MODELLING CIRCULAR DATA USING A MIXTURE OF VON MISES AND UNIFORM DISTRIBUTIONS

by

John Bentley

B.Sc., Simon Fraser University, 1998.

A PROJECT SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
in the Department
of
Statistics and Actuarial Science

SIMON FRASER UNIVERSITY
Fall 2006

APPROVAL

Name: John Bentley

Degree: Master of Science

Title of project: Modelling Circular Data Using a Mixture of von Mises and Uniform Distributions

Examining Committee: Dr. Carl Schwarz
Chair

Dr. Richard Lockhart
Simon Fraser University
Senior Supervisor

Dr. Michael Stephens
Simon Fraser University

Dr. Derek Bingham
External Examiner
Simon Fraser University

Date Approved: December 7, 2006

Abstract

The von Mises distribution is often useful for modelling circular data problems. We consider a model for which von Mises data is contaminated with a certain proportion of points uniformly distributed around the circle. Maximum likelihood estimation is used to produce parameter estimates for this mixture model. Computational issues involved with obtaining the maximum likelihood estimates for the mixture model are discussed. Both parametric and goodness-of-fit based test procedures are presented for selecting the appropriate model (uniform, von Mises, mixture) and determining its adequacy. Parametric tests presented in this project are based on the likelihood ratio test statistic and goodness-of-fit tests are based on Watson's goodness-of-fit statistic for the circle. A parametric bootstrap is performed to obtain the approximate distribution of Watson's statistic in situations where the true parameter values are unknown.

Keywords: goodness-of-fit; model selection; parametric bootstrap

Acknowledgments

A special thanks go to both my supervisors who were extremely helpful throughout this project and without whom I would not have made it through the program.

Dr. Lockhart was very patient with me and was willing to take time away from his busy schedule to meet with me twice a week throughout the project. Dr. Lockhart provided me with extremely valuable advice on how the project should be structured and a list of recommended readings to help familiarize myself with the topic. Throughout the project, Dr. Lockhart supplied useful suggestions for items to include, and an abundance of theoretical help. Advice I received from Dr. Lockhart on strategies for model selection and tests of fit for Chapter 5 was particularly helpful.

The idea for the mixture model was provided by Dr. Stephens and I am very grateful for him giving me such an enjoyable project to work on. Dr. Stephens also provided me with valuable advice on relevant papers to read and suggested including Examples 1 and Examples 3 for this project. Example 3 was particularly useful in illustrating some complications that can arise when trying to find the maximum likelihood estimates for the mixture distribution.

I am also very appreciative of all the assistance I have received from all the professors and support staff in the department. Other graduate students in the department have also been very helpful to me on many occasions. The Statistics and Actuarial Science Department at Simon Fraser University is truly first class and I feel very fortunate to have been accepted into the program.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

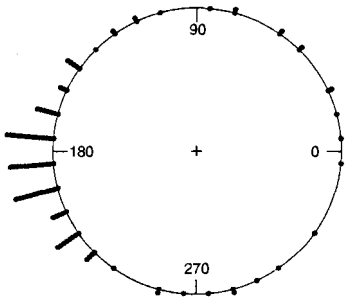
Most scientific fields (Biology, Chemistry, Physics, Medicine, ...) have applications in which directions are collected and statistically analysed. Some examples of directional data include animal orientations (associated with migration, homing, escape or exploratory activity) [2] and wind directions. Circular representations are also often used with cyclic time series data. For example the times in which patient deaths occur can be recorded and given a circular representation with a full 24 hour period corresponding to 360°.

The von Mises distribution is commonly used as a model for many circular data problems. In some situations, the von Mises model appears appropriate but is unable to sufficiently model both the number of points that are tightly concentrated around the mean direction and the number of points that are more dispersed at the opposite end of the circle. One can potentially explain the above situation as resulting from a certain proportion of data coming from a von Mises distribution while the remaining proportion is randomly (or uniformly) scattered around the circle. In this project a mixture of von Mises and circular distributions are used to provide a model that is suitable for data in which the von Mises model appears appropriate but for the reasons described above is not able to sufficiently model the data.

An experiment done on the behavior of ants in the presence of a light source (see Chapter 2, Section 2.6 for references and details) is an example for which the mixture model is appropriate. Ants were placed into an arena one at a time, and the directions they chose relative to an evenly illuminated black light source placed at 180° were recorded. The orientations of 100 ants are illustrated in Figure 1.1.

In Chapter 2, the von Mises model is fit to the ant data. Maximum likelihood parameter

Figure 1.1: Circular data plot of orientations of 100 ants



estimates are provided and large sample theory is used to provide the asymptotic distribution of the MLE. The P-P plot is discussed as a way of graphically assessing the fit of the model. A shortcoming of the use of the P-P plot for circular data is that visual assessment of the goodness-of-fit of the data may depend on how the data has been oriented around the circle and this shortcoming is also discussed.

The von Mises distribution is not able to sufficiently model both the number of ants that are concentrated in directions around the light source and the number of ants that are scattered about in the opposite direction. In Chapter 3, the mixture model that is introduced provides a better model for explaining the behavior of the ants. Maximum likelihood parameter estimates are provided for the mixture model and large sample theory is used to provide the asymptotic distribution of the MLE.

In Chapter 4 various computational details concerned with the calculation of the MLE for the mixture distribution are discussed. A simple algorithm is provided for obtaining the maximum likelihood parameter estimates for the von Mises distribution. A discussion is provided on the ill behavior of the likelihood function of the mixture distribution in certain regions of the parameter space. While it does not fit in with our motivation for the mixture model, the mathematical possibility of the mixture distribution having a proportion of von Mises distributed data greater than 1 is discussed. A circular data example is provided where

the likelihood function has higher values when the proportion of von Mises distributed data is allowed to be more than 1. This situation does not fit in with our initial motivation for the model, and von Mises proportions greater than 1 are not practical for use with the mixture model. Thus we provide a simple way of detecting wether or not higher likelihoods exist for von Mises proportions greater than 1 and in that event the von Mises model can be used instead. Finally we provide an algorithm for obtaining the MLE for the mixture distribution, including the calculation of initial parameter estimates.

In Chapter 5 we examine the goodness-of-fit of the ant data to the different models. Two different approaches for testing fit and selecting the appropriate model are discussed. A non-parametric approach is discussed in which Watson's U^2 statistic is used to assess the fit of a model. To obtain the approximate distribution of the U^2 statistic in situations where the true parameter values are unknown, a parametric bootstrap sample is taken. A parametric based approach is also provided in which likelihood ratio tests are used for testing for uniformity against the von Mises alternative and for testing for von Misesness against the mixture alternative.

Chapter 2

The von Mises Distribution

In this chapter we discuss modelling circular data using the von Mises distribution. An introduction to the von Mises distribution is given in Section 2.1. We provide the maximum likelihood estimator for the parameters of the von Mises distribution in Section 2.2 and its asymptotic distribution is given in Section 2.3. A graphical method for assessing the goodness-of-fit of the von Mises model is discussed in Section 2.4. A circular data example is presented in Section 2.5 and we fit the von Mises model to this data. In Section 2.6 we conclude with an example of circular data for which the von Mises model alone does not provide a good fit and a better fit for the data would be the mixture model discussed in Chapter 3.

2.1 Introduction

The most commonly used distribution for modelling circular data is the von Mises distribution. The probability density function of the von Mises distribution is given by

$$f_{VM}(\theta;\mu,\kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa\cos(\theta-\mu)\},\ 0 \le \theta < 2\pi,\ \kappa \ge 0,\ 0 \le \mu < 2\pi,$$

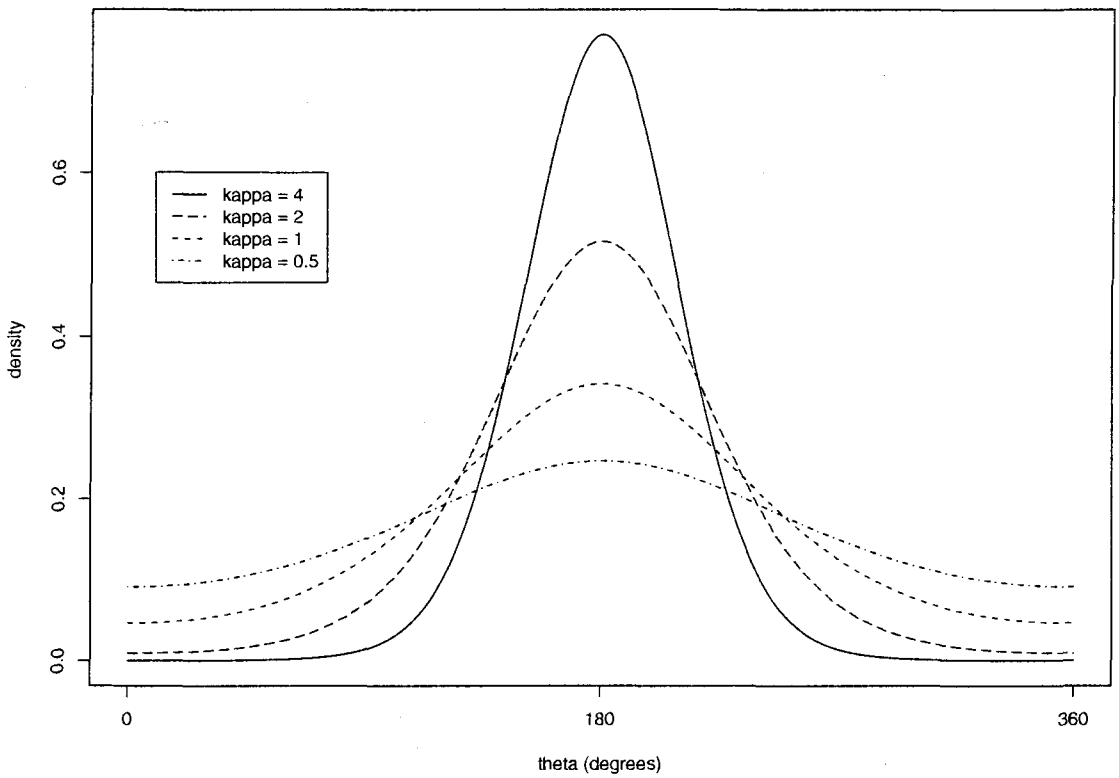
where $I_0(\kappa)$ is the zeroth order modified Bessel function of the first kind, and can be expressed as

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

The mean direction is specified by the μ parameter. The parameter κ influences how concentrated the distribution is around the mean direction. Larger values of κ result in the

distribution being more tightly clustered about the mean direction.

Figure 2.1: Probability density functions of several von Mises distributions



The density functions of several von Mises distributions with mean direction, $\mu = 180°$, and various concentration parameters, κ, are plotted in Figure 2.1.

2.2 Maximum Likelihood Estimation

Often we wish to model circular data according to a von Mises distribution for which the parameters μ and κ are unknown. Suppose that $\theta_1, \ldots, \theta_n$ are n independent random

directions drawn from a von Mises distribution with unknown parameters μ and κ. Let $\phi_{VM} = (\mu, \kappa)$. The log-likelihood function is given by

$$l_{VM}(\phi_{VM}) = \kappa \sum_{i=1}^{n} \cos(\theta_i - \mu) - n \left[\log(2\pi) + \log\left\{I_0(\kappa)\right\}\right].$$

We will give the maximum likelihood parameter estimates shortly but first we need to define the resultant vector and mean angular direction. Definitions for the resultant vector and mean angular direction have been given by Jammalamadaka and SenGupta in [8] (p. 13), for example, and for convenience have been repeated below.

