum(xihat)/mc

mon2.xi<-sum(xihat^2)/mc

mon3.xi<-sum(xihat^3)/mc

mon4.xi<- sum(xihat^4)/mc

mon1.xi

mon2.xi

mon3.xi

mon4.xi
```

Bibliography

[1] Abramowitz, M. and Stegun, I. A. (1970). Handbook of Mathematics Functions. New York: National Bureau of Standards. (50, 51).

[2] Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain goodness-of-fit criteria based on stochastic process. Ann. Math. Stat., 23, 193-212.

[3] Ajne, B. (1968). A simple test for uniformity of a circular distribution. Biometrika, 55, 343-354.

[4] Azzalini, A. (1985). A class of distributions which includes the normal ones. Scand. J. Statist., 12, 171-178.

[5] Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. J. R. Statist. Soc., B, 61, 579-602.

[6] Barndorff-Nielsen, O. and Cox, D. R. (1980). Edgeworth and saddle-point approximations with statistical applications (with discussion). J. R. Statist. Soc., B, 41, 279-312.

[7] Batschelet, E. (1981). Circular statistics in Biology. London: Academic Press.

[8] Bentley, J. (2006). Modelling circlar data using a mixture of von Mises and Uniform distributions. Master's Thesis, Simon Fraser Univ..

[9] Berkson, J. (1980). Minimum chi-square, not maximum likelihood! Ann. Stat., 8(3), 457-487.

[10] Best, D. J. and Fisher, N. I. (1979). Efficient simulation of the von Mises distribution. Appl. Statist., 28, 152-157.

[11] Best, D. J. and Fisher, N. I. (1981). The bias of the maximum likelihood estimators of the von Mises-Fisher concentration parameters. Commun. Statist-Simul. Comput. B10(5), 493-502.

[12] Bruderer, B. and Jenni, L. (1990). Migration across the Alps. In Bird Migration: Physiology and Ecophysiology, 60-77. Berlin:Springer-Verlag.

[13] Cox, D. R. (1975). Contribution to discussion of Mardia (1975). J. Roy. Statist. Soc., B, 37, 380-381.

[14] Chambers, J. M., Mallows, C. and Stuck, B. W. (1976). A method for simulating stable random variables. J. Amer. Statist. Assoc., 71, 340-344.

[15] Dagpunar, J. S. (1990). Sampling from the von Mises distribution via a comparison of random numbers. J. Appl. Statist., 17, 165-168.

[16] D' Agostino, R. B. and Stephens, M. A. (1986). Goodness-of-fit techniques. New York: Marcel Decker.

[17] Fieller, N. R. J., Flenley, E. C. and Olbricht, W. (1992). Statistics of particle size data. Appl. Statist., 41, 127-146.

[18] Fisher, N. I. (1993). Statistical analysis of circular data. Cambridge: Cambridge University Press.

[19] Fisher, R. A. (1953). Dispersion on a sphere. Proc. Roy. Soc. London, A, 217, 295-305.

[20] Gatto, R. and Jammalamadaka, S. R. (2003). Inference for wrapped symmetric α stable circular models. Sanhky$\bar{a}$, 65, 333-355.

[21] Gatto, R. and Jammalamadaka, S. R. (2006). The Generalized von Mises distribution. Statistical Methodology, 4(3), 341-353.

[22] Greenwood, J. A. and Durand, D. (1955). The distribution of length and components of the sum of n random unit vectors. Ann. Math. Stat., 26, 233-246.

[23] Gumbel, E. J., Greenwood, J. A. and Durand, D. (1953). The circular normal distribution: theory and tables. J. Amer. Stat. Assoc., 48, 131-152.

[24] Henze, N. (1986). A probabilistic representation of the 'skew-normal' distribution. Scand. J. Stat., 13: 271-275.

[25] Hinkley, D. V. and Revankar, N. S. (1977). On the estimation of the Pareto law from underreported data. A further analysis. J. Econometr., 5, 1-11.

[26] Hodges, J. L., Jr. (1955). A bivariate sign test. Ann. Math. Stat., 26, 523-527.