Each of the angles $\theta_1, \ldots, \theta_n$ can be converted from polar to rectangular co-ordinates by using the transformation, $\boldsymbol{x}_i = (\cos(\theta_i), \sin(\theta_i))$, $i = 1, \ldots, n$. The component-wise sums of these unit vectors are defined below.

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

The resultant vector and resultant length are defined as

$$\boldsymbol{R} = (C, S), \quad \text{and} \quad R = \sqrt{C^2 + S^2},$$

respectively.

The mean angular direction, $\bar{\theta}$, is not defined for $R = 0$ (or equivalently, for $C = 0$ and $S = 0$). For $R > 0$, the mean angular direction is given by

$$\bar{\theta} = \begin{cases} \arctan(S/C) & \text{if } C > 0,\ S \geq 0, \\ \pi/2, & \text{if } C = 0,\ S > 0, \\ \arctan(S/C) + \pi & \text{if } C < 0, \\ 3\pi/2, & \text{if } C = 0,\ S < 0, \\ \arctan(S/C) + 2\pi & \text{if } C > 0,\ S < 0. \end{cases}$$

The angle $\bar{\theta}$ is the angle between the resultant vector and the positive x-axis; that is, the vector $\boldsymbol{R}_x$, points in the direction of $\bar{\theta}$. Notice that $0 \leq \bar{\theta} < 2\pi$.

Now, we give a derivation for the maximum likelihood parameter estimates, we follow the presentation of Mardia and Jupp in [13] (p. 85). First note that by using the trigonometric identities

$$\cos(\theta - \bar{\theta} + \bar{\theta} - \mu) = \cos(\theta - \bar{\theta})\cos(\bar{\theta} - \mu) - \sin(\theta - \bar{\theta})\sin(\bar{\theta} - \mu),$$

$$\cos(\theta - \bar{\theta}) = \cos(\theta)\cos(\bar{\theta}) + \sin(\theta)\sin(\bar{\theta}), \quad \text{and}$$

$$\sin(\theta - \bar{\theta}) = \sin(\theta)\cos(\bar{\theta}) - \sin(\bar{\theta})\cos(\theta),$$

the log-likelihood can be re-expressed in the form

$$l_{VM}(\phi_{VM}) = \kappa R\cos(\bar{\theta} - \mu) - n\left[\log(2\pi) + \log\left\{I_0(\kappa)\right\}\right].$$

In this form, the maximum likelihood estimate of μ can be seen to be $\hat{\mu} = \bar{\theta}$, since $\cos(x)$ has its maximum at $x = 0$.

We will also need some results concerning Bessel functions. In general, the j^{th} order modified Bessel function of the first kind, $I_j(\kappa)$, is given by

$$I_j(\kappa) = \frac{1}{2\pi}\int_0^{2\pi} \cos(j\theta)\exp\{\kappa\cos(\theta)\}\, d\theta.$$

The derivative of the zeroth order modified Bessel function of the first kind is equal to the first order modified Bessel function of the first kind.

$$\frac{d}{d\kappa}I_0(\kappa) = \frac{1}{2\pi}\int_0^{2\pi} \cos(\theta)\exp\{\kappa\cos(\theta)\}\, d\theta = I_1(\kappa).$$

Now, differentiating the log-likelihood function with respect to κ gives

$$\frac{\partial l(\phi_{VM})}{\partial\kappa} = R\cos(\bar{\theta} - \mu) - nA(\kappa),$$

where $A(\kappa) = I_1(\kappa)/I_0(\kappa)$. Thus the maximum likelihood estimate $\hat{\kappa}$ of κ is the unique solution of

$$A(\hat{\kappa}) = R/n,$$

i.e.

$$\hat{\kappa} = A^{-1}(R/n).$$

2.3 Large Sample Asymptotic Distribution of the MLE

Let μ_0 and κ_0 be the true parameter values of μ and κ respectively. As mentioned by Mardia and Jupp in [13], standard theory of maximum likelihood estimators, as can be found in [4] (pp. 294-296), can be used to show that $(\hat{\mu}, \hat{\kappa})$, is asymptotically normally distributed,

$$\sqrt{n}(\hat{\mu} - \mu_0, \hat{\kappa} - \kappa_0) \sim N(\mathbf{0}, \boldsymbol{I}^{-1}),$$

where $\boldsymbol{I}$ denotes the Fisher information matrix,

$$\boldsymbol{I} = E\left[-\begin{pmatrix} \frac{\partial^2 f_{VM}}{\partial \mu^2} & \frac{\partial^2 f_{VM}}{\partial \mu \partial \kappa} \\ \frac{\partial^2 f_{VM}}{\partial \kappa \partial \mu} & \frac{\partial^2 f_{VM}}{\partial \kappa^2} \end{pmatrix}\right] = \begin{pmatrix} \kappa A(\kappa) & 0 \\ 0 & 1 - A(\kappa)^2 - A(\kappa)/\kappa \end{pmatrix},$$

based on a single observation. Thus $\hat{\mu}$ and $\hat{\kappa}$ are asymptotically independently normally distributed with means variances given by

$$E(\hat{\mu}) = \mu_0, \quad \text{Var}(\hat{\mu}) \simeq \frac{1}{n\kappa A(\kappa)}, \quad \text{and}$$

$$E(\hat{\kappa}) = \kappa_0, \quad \text{Var}(\hat{\kappa}) \simeq \frac{1}{n[1 - A(\kappa)^2 - A(\kappa)/\kappa]},$$

respectively.

An algorithm for obtaining the maximum likelihood parameter estimates for the von Mises distribution is provided in Section 4.1.

2.4 Graphical Assessment of Goodness-of-fit

One method of graphically assessing the goodness-of-fit of the von Mises model is to construct a Probability-Probability (P-P) plot. To construct a P-P plot we first sort our n angular data values $\theta_1, \ldots, \theta_n$, in order from smallest to largest to obtain $\theta_{(1)}, \ldots, \theta_{(n)}$. We then calculate the cumulative distribution function, $F_{VM}(\theta_{(i)}; \pi, \hat{\kappa})$ and plot it against $(i - 0.5)/n$, for $i = 1, \ldots, n$. If the von Mises model is a good model, then the points on the plot should approximately be along the line $y = x$.

It is, however, important to mention that our visual perceptions as to the goodness-of-fit od the model based on a P-P plot may depend on rotations of the data. We can potentially obtain quite different looking P-P plots simply by rotating the data as illustrated in Figures 2.3 and 2.4 in Section 2.5. Since typically circular data problems are arbitrary assigned starting directions (ie 0° is arbitrarily assigned to a direction), the lack of consistency in appearance of P-P plots based on rotations of the data is not a particularly desirable characteristic. More formal goodness-of-fit tests that are based on Watson's U^2 statistic and do not depend on rotations of the data and are presented in Chapter 5.

The P-P plot can still be useful in identifying situations in which the model is quite clearly inadequate. If the P-P plot deviates significantly both above and below the line

$y = x$, then Watson's U^2 statistic will be relatively large. Large deviations both above and below the line are indication that the model may not be appropriate.

2.5 Example 1

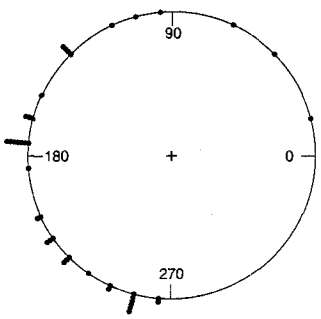
We will now provide an example of data that are approximately von Mises distributed and provide the maximum likelihood parameter estimates.

The directions of slope of 44 lamination surfaces of sandstone rock are given in Table 2.1 and have also been illustrated in the circular data plot in Figure 2.2. The data in Table 2.1 are taken from the first of two samples that Pearson and Stephens [15] (p. 129) use in determining whether or not the samples come from the same von Mises population. Pearson and Stephens originally took the data from Kiersch [10].

Table 2.1: Directions of slope of 44 lamination surfaces of sandstone rock

0	0	0	15	45	68	100	110	113	135	135	140
140	155	165	165	169	180	180	180	180	180	180	180
189	206	209	210	214	215	225	226	230	235	245	250
255	255	260	260	260	260	270	270				

Figure 2.2: Circular data plot of directional sandstone rock data



A von Mises model can be fit to this data by using the maximum likelihood parameter estimates in place of the true parameters. The maximum likelihood parameter estimates and associated standard errors are provided in Table 2.2.

Table 2.2: von Mises maximum likelihood parameter estimates for Example 1

$\hat{\mu}$	stderr($\hat{\mu}$)	$\hat{\kappa}$	stderr($\hat{\kappa}$)
199.4°	12.2°	1.07	0.26

Figures 2.3 and 2.4 show two different P-P plots for the same von Mises model with parameters given above. In Figure 2.3, the fit appears poor but by simply rotating the data by $(180° - \hat{\mu}) = -19.4°$ in Figure 2.4, the fit appears much more reasonable. Thus, in assessing goodness-offit to the von Mises distribution we need methods which do not depend on which angle is chosen to be the 0° point on the circle. As mentioned in the Section 2.4, goodness-of-fit tests based on Watson's U^2 statistic can be used to make such rotationally independent assessments of the fit of the model. These tests are presented in Chapter 5.

Figure 2.3: von Mises P-P plots of directional sandstone rock data (not rotated)

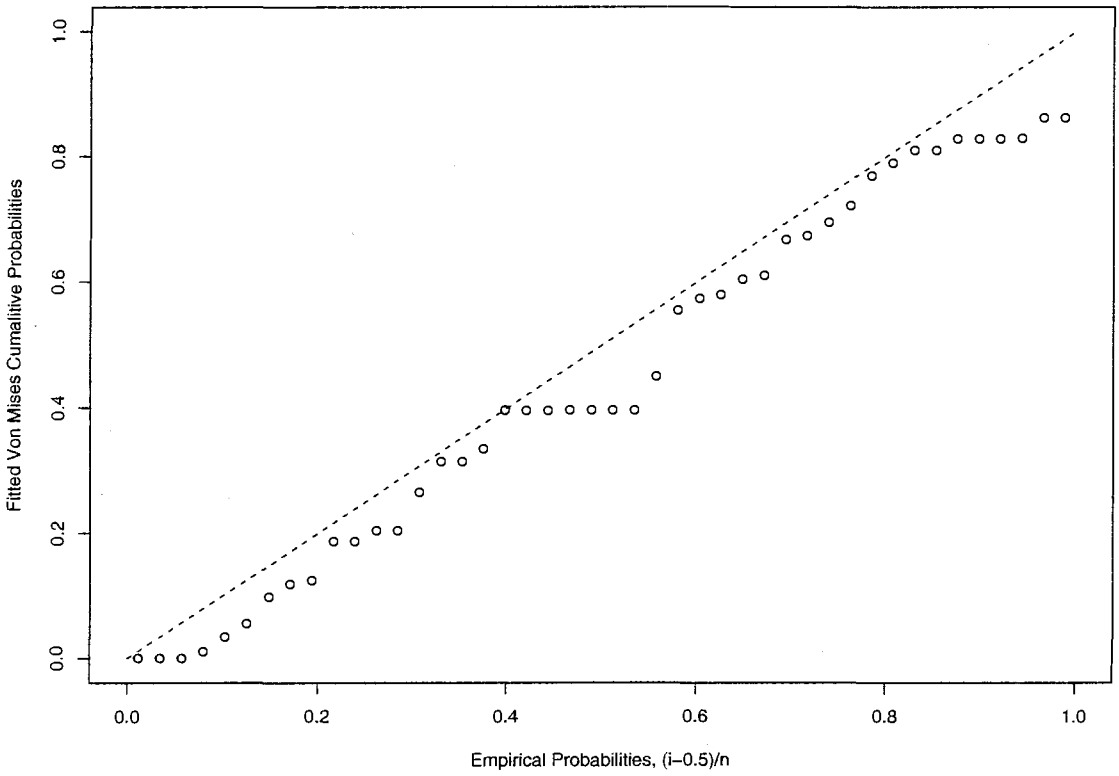
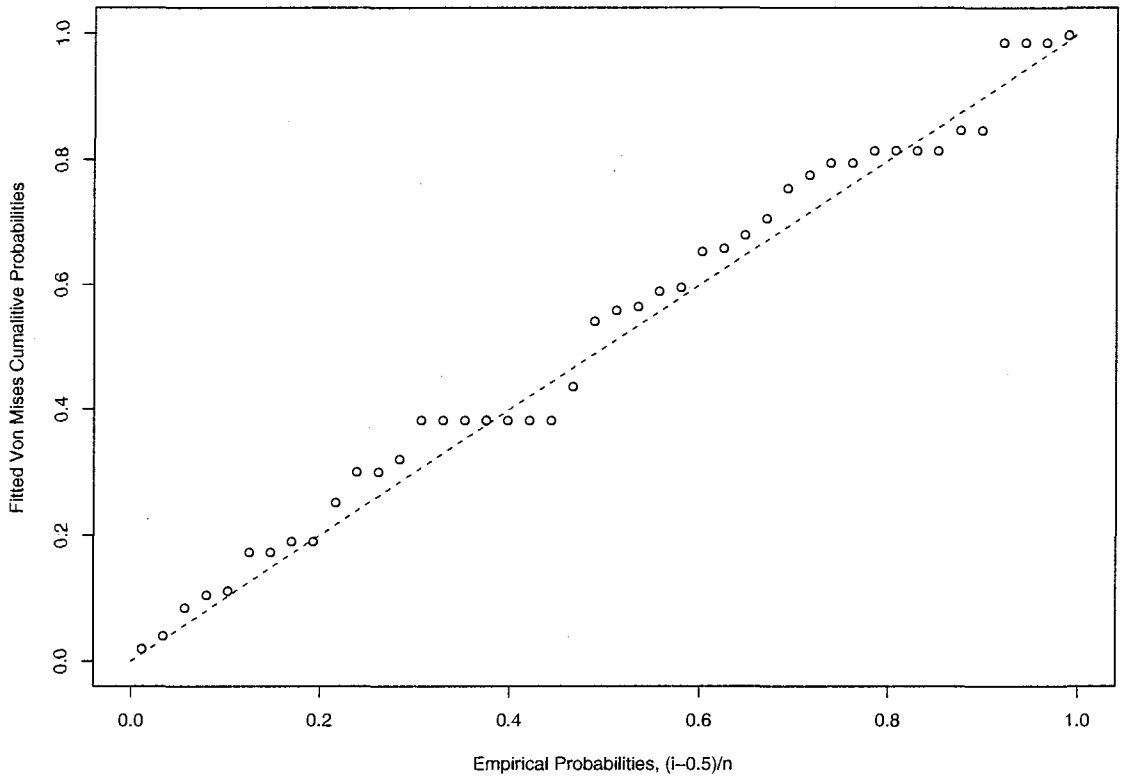


Figure 2.4: von Mises P-P plots of directional sandstone rock data (rotated by $-19.4°$)



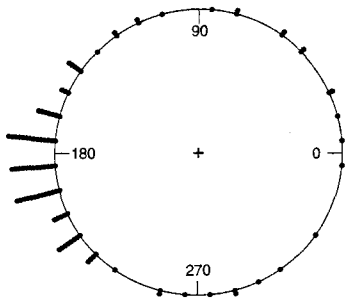
2.6 Example 2

For some data sets, the von Mises distribution does not sufficiently model the number of data points that are observed to fall far away from the mean direction. In this example we consider data collected from ants that were placed in an arena. An evenly illuminated black target was placed in a position centered at 180° and each ant was then placed individually into the arena and the optical orientation of the ant was recorded. The ants tended to run towards the illuminated black target. These data are taken from Fisher [7] (p. 243) and are a random sample of size 100 taken from Jander's larger data set in [9] (Jander's figure 18A). The data in [7] are grouped in the sense that the directions have been recorded to the nearest 10°. We do not yet have a method for analyzing grouped data with the mixture distribution. We have therefore adjusted Fisher's data by adding, to each angle given by Fisher, an independent random quantity uniformly distributed between −5° and 5°. For this data set, the grouping is not severe. It is therefore not expected that analyzing the grouped data instead would make a dramatic difference. The adjusted ant data, rounded to the nearest 0.1°, is given in Table 2.3 and is displayed graphically in the circular data plot in Figure 2.5.

Table 2.3: Orientations of 100 ants

1.9	12.4	28.1	28.9	41.5	46.0	55.5	56.6	72.6	75.5
86.1	109.6	111.0	115.3	123.3	127.6	139.6	140.8	142.0	147.5
147.7	149.8	150.3	154.1	160.0	161.9	162.1	162.4	162.7	163.1
163.7	168.1	170.2	170.4	171.9	172.2	172.5	175.4	175.6	175.7
176.5	177.1	177.7	179.0	179.4	179.7	180.6	180.7	180.7	181.1
181.7	182.0	183.8	184.1	185.3	188.5	188.8	189.0	189.8	192.6
193.9	194.9	195.5	195.7	195.9	196.0	196.2	196.4	196.6	198.0
199.2	199.8	202.4	202.9	204.8	206.7	207.6	210.5	210.9	212.4
212.5	213.1	214.8	218.0	219.6	220.6	220.7	224.5	228.2	232.4
254.5	255.0	266.0	277.4	282.9	289.4	295.4	301.1	326.4	354.9

Figure 2.5: Circular data plot of orientations of 100 ants



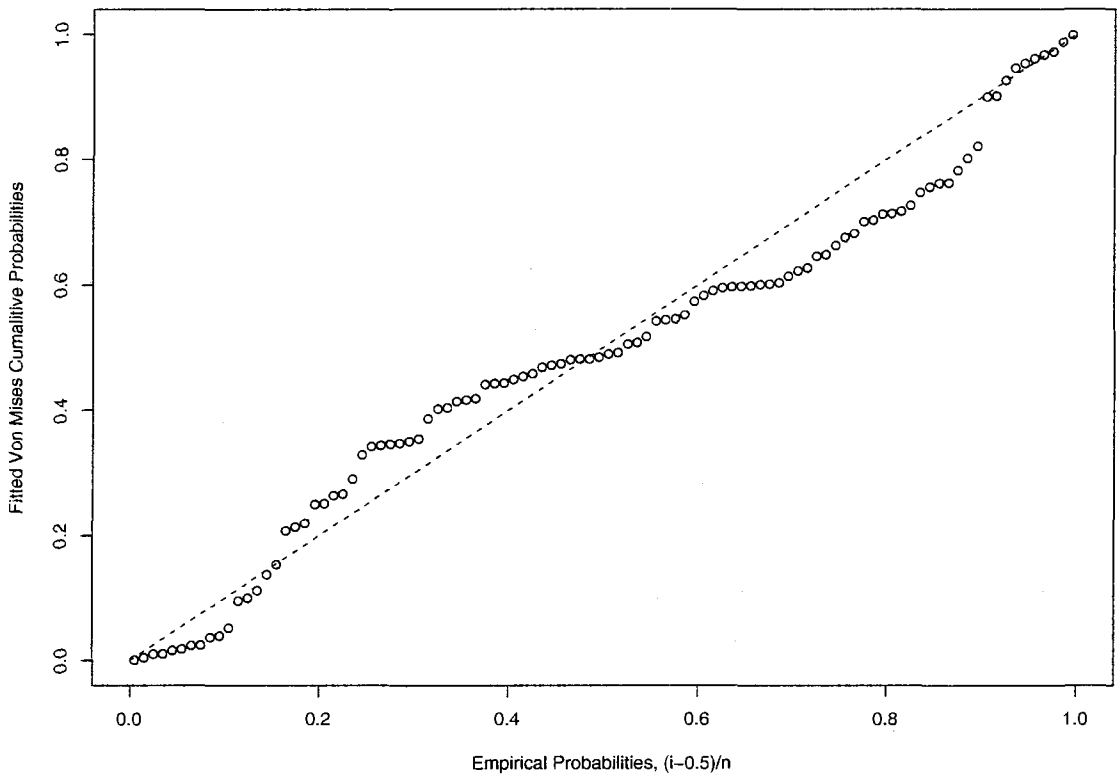
A von Mises model was fit to the adjusted ant data. The maximum likelihood parameter estimates and associated standard errors are provided in Table 2.4.

Table 2.4: Von Mises maximum likelihood parameter estimates for Example 2

$\hat{\mu}$	stderr($\hat{\mu}$)	$\hat{\kappa}$	stderr($\hat{\kappa}$)
183.3°	5.9°	1.55	0.21

The P-P plot in Figure 2.6 shows that the von Mises model is not a particularly good fit for the ant data since there a points both significantly above and below the $y = x$ line. While many of the ants appear to be heading in the approximate direction of the illuminated black target, there are also several ants heading in directions far away from the target. The von Mises model is not sufficient to capture both the concentration of the ants heading in directions that are close to the target and the frequency of ants that are heading in directions far away from the target. In Chapter 3 we will introduce a mixture model that contains both a von Mises and a uniform component. This more flexible model is more adequately able to model the ant data.

Figure 2.6: von Mises P-P plot of directional ant data



Chapter 3

The Mixture Distribution

In this chapter we consider mixture models for data sets where the von Mises distribution does not provide an adequate fit. An introduction to the mixture model is given in Section 3.1. A rough outline of a method for obtaining the maximum likelihood parameter estimates of the mixture distribution is provided in Section 3.2 and in Section 3.3 we give the asymptotic distribution and variance of these estimates. In Section 3.4 we revisit Example 2 from Section 2.6 and provide maximum likelihood estimates for the parameters of the mixture model along with P-P plots that suggest an improved fit is obtained when using the mixture model rather than the von Mises model.