[27] Hurst, S. (1995). The characteristic function of the Student t distribution. Financial Mathematics Research Report No. FMRR 006-95, Statistics Research Report No. SRR 044-95, Centre for Financial Mathematics, School of Mathematical Sciences, Australian National Univ., Canberra.

[28] Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables. Biometrika, 48, 419-426.

[29] Jander, R. (1957). Die Optische Richtungsorientierung der roten Waldameises (Formica rufa). Zeitschrift fur vergleichende Physiologie, 40, 162-238.

[30] Jammalamadaka and SenGupta (2001). Topics in circular statistics. New Jersey: World Scientific.

[31] Jammalamadaka, S. R. and Kozubowski, T. J. (2003). A new family of circular models: The wrapped Laplace distributions. Advances and Applications in Statistics, 3, 77-103.

[32] Jammalamadaka, S. R. and Kozubowski, T. J. (2004). New families of wrapped distributions for modeling skew circular data. Commun. Statist. Theory Meth., 33, 2059-2074.

[33] Johnson, R. A. and Wehrly, T. E. (1987). Some angular-linear distributions and related regression models. J. Amer. Stat. Assoc., 73, 602-606.

[34] Julia, O. and Vives-Rego, J. (2005). Skew-Laplace distribution in Gram-negative bacterial axenic cultures: new insights into intrinsic cellular heterogeneity. Microbiology, 151, 749-755.

[35] Kato, S. and Shimizu, K. (2005). A further study of t-distributions on spheres. Technical report, School of Fundamental Science and Technology, Keio Univ. , Yokohama.

[36] Stuart, A., Ord, J. K. and Arnold, S. (1994). Kendall's advanced theory of statistics. London: Arnold.

[37] Kotz. S., Kozubowski. T. J., and Podgorski K. (2001). The Laplace distribution and generalizations. Boston: Birkhauser.

[38] Keiper, N. H. (1960). Tests concerning random points on a circle. Konikl. Nederl. Akad. Wet. Proc., A, 63, 38-47.

[39] Lai, M. (1994). Some results in the statistical analysis of directional data. Master's thesis, Univ. of Hong Kong.

[40] Laubscher, N. F. and Rudolph, G. J. (1968). A distribution arising from random points on the circumference of a circle. Nat. Res. Inst. Math. Sci. Report 268, Pretoria, South Africa, 1-15.

[41] Levy, P. (1924). Theorie des erreurs la loi de Gauss et les lois exceptionnelles. Bull. Soc. Math. France, 52, 49-85.

[42] Lebedev, N. N. (1965). Special functions and their applications. Englewood Cliffs, New Jersey: Prentice-Hall.

[43] Lockhart, R. A., Spinelli, J. J. and Stephens, M. A. (2007). Cramér-von Mises statistics for discrete distributions with unknown parameters. Canad. J. Statist., 35, 125-133.

[44] Lockhart, R. A. and Stephens, M. A. (1985). Tests of fit for the von Mises distribution. Biometrika, 72, 647-652.

[45] Mardia, K.V. (1975). Statistics of directional data (with discussion). J. Royal Statist. Soc., B, 37, 349-393.

[46] Mardia, K. V. and Jupp P. E. (2000). Directional statistics. Chichester: Wiley.

[47] Neader, J. A. and Mead, R. (1965). A simplex method for function minimization. Comput. J., 7, 308-313.

[48] Nolan, J. P. (1998). Parameterizations and modes of stable distributions. Statist. Probab. Lett., 38, 187-195.

[49] Pearson, K. (1905). The problem of the random walk. Nature, 72, 294.

[50] Pewsey, A. (2000). The wrapped skew normal distribution on the circle. Commun. Statist. Theory Meth., 29, 2459-2472.

[51] Pewsey, A. (2006). Modelling asymmetrically distributed circular data using the wrapped skewed-normal distribution. Environ. Ecol. Stat., 13, 257-269.