3.1 Introduction

One possible way to explain the behavior of the ants shown in the circular data plot in Figure 2.5 of Section 2.6, is that some of them are influenced by the illuminated black target while others are not. Suppose an ant is influenced by the target with probability p and heads off in the general direction of the target according to a von Mises distribution. In addition, suppose the same ant has a probability $1-p$ of remaining uninfluenced by the target, in which case it heads off in a random direction. Then, the distribution of directions an ant will travel in is a mixture distribution and has probability density function given by

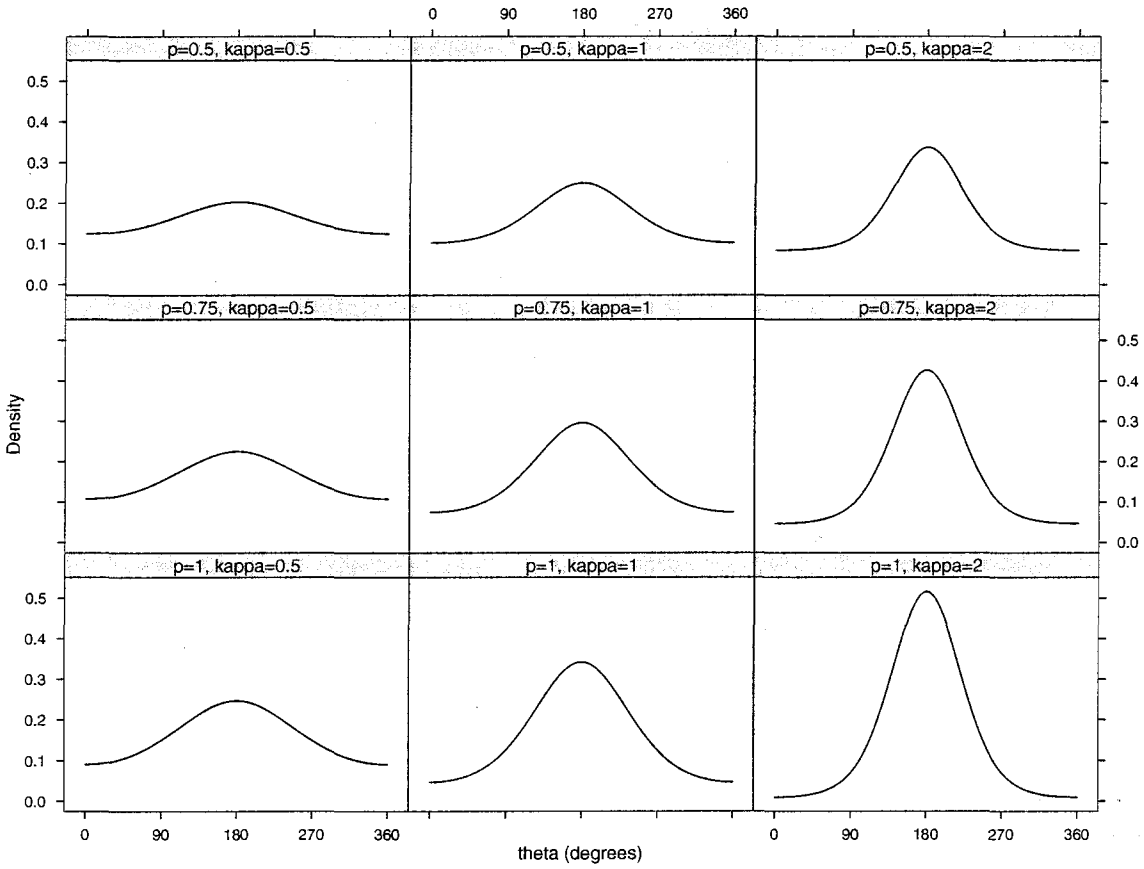
$$f(\theta; p, \mu, \kappa) = p\, f_{VM}(\theta; \mu, \kappa) + (1-p) f_U(\theta),\ 0 \leq \theta < 2\pi,\ 0 \leq p \leq 1,\ \kappa > 0,$$

where $f_{VM}(\theta; \kappa, \mu)$ is the von Mises density given in Section 2.1 and $f_U(\theta) = 1/2\pi$, is the uniform density. When $p = 1$ the model simplifies to the von Mises model discussed in

Chapter 2 and when $p = 0$ the model simplifies to a uniform model.

Several probability density functions of mixture distributions with various parameter values of p and κ have been plotted in Figure 3.1. In the figure, the columns correspond, from left to right, to $\kappa = 0.5, 1$ and 2 while the rows correspond, from top to bottom, to $p = 0.5, 0.75$ and 1. In each panel μ is 180°.

Figure 3.1: Probability density functions of various mixture distributions



Notice that an increase in either p or κ increases the density around the mean angular direction and decreases the density in the tails of the distribution. The last row of plots correspond to $p = 1$ and are von Mises distributions. One can also observe that the plot with parameters $p = 0.5$ and $\kappa = 2$ has a similar density around the mean angular direction as the von Mises density with parameters $p = 1$ and $\kappa = 1$. However, in the plot that has $p = 0.5$, the density drops off more rapidly as we move away from the mean angular direction and a higher density is left in the tails of the distribution. The mixture model can be a good alternative to the von Mises distribution in situations where it looks as though a portion of the data is von Mises distributed but a fitted von Mises distribution is not able to model adequately both the number of data points that are concentrated around the mean angular direction and the number of points that are in the tails of the distribution.

3.2 Maximum Likelihood Estimation

In this section we discuss a method for obtaining maximum likelihood parameter estimates for the mixture model. Unfortunately, the mixture distribution does not belong to the exponential family so we can not use the general method of obtaining maximum likelihood parameter estimates for exponential family distributions as can be done for the von Mises distribution.

We can obtain maximum likelihood parameter estimates for the mixture model by identifying the parameter values which maximize the likelihood function or equivalently maximize the log-likelihood function. Suppose that $\theta_1, \ldots, \theta_n$ are n independent random directions drawn from the mixture distribution above and we wish to estimate the parameters p, μ, and κ.

Let $\phi = (p, \mu, \kappa)$, then the log-likelihood function is given by

$$\begin{aligned} l(\phi) &= \log\left\{\prod_{i=1}^{n} f(\theta_i; p, \mu, \kappa)\right\} = \sum_{i=1}^{n} \log\{f(\theta_i; p, \mu, \kappa)\} \\ &= \sum_{i=1}^{n} \left[\log\left\{p f_{VM}(\theta_i; p, \mu, \kappa) + (1-p) f_U(\theta_i)\right\}\right]. \end{aligned}$$

The maximum of the log-likelihood function will either be on the boundary of the parameter space, or a point within the interior of the parameter space which is a relative maximum. In the latter case, the maximum will have first derivatives of l equal to $\mathbf{0}$.

The score function, $\boldsymbol{U}(\boldsymbol{\phi})$, is the vector of first derivatives of the log-likelihood function and can be useful for identifying potential relative maxima. It is given by

$$\boldsymbol{U}(\boldsymbol{\phi}) = \begin{bmatrix} \frac{\partial l(\boldsymbol{\phi})}{\partial p} \\ \\ \frac{\partial l(\boldsymbol{\phi})}{\partial \mu} \\ \\ \frac{\partial l(\boldsymbol{\phi})}{\partial \kappa} \end{bmatrix},$$

where

$$\begin{aligned} \frac{\partial l(\boldsymbol{\phi})}{\partial p} &= \sum_{i=1}^{n} \frac{f_{VM}(\theta_i;\mu,\kappa) - f_U(\theta_i)}{f(\theta_i;p,\mu,\kappa)}, \\ \frac{\partial l(\boldsymbol{\phi})}{\partial \mu} &= p\kappa \sum_{i=1}^{n} \frac{f_{VM}(\theta_i;\mu,\kappa)}{f(\theta_i;p,\mu,\kappa)} \sin(\theta_i - \mu), \quad \text{and} \\ \frac{\partial l(\boldsymbol{\phi})}{\partial \kappa} &= p \sum_{i=1}^{n} \frac{f_{VM}(\theta_i;\mu,\kappa)}{f(\theta_i;p,\mu,\kappa)} \{\cos(\theta_i - \mu) - A(\kappa)\}, \end{aligned}$$

are the first order partial derivatives of the log-likelihood function and where $A(\kappa) = I_1(\kappa)/I_0(\kappa)$; the j^{th} order modified Bessel function of the first kind, $I_j(\kappa)$, is defined in Section 2.2. We can identify one or potentially more critical points, by finding the solutions to the likelihood equations given by $\boldsymbol{U}(\boldsymbol{\phi}) = \mathbf{0}$. The solutions to the likelihood equations will be relative maxima if the second derivative matrix of the log-likelihood function is negative definite.

3.3 Large Sample Asymptotic Distribution of the MLE

Let $\hat{\boldsymbol{\phi}} = (\hat{p}, \hat{\mu}, \hat{\kappa})$ be the maximum likelihood parameter estimate for the vector of true parameters for the mixture distribution, $\boldsymbol{\phi}_0 = (p_0, \mu_0, \kappa_0)$. If $\boldsymbol{\phi}_0$ is an interior point of the parameter space, then standard theory of maximum likelihood estimators [4] (pp. 294-296) can be used to show that $\hat{\boldsymbol{\phi}}$ is asymptotically normally distributed,

$$\sqrt{n}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0) \sim N(\mathbf{0}, \boldsymbol{I}^{-1}),$$

where $\boldsymbol{I}$ denotes the Fisher information matrix, based on a single observation,

$$\boldsymbol{I} = E\left[-\boldsymbol{H}_1(\boldsymbol{\phi}_0)\right].$$

Here, $\boldsymbol{H}_1(\boldsymbol{\phi}_0)$, is the second derivative matrix of the logarithm of the mixture density function evaluated at the true parameter values, and is given by

$$\boldsymbol{H}_1(\boldsymbol{\phi}_0) = \left(\begin{array}{ccc} \frac{\partial^2 \log f(\theta;\boldsymbol{\phi})}{\partial p^2} & \frac{\partial^2 \log f(\theta;\boldsymbol{\phi})}{\partial p \partial \mu} & \frac{\partial^2 \log f(\theta;\boldsymbol{\phi})}{\partial p \partial \kappa} \\ \frac{\partial^2 \log f(\theta;\boldsymbol{\phi})}{\partial \mu \partial p} & \frac{\partial^2 \log f(\theta;\boldsymbol{\phi})}{\partial \mu^2} & \frac{\partial^2 \log f(\theta;\boldsymbol{\phi})}{\partial \mu \partial \kappa} \\ \frac{\partial^2 \log f(\theta;\boldsymbol{\phi})}{\partial \kappa \partial p} & \frac{\partial^2 \log f(\theta;\boldsymbol{\phi})}{\partial \kappa \partial \mu} & \frac{\partial^2 \log f(\theta;\boldsymbol{\phi})}{\partial \kappa^2} \end{array} \right) \Bigg|_{\boldsymbol{\phi}=\boldsymbol{\phi}_0}.$$

Numerical solutions can be obtained for $\boldsymbol{I}$ can be obtained by numerical integration of

$$\boldsymbol{I} = -\int_0^{2\pi} \boldsymbol{H}_1(\boldsymbol{\phi}_0) f(\theta;\boldsymbol{\phi}_0) d\theta.$$

There is, however, no real need to evaluate this integral because we can use the result that $\boldsymbol{V}(\hat{\boldsymbol{\phi}})/n$ is asymptotically equal to $\boldsymbol{I}$, where $\boldsymbol{V}(\hat{\boldsymbol{\phi}})$ is the observed information matrix evaluated at $\hat{\boldsymbol{\phi}}$. The observed information matrix can be expressed as $\boldsymbol{V}(\hat{\boldsymbol{\phi}}) = -\boldsymbol{H}_n(\hat{\boldsymbol{\phi}})$, where $\boldsymbol{H}_n(\hat{\boldsymbol{\phi}})$ is the hessian matrix (second derivative matrix of the log-likelihood function) and is given by

$$\boldsymbol{H}_n(\hat{\boldsymbol{\phi}}) = \left(\begin{array}{ccc} \frac{\partial^2 l(\boldsymbol{\phi})}{\partial p^2} & \frac{\partial^2 l(\boldsymbol{\phi})}{\partial p \partial \mu} & \frac{\partial^2 l(\boldsymbol{\phi})}{\partial p \partial \kappa} \\ \frac{\partial^2 l(\boldsymbol{\phi})}{\partial \mu \partial p} & \frac{\partial^2 l(\boldsymbol{\phi})}{\partial \mu^2} & \frac{\partial^2 l(\boldsymbol{\phi})}{\partial \mu \partial \kappa} \\ \frac{\partial^2 l(\boldsymbol{\phi})}{\partial \kappa \partial p} & \frac{\partial^2 l(\boldsymbol{\phi})}{\partial \kappa \partial \mu} & \frac{\partial^2 l(\boldsymbol{\phi})}{\partial \kappa^2} \end{array} \right) \Bigg|_{\boldsymbol{\phi}=\hat{\boldsymbol{\phi}}}.$$

Thus if $\boldsymbol{\phi}_0$ is an interior point of the parameter space, then $\hat{\boldsymbol{\phi}}$ will have an asymptotically multivariate normal distribution, with mean vector $\boldsymbol{\phi}_0$, and approximate variance covariance matrix $\boldsymbol{V}^{-1}(\hat{\boldsymbol{\phi}})$. If $\boldsymbol{\phi}_0$ is not an interior point in the parameter space, then $\boldsymbol{\phi}_0$ must either lie outside or on the boundary of the parameter space and the above limiting distributional theory does not hold. In the case where $\boldsymbol{\phi}_0$ is on the boundary of the parameter space, then either $p_0 = 0$ or $p_0 = 1$. If $p_0 = 0$, then the uniform model applies and there are no parameters to estimate. When, $p_0 = 1$, then the von Mises model applies, and we need only estimate μ and κ and the theory presented in Chapter 2 applies. In Chapter 5, tests are provided for making the decision whether $p_0 = 0$, $p_0 = 1$, or p_0 is somewhere between 0 and 1. It is recommended that these tests first be used to decide which of the three models (uniform, von Mises, mixture) is most appropriate.

For convenience, the second order partial derivatives of the log-likelihood function are provided below. For simplification, f_U, f_{VM}, and f will be used as shorthand for $f_U(\theta_i)$, $f_{VM}(\theta_i;\mu,\kappa)$, and $f(\theta_i;p,\mu,\kappa)$, respectively.

$$\begin{aligned}
\frac{\partial^2 l(\phi)}{\partial p^2} &= -\sum_{i=1}^{n}\left(\frac{f_{VM}-f_U}{f}\right)^2, \\
\frac{\partial^2 l(\phi)}{\partial \mu^2} &= p\kappa\sum_{i=1}^{n}\frac{f_{VM}}{f}\left\{\kappa\left(\frac{f_{VM}}{f}-1\right)\sin^2(\theta_i-\mu)-\cos(\theta_i-\mu)\right\}, \\
\frac{\partial^2 l(\phi)}{\partial \kappa^2} &= \sum_{i=1}^{n}\left(p\frac{f_{VM}}{f}\left[A(\kappa)\left\{\frac{1}{\kappa}-2\cos(\theta_i-\mu)\right\}-\sin(\theta_i-\mu)^2\right]\right. \\
&\qquad \left.-\left(p\frac{f_{VM}}{f}\right)^2\{\cos(\theta_i-\mu)-A(\kappa)\}^2\right), \\
\frac{\partial^2 l(\phi)}{\partial p\partial \mu} &= \kappa\sum_{i=1}^{n}\frac{f_{VM}}{f}\left\{1-p\frac{f_{VM}-f_U}{f}\right\}\sin(\theta_i-\mu), \\
\frac{\partial^2 l(\phi)}{\partial p\partial \kappa} &= \sum_{i=1}^{n}\frac{f_{VM}}{f}\left(1-p\frac{f_{VM}-f_U}{f}\right)\{\cos(\theta_i-\mu)-A(\kappa)\}, \\
\frac{\partial^2 l(\phi)}{\partial \mu\partial \kappa} &= p\sum_{i=1}^{n}\frac{f_{VM}}{f}\left[\kappa\left(1-p\frac{f_{VM}-f_U}{f}\right)\{\cos(\theta_i-\mu)-A(\kappa)\}+1\right]\sin(\theta_i-\mu).
\end{aligned}$$

The second order partial derivatives are symmetric so $\partial^2 l(\phi)/\partial\mu\partial p = \partial^2 l(\phi)/\partial p\partial\mu$, $\partial^2 l(\phi)/\partial\kappa\partial p = \partial^2 l(\phi)/\partial p\partial\kappa$, and $\partial^2 l(\phi)/\partial\kappa\partial\mu = \partial^2 l(\phi)/\partial\mu\partial\kappa$.

An algorithm for obtaining the maximum likelihood parameter estimates for the mixture distribution is provided in Section 4.6.

3.4 Revisiting Example 2 from Section 2.6

We now revisit the data that was collected for the directions chosen by 100 ants in response to an evenly illuminated black target and previously given as Example 2 in Section 2.6. We fit both the von Mises and mixture models and make a graphical comparison of fit using P-P plots. Maximum likelihood parameter estimates and their associated standard errors for each of the two models are provided in Table 3.1.

Table 3.1: Maximum likelihood parameter estimates for Example 2

Model	$\hat{p}$	stderr($\hat{p}$)	$\hat{\mu}$	stderr($\hat{\mu}$)	$\hat{\kappa}$	stderr($\hat{\kappa}$)
von Mises	-	-	183.3°	5.9°	1.55	0.21
mixture	0.646	0.065	185.5°	3.2°	7.34	1.86

Comparing the von Mises and mixture maximum likelihood estimates one can make several observations

1. The $\hat{\kappa}$ estimate is much higher for the mixture model than for the von Mises model. This is because, having the added flexibility of modelling a proportion of directions to be randomly dispersed around the circle, allows the model to more accurately reflect the tightness of observations that are dispersed around the mean direction in the von Mises component.

2. Only approximately 65% of ants travel in a direction that is influenced by the black illuminated target. The other 35% of the ants remain uninfluenced by the target.

3. The estimated mean direction, $\hat{\mu}$, differs little between the von Mises and mixture estimates. This is to be expected because $E_{\text{mixture}}(\boldsymbol{R}) = p_0 E_{VM}(\boldsymbol{R})$ and the expectant resultant vectors point in the same mean direction. In large samples, we therefore expect the estimated mean directions to be approximately equal for both models.

The P-P plots for the von Mises and mixture models are given in Figures 3.2 and 3.3. One can clearly see that the P-P plot in Figure 3.3 appears to fit fairly well along the line $y = x$ while there clearly seems to be some curvature in the P-P plot in Figure 3.2, indicating that the fit is not as good. Therefore the mixture data model appears to be the more appropriate model based on the P-P plots. The improved fit in the mixture model is obtained from its added flexibility of allowing many directions to be fairly tightly dispersed around the mean direction, while still being able to model sufficiently the proportion of points that are far away from the mean direction. A more formal assessment of the goodness-of-fit of the von Mises and mixture models using Watson's U^2 statistic is provided in Chapter 5.

Figure 3.2: von Mises P-P plot of directional ant data

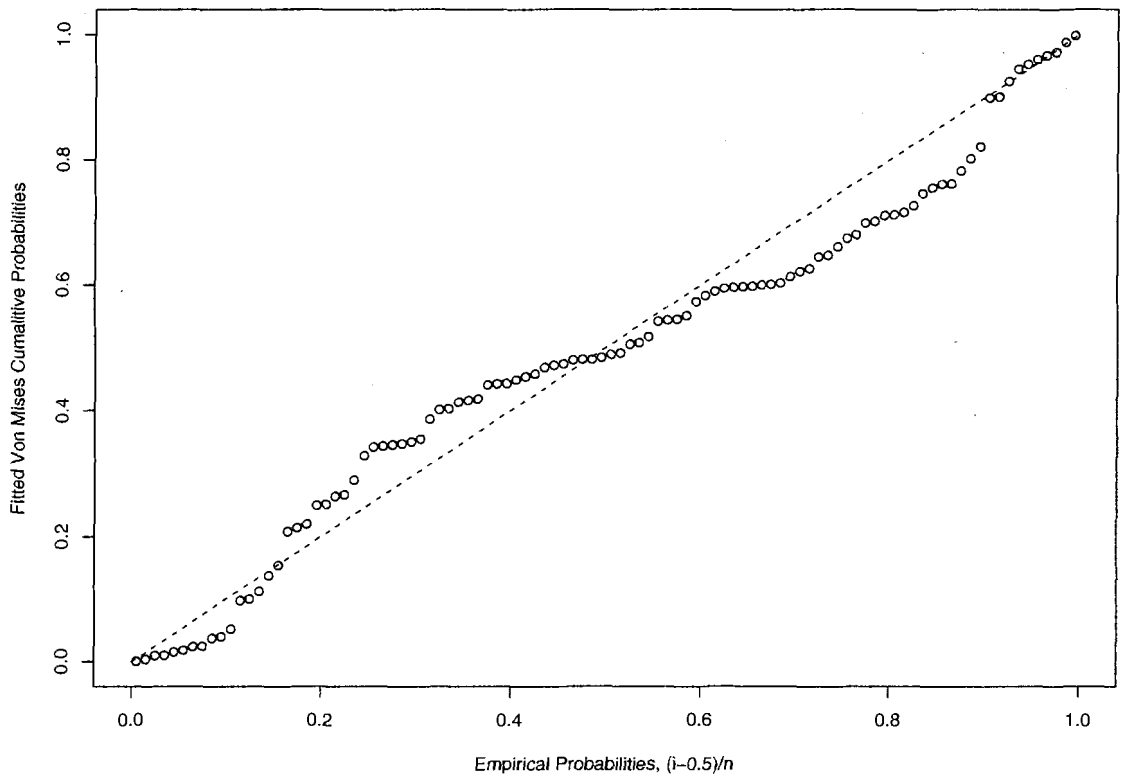
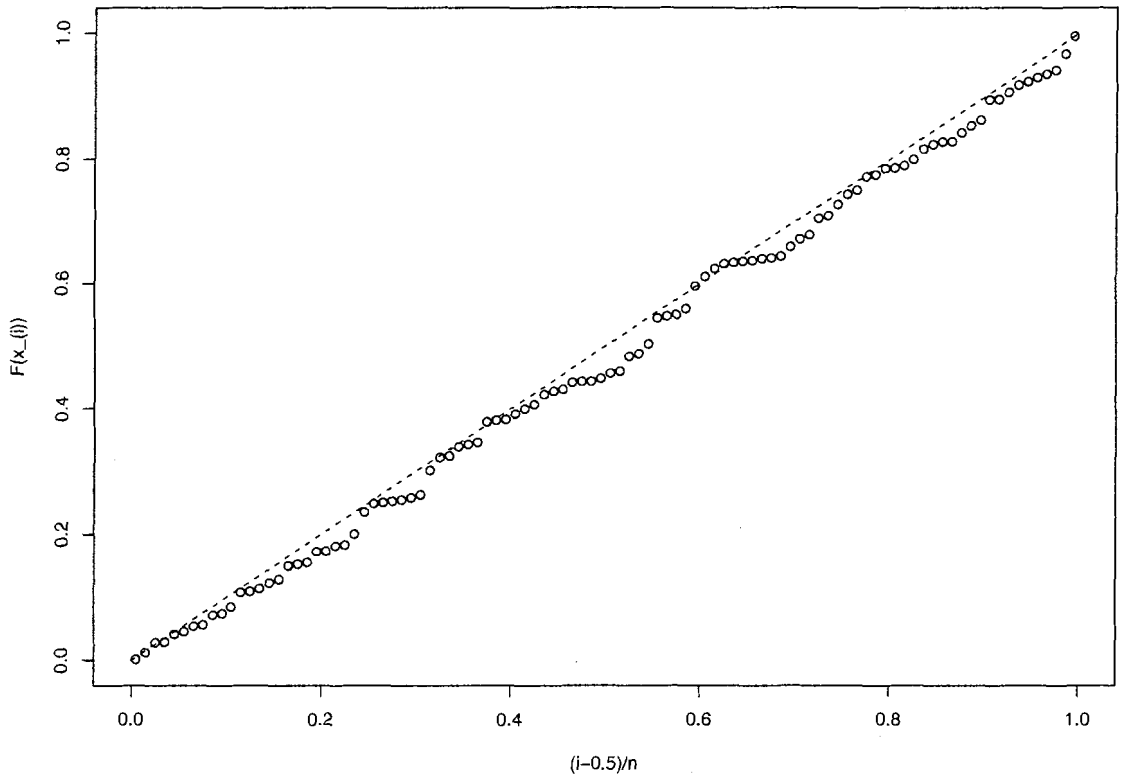