[52] Pewsey, A., Lewis, T. and Jones, M. C. (2007). The wrapped t Family of circular distributions. Aust. N. Z. J. Stat., 49, 79-91.

[53] Pewsey, A. (2008). The wrapped stable family of distributions as a flexible model for circular data. Comput. Stat. Data. Anal., 52,1516-1523.

[54] Rao, C. R. (1958). Maximum likelihood for the multinomial distribution with infinite number of cells. Sankhyā, 20, 211-218.

[55] Rao, J. S. (1976). Some tests based on arc-lengths for the circle. Sankhyā, B, 33, 1-10.

[56] Rayleigh, Lord. (1880). On the resultant of a large number of vibrations of the same pitch and of arbitrary plane. Phil. Mag., 10, 73-78.

[57] Reed, W. J. (2003). The Pareto law of incomes - an explanation and an extension. Physica. A, 319, 579-597.

[58] Reed, W. J. (2004). The Normal-Laplace Distribution and its Relatives. In Order Statistics and Inference, N.Balakrishna et al. (eds) Birkhauser.

[59] Reed, W. J. and M. Jorgensen (2004a). The double Pareto-lognormal distribution- A new parametric model for size distribution. Commun. Statist. Theory Meth., 33, 1733-1753.

[60] Reed, W. J. (2007). Brownian-Laplace motion and its application in financial modeling. Commun. Statist. Theory Meth., 36. 473-484.

[61] Reed, W. J. and Pewsey, A. (2008). Two nested families of skew-symmetric circular distributions. Test. 18, 1133-0686.

[62] Saw, J. G. (1978). A family of distributions on the m-sphere and some hypothesis tests. Biometrika, 70, 665-671.

[63] Choulakian, V., Lockhart, R. A. and Stephens, M. A. (1994). Cramér-von Mises statistics for discrete distributions, Canad. J. Statist., 22, 125-137.

[64] Spinelli, J. J. (2001). Testing fit for the grouped exponential distribution. Canad. J. Statist., 29, 451-458.

[65] Stephens, M. A. (1963). Random walk on a circle. Biometrika, 50, 385-390.

[66] Stephens, M. A.(1964). The distribution of the goodness-of-fit statistic, U_N^2. II. Biometrika, 51, 393-397.

[67] Stephens, M. A. (1965). The goodness-of-fit statistic V_N: distribution and significance points. Biometrika, 52, 309-321.

[68] Stephens, M. A. (1969). Techniques for directional data. Technical Report 150, Department of Statistics, Stanford Univ..

[69] Stephens, M. A. (1972). Multisample tests for the von Mises distribution. J. Amer. Statist. Assoc. 67, 456-461.

[70] Stephens, M. A. (1976a). Asymptotic results for goodness-of-fit statistics with unknown parameters. Ann. Statist. 4, 357-369.

[71] Stephens, M. A. (1976b). Asymptotic power of EDF statistics for exponentiality against Gamma and Weibull alternatives. Technical report No. 297, Department of Statistics, Stanford Univ..

[72] Stephens, M. A. (1985). Tests for the Cauchy distribution based on the EDF. Technical report, Department of Statistics, Stanford Univ..

[73] Schmidt, W. (1917). Statistische Methoden beim Gefugestudium Krisalliner Schiefer. Sitz. Kaiserl. Akad. Wiss. Wien, Math. nat. Kl. Abt. 1,126, 515-538.

[74] von Mises, R. (1918). Uber die "Ganzzahligkeit" der Atomgewichte und verwandte Fragen. Physikal. Z. 19, 490-500.

[75] Watson, G. S. (1961). Goodness-of-fit tests on a circle. Biometrika, 48, 109-114.

[76] Watson, G. S. and Williams, E. J. (1956). On the construction of significance tests on the circle and the sphere. Biometrika, 43, 344-352.

[77] Yfantis, E. A. and Borgman, L. E. (1982). An extension of the von Mises distribution. Commun. Statist. Theory Meth., 11, 1695-1706.