Figure 3.3: Mixture P-P plot if directional ant data



Chapter 4

Computational Details

In this chapter, we discuss some of the computational details that were used in obtaining maximum likelihood estimates for the von Mises and mixture distribution. In Section 4.1 we provide and algorithm for obtaining maximum likelihood estimates for the von Mises distribution. A characteristic of the mixture distribution having ill-behaved and unbounded likelihoods when values of κ are very large is discussed in Section 4.2. An example is provided in Section 4.3 of a situation where the maximum likelihood estimate for p under the mixture model is greater than 1 and falls outside the allowable parameter space. In Section 4.4 we elaborate on this phenomena in a little more detail, discussing what values of p and κ are required for $f(\theta; p, \mu, \kappa)$ to be a valid probability density function. We discuss how to identify situations, where higher likelihoods exist for values of $p > 1$ in Section 4.5. In Section 4.6 we discuss how to obtain initial approximations for the parameters of the mixture distribution. An algorithm for obtaining the MLE for the mixture distribution is presented in Section 4.7.

4.1 Algorithm for von Mises Maximum Likelihood Estimation

Prior to discussing the algorithm for obtaining maximum likelihood estimates for the von Mises distribution, some details from Chapter 2 are required. The von Mises density along with a description of its parameters is given in Section 2.1 and maximum likelihood estimation for the von Mises distribution is covered in Section 2.2. Definitions for the length of the resultant vector, R, the mean angular direction, $\bar{\theta}$, and the j^{th} order modified Bessel function of the first kind, $I_j(\kappa)$, are also provided in Section 2.2.

The maximum likelihood estimate for μ is given by $\hat{\mu} = \bar{\theta}$ and the maximum likelihood estimate for κ is implicitly given by $A(\hat{\kappa}) = R/n$, where $A(\kappa) = I_1(\kappa)/I_0(\kappa)$. Equivalently, the maximum likelihood estimate for κ can be found by differentiating the log-likelihood function with respect to κ and solving the likelihood equation,

$$U_{VM}(\kappa) = \frac{\partial l_{VM}(\hat{\mu}, \kappa)}{\partial \kappa} = R - nA(\kappa) = 0.$$

By using the relations

$$\frac{d}{d\kappa} I_0(\kappa) = I_1(\kappa) \quad \text{and} \quad \frac{d}{d\kappa} I_1(\kappa) = I_0(\kappa) - \frac{I_1(\kappa)}{\kappa},$$

the derivative of U_{VM}, can be seen to be

$$H_{VM}(\kappa) = \frac{d}{d\kappa} U_{VM}(\kappa) = n\left\{A(\kappa)^2 + A(\kappa)/\kappa - 1\right\}.$$

Let $\bar{R} = R/n$. A good initial approximation, $\hat{\kappa_0}$, for $\hat{\kappa}$ is given by Fisher [7](p.88),

$$\hat{\kappa_0} = \begin{cases} 2\bar{R} + \bar{R}^3 + 5\bar{R}^5/6 & \bar{R} < 0.53 \\ -0.4 + 1.39\bar{R} + 0.43/(1 - \bar{R}) & 0.53 \leq \bar{R} < 0.85 \\ 1/(\bar{R}^3 - 4\bar{R}^2 + 3\bar{R}) & \bar{R} \geq 0.85. \end{cases}$$

An algorithm for obtaining maximum likelihood estimates from the von Mises distribution is given below.

Algorithm

Step 1: Calculate $\bar{\theta}$, and R as given in Section 2.2.

Step 2: Calculate $\hat{\kappa_0}$ as previously described.

Step 3: Set $\hat{\mu} = \tilde{\theta}$ and $\hat{\kappa} = \text{Newton-Raphson}(\hat{\kappa_0}, U_{VM}, H_{VM})$.

Step 4: Calculate the estimated variances for $\hat{\mu}_{VM}$ and $\hat{\kappa}_{VM}$ as

$$\hat{\text{Var}}(\hat{\mu}) = \frac{1}{n\hat{\kappa}A(\hat{\kappa})}, \quad \text{and} \quad \hat{\text{Var}}(\hat{\kappa}) = \frac{1}{n[1 - A(\hat{\kappa})^2 - A(\hat{\kappa})/\hat{\kappa}]}.$$

The Newton-Raphson algorithm used in step 3 is provided in Appendix A.

4.2 Behavior of Mixture Likelihood for Large Values of κ

The mixture distribution was previously introduced in Section 3.1 along with a description of its parameters, p, μ and κ. The log-likelihood function for the mixture distribution is given in Section 3.2. As with many mixture models, the likelihood function is ill-behaved for values of $0 < p < 1$ when a concentration parameter, κ in the von Mises case, is large. If we allow the values of κ to approach infinity, then the log-likelihood function will also approach infinity for values of μ that are equal to one of the data points. To illustrate this characteristic of the mixture distribution a little better, a simple example is illustrated below.

Consider a situation in which we have only 3 angular data points: 135°, 180°, and 225°. A contour plot of the log-likelihood function surface when p is fixed at 1/3 and κ and μ are allowed to vary is given in Figure 4.1. Figure 4.2, plots the same data but is a 3-dimensional plot and provides a different view of the log-likelihood function surface.

As can be seen in Figures 4.1 and 4.2, the log-likelihood function is highest for large values of κ at values of μ corresponding to each of the 3 angular data points. As previously mentioned, if κ is allowed to increase indefinitely, then the log-likelihood function will approach infinity at values of μ that are equal to any of the 3 angular data points.

Figure 4.1: Contour plot of log-likelihood function when p is fixed at $1/3$

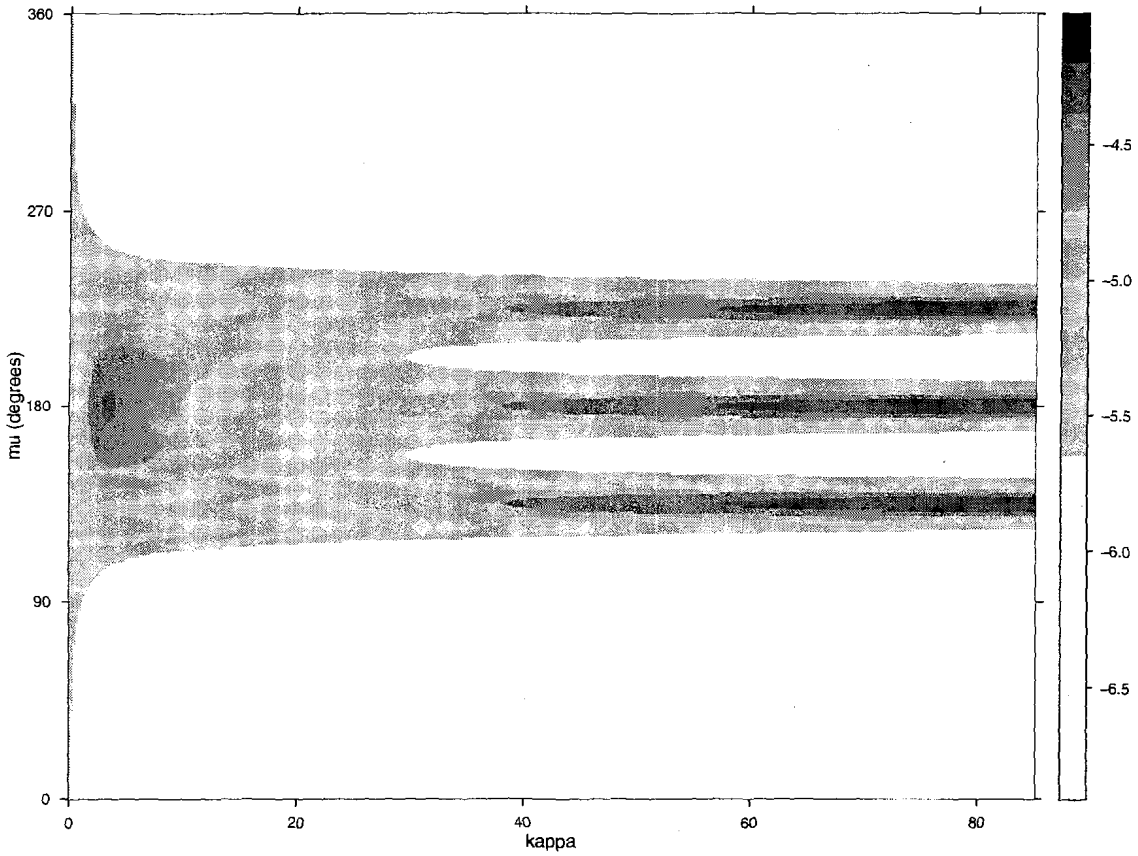
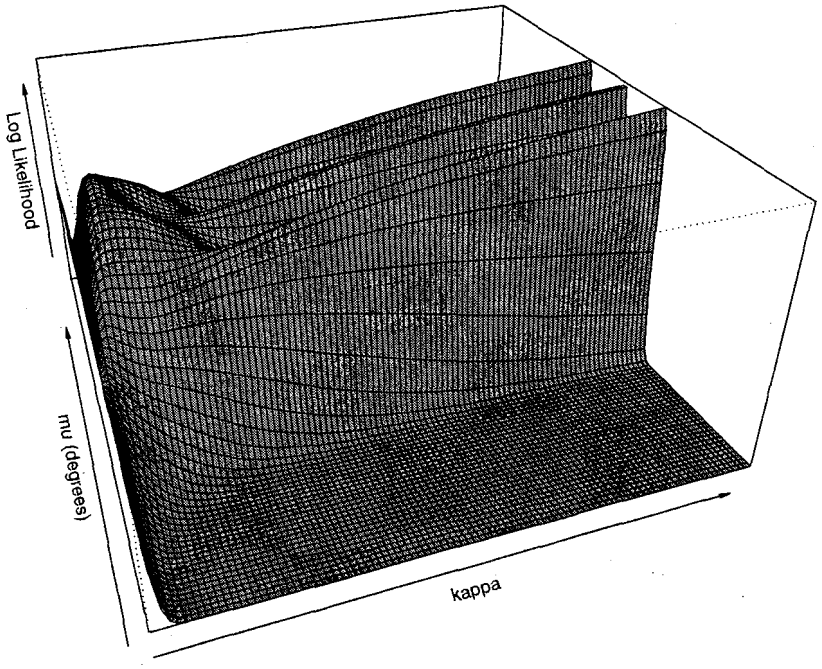


Figure 4.2: 3-dimensional plot of log-likelihood function when p is fixed at $1/3$



Although the parameter p was set to $1/3$ in the plots, the same ridges are present for every value of $0 < p < 1$. A simple example was chosen with only 3 data points so that the plots do not become too cluttered but the same concept applies to problems with more data. It is not particularly reasonable to consider that the true parameter value for κ is very large and the true parameter value for μ is most likely to be arbitrarily any one of the data points. Most data problems for which the mixture distribution is a natural choice do however have a local maximum of the log-likelihood function at a point that is a solution to the likelihood equations given in Section 3.2. Fortunately, the log-likelihood function is well behaved near the local maximum and standard maximum likelihood theory can be applied.

If an algorithm such as the Newton-Raphson algorithm is used to obtain the solution to the likelihood equations, then some care needs to be taken with the initial choice of parameter values. As long as the initial values are close enough to the solution of the likelihood equations as to be in a region near the solution, the Newton-Raphson algorithm should work well. If however, starting values are far away from the solution or chosen in a region where the log-likelihood function is ill-behaved, the Newton-Raphson algorithm will not converge. Typically one should be cautious using large initial values for the κ parameter, particularly when the initial values for the p parameter are small.

4.3 Example of MLE for p that is Greater than 1

The mixture model is a natural choice in situations where it appears as though there is both a good proportion of von Mises data that is clustered fairly tightly around the mean and there is also a significant proportion of uniform data dispersed throughout the rest of the circle. If the data is not tightly clustered around the mean and thus has a low value of κ, then there is little difference between the von Mises and uniform distributions. In such situations, the mixture model has less appeal than using either the von Mises or the uniform distributions. For most problems for which the mixture model is a natural choice, the likelihood function has a local maximum for some value of p between 0 and 1 and a reasonable value of κ. However, in the example below, the mixture distribution is not a natural fit for the data and there is no solution to the likelihood equations within the allowable parameter space.

4.3.1 Example 3

The counts of the number of births of children born with anecephalitis in Birmingham, England were recorded for the years 1940-1947. These data were originally obtained from Edwards in [6] and tested for discrete uniformity using Watson's U^2 statistic by Choulakian, Lockhart and Stephens in [3]. The p-value for Choulakian, Lockhart and Stephens' test is 0.031 and thus the null hypothesis of these data belonging to a discrete uniform distribution is marginally rejected at the typically used 0.05 significance level. Although there does seem to be significant evidence against these data belonging to a discrete uniform distribution, the evidence is not overwhelming.

The count data for each month is given in Table 4.1. Since monthly data is recorded for each birth rather than the exact day and time, the data is discrete.

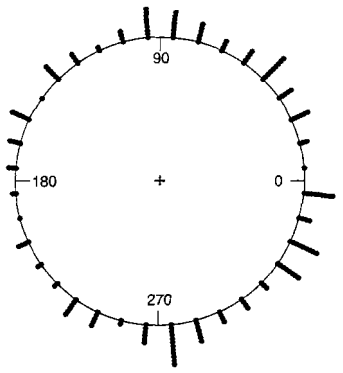
Table 4.1: Counts of births of children born with anecephalitis

Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.
10	19	18	15	11	13	7	10	13	23	15	22

Counts of monthly data are often displayed around a circle divided in 12 sectors. Prior to analyzing the data each birth was assigned an angle between 0° and 360°, based on the months that the births were recorded in. Each of the 12 sectors in the circle corresponds to a 30° slice with the months January to December assigned the slices 0° to 30°, ..., 330° to 360° respectively. Note that Choulakian, Lockhart and Stephens conducted the goodness-of-fit test for uniformity using discrete birth count data. Since we have not yet developed the theory yet for extending the mixture model to discrete data we use a simplistic transformation of the discrete birth count data to a continuous scale. Each birth is randomly assigned an angle within the interval corresponding to the month the birth was in.

The data are displayed in the circular data plot in Figure 4.3.

Figure 4.3: Circular data plot of anecephalitis birth count data



Visually this data do not seem to lend itself naturally to the mixture model as a good choice since the data do not seem tightly concentrated around any particular point in the circle. The von Mises maximum likelihood parameter estimates for the anecephalitis birth count data are $\hat{\mu} \approx 1.52°$, and $\hat{\kappa} \approx 0.288$. The value for κ is quite close to 0. Recall from Chapter 2 that the von Mises distribution is equivalent to the circular uniform distribution when $\kappa = 0$, so this is another indication that there is little difference between the von Mises and uniform models.

Unlike in most problems where the mixture distribution is a natural choice for the data, we do not have a local maximum for the likelihood function that is a solution to the likelihood equations in this example. Figure 4.4 shows the maximized profile log-likelihood function for the mixture distribution when the parameter p is held fixed and the likelihood function is maximized with respect to the other 2 parameters. As can be seen in the figure, the maximum likelihood estimate for p is undefined since the maximized profile likelihood function asymptotically approaches its maximum as p approaches infinity.

Figures 4.5 and 4.6 show how the restricted maximum likelihood estimates change for κ and μ when the profile likelihood function is maximized at different fixed levels of p.

Figure 4.4: Profile log-likelihood function for mixture distribution when p is held fixed

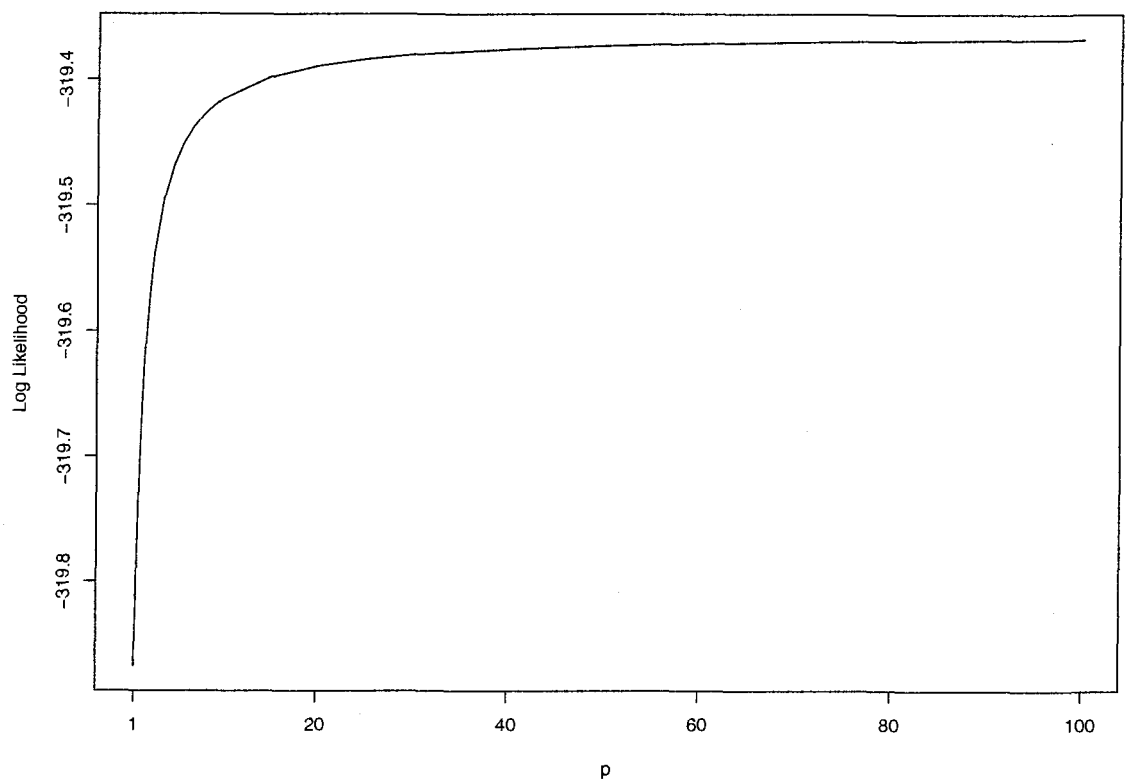


Figure 4.5: Values of κ that maximize profile likelihood for fixed values of p

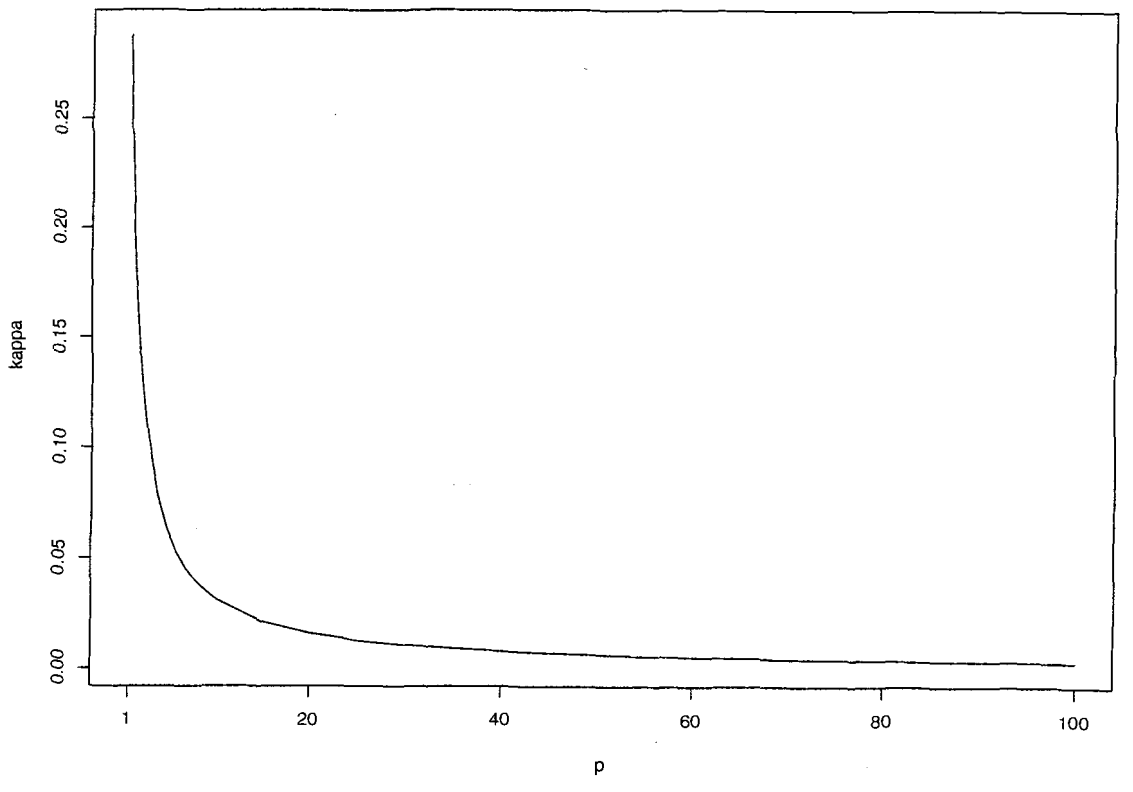
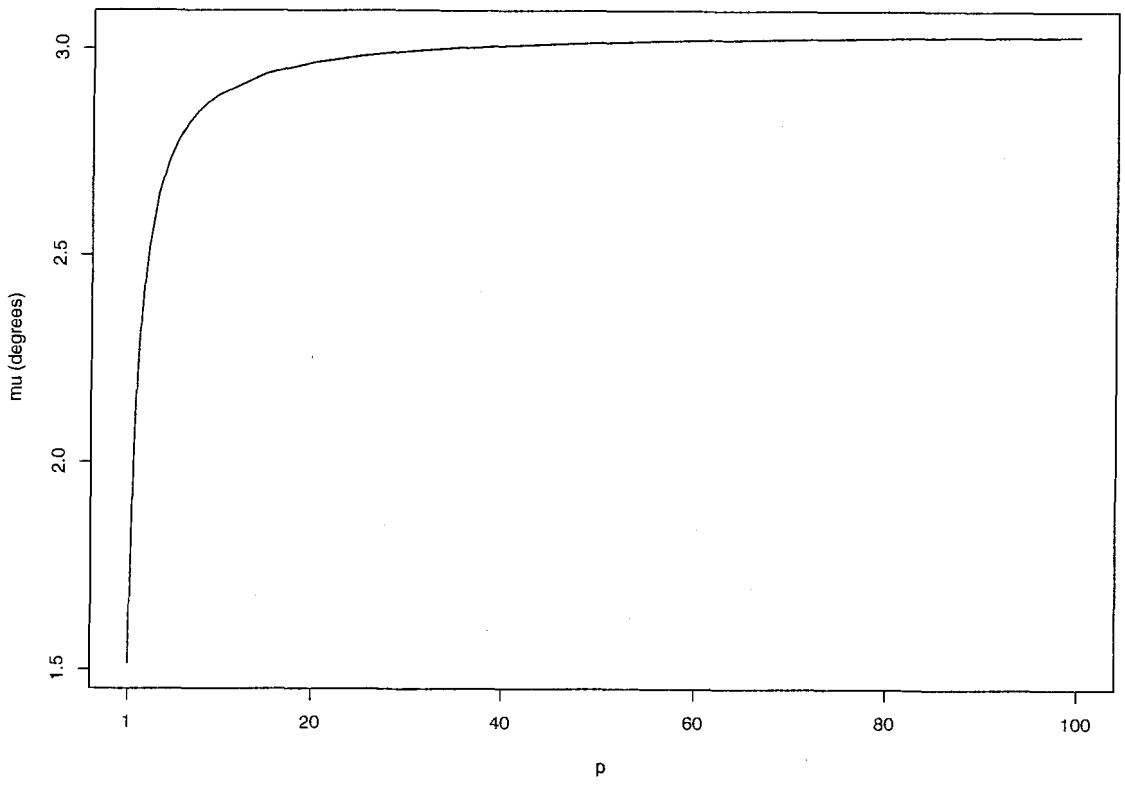


Figure 4.6: Values of μ that maximize profile likelihood for fixed values of p



In addition to not fitting in with out initial motivation for constructing the mixture model, it is not convenient to permit the parameter estimates for p to approach ∞ and κ to approach 0. As such, it is better to either to go with one of the simpler uniform or von Mises models (provided there is a good fit) or to consider other models which fall outside the scope of this paper. Model selection, including discussion of goodness-of-fit based tests, are considered in Chapter 5.

4.4 Further Discussion of Values of p that Fall Outside Parameter Space

Although it does not fit with our original motivation for using the mixture model, values of p that are greater than 1 can still be legitimate distributions provided that the corresponding values of κ are sufficiently small to keep the density positive everywhere from 0 to 2π. Recall that the probability density function for the mixture distribution is

$$f(\theta; p, \mu, \kappa) = p\, f_{VM}(\theta; \mu, \kappa) + (1-p) f_U(\theta),\ 0 \leq \theta < 2\pi,\ 0 \leq p \leq 1,\ \kappa > 0,$$

where $f_{VM}(\theta; \kappa, \mu)$ is the von Mises density and $f_U(\theta)$, is the uniform density.

In order for $f(\theta; p, \mu, \kappa)$ to be a continuous probability density function, it must satisfy both of the following properties:

1.
$$f(\theta; p, \mu, \kappa) \geq 0, \quad \forall\, \theta \in [0, 2\pi).$$

2.
$$\int_0^{2\pi} f(\theta; p, \mu, \kappa) d\theta = 1.$$

The first property is satisfied as long as values of p and κ satisfy the following equation,

$$\begin{aligned} f(\theta; p, \mu, \kappa) &\geq 0, \\ p\, f_{VM}(\theta; \mu, \kappa) + (1-p) f_U(\theta) &\geq 0, \\ p\, \frac{\exp\{\kappa \cos(\theta - \mu)\}}{2\pi\, I_0(\kappa)} + (1-p)\frac{1}{2\pi} &\geq 0, \\ p\, \frac{\exp\{\kappa \cos(\theta - \mu)\}}{I_0(\kappa)} + (1-p) &\geq 0. \end{aligned}$$

Note that $\cos(\theta - \mu) \geq -1, \forall\, \theta \in [0, 2\pi)$.

Thus,

$$p\,\frac{\exp(-\kappa)}{I_0(\kappa)} + (1-p) \geq 0,$$

and it is then easily shown that

$$p \leq \frac{I_0(\kappa)}{I_0(\kappa) - \exp(-\kappa)}.$$

Maximum allowable values of p for corresponding values of κ so that $f(\theta; p, \mu, \kappa)$ is still a valid probability density function are given in Table 4.2.

Table 4.2: Maximum allowable values of p for corresponding values of κ

κ	p
10	1.000
1	1.410
0.1	10.265
0.01	100.252
0.001	1000.250

It is straightforward to show the second property is always satisfied by noting that both $f_{VM}(\theta; \mu, \kappa)$ and $f_U(\theta)$ are probability density functions and thus

$$\begin{aligned}\int_0^{2\pi} f(\theta; p, \mu, \kappa)d\theta &= p\int_0^{2\pi} f_{VM}(\theta; \mu, \kappa)d\theta + (1-p)\int_0^{2\pi} f_U(\theta)d\theta \\ &= p + (1-p) = 1.\end{aligned}$$

4.5 Identification when the MLE for p is Outside Parameter Space

To avoid unnecessary computation, it is useful to be able to identify situations in which higher likelihoods exist for values of $p > 1$ than for $0 \leq p \leq 1$. It is relatively easy to identify these situations by simply examining the partial derivative of the log-likelihood function with respect to the parameter p, evaluated at $p = 1$ and the other parameters, μ and κ equal to their von Mises MLE values. If this derivative if positive, then higher likelihoods do exist for values of $p > 1$. We do not want to consider values of $p > 1$ out of practical considerations, so the parameter estimate for p can then simply be set to 1 and the parameter estimates for μ and κ can be set to their von Mises MLE values.

4.6 Initial Parameter Estimates for Mixture Distribution

As mentioned in Section 4.2, initial parameter values that are too far away from the solution to the likelihood equations can result in problems for the Newton-Raphson algorithm converging on the correct solution. Convergence can be particularly problematic if the initial parameter value for p is significantly underestimated and the initial parameter value for κ is significantly overestimated. Thus some care needs to be taken in choosing initial parameter values that are reasonable enough to result in the Newton-Raphson algorithm performing as desired.

To avoid confusion with mixture distribution parameter estimates, we will refer to the maximum likelihood parameter estimates for the von Mises model as $\hat{\mu}_{VM}$ and $\hat{\kappa}_{VM}$ throughout the rest of this chapter.

In Section 3.4 it was mentioned that we expect the parameter estimates for the mean direction parameter, μ to be approximately equal for both the von Mises and mixture models. Thus a reasonable initial estimate for the mean direction of the mixture distribution is $\hat{\mu_0} = \hat{\mu}_{VM}$.

A simple way of obtaining an initial estimate for the proportion, p, of data that is von Mises distributed is to first count the number, n_q, of points in the quadrant of the circle that is centered 180° away from $\hat{\mu_0}$. As long as the concentration parameter of the von Mises component is reasonably large, we expect that most of these points will be from the uniform component. Since n_q counts only the number of points in one of the four quadrants of the circle, a reasonable initial estimate for $q = 1 - p$ is

$$\hat{q_0} = \frac{4n_q}{n}.$$

An initial estimate for p is then

$$\hat{p_0} = 1 - \hat{q_0} = \frac{n - 4n_q}{n}.$$

Since the uniform data is mixed with von Mises data, some points we count as being uniform distributed may in fact be von Mises rather than uniformly distributed. Thus our estimate $\hat{q_0}$ is positively biased and our estimate $\hat{p_0}$ is negatively biased. However, as long as the concentration parameter, κ, for the von Mises component is reasonably large, this bias won't be too extreme. Under certain circumstances there is still the risk that our estimate for p will be low to such an extent as to result in convergence problems. In an extreme example,

$\hat{p_0}$ could even potentially be negative. It is therefore a good idea not to allow $\hat{p_0}$ to be less than 0.5. Thus our initial estimate for p becomes

$$\hat{p_0} = \max\left(0.5, \frac{n - 4n_q}{n}\right).$$

The calculation of an initial estimate for p is illustrated in Figure 4.7 using the ant data from Example 2 in Section 2.6. As can be seen in the figure there are $n_q = 8$ points in the quadrant of the circle that is centered 180° away from $\hat{\mu_0}$. Thus our initial estimate for p becomes,

$$\hat{p_0} = \frac{100 - 4(8)}{100} = 0.68.$$

This is reasonably close to the final maximum likelihood estimate for the parameter p, namely, $\hat{p} \approx 0.646$.

If the parameter, p, for the proportion of von Mises distributed data is close to 1, then $\hat{\kappa}_{VM}$ would be a reasonable initial estimate for the concentration parameter, κ. Typically, $\hat{\kappa}_{VM}$ underestimates κ because inclusion of the uniform component in the model tightens the concentration of the remaining von Mises component. This is, however, not a serious problem because using initial parameter estimates for κ that are too small does not typically put us in a region where convergence would not be possible as could be the case if κ were significantly overestimated as discussed previously in Section 4.2. We can still improve a little on this crude initial estimate by maximizing the profile likelihood when p is fixed at $\hat{p_0}$ and μ is fixed at $\hat{\mu_0}$ to obtain a better initial estimate for κ.
Our initial estimate for κ then becomes

$$\hat{\kappa_0} = \text{Newton-Raphson}(\hat{\kappa}_{VM}, U_\kappa, H_\kappa),$$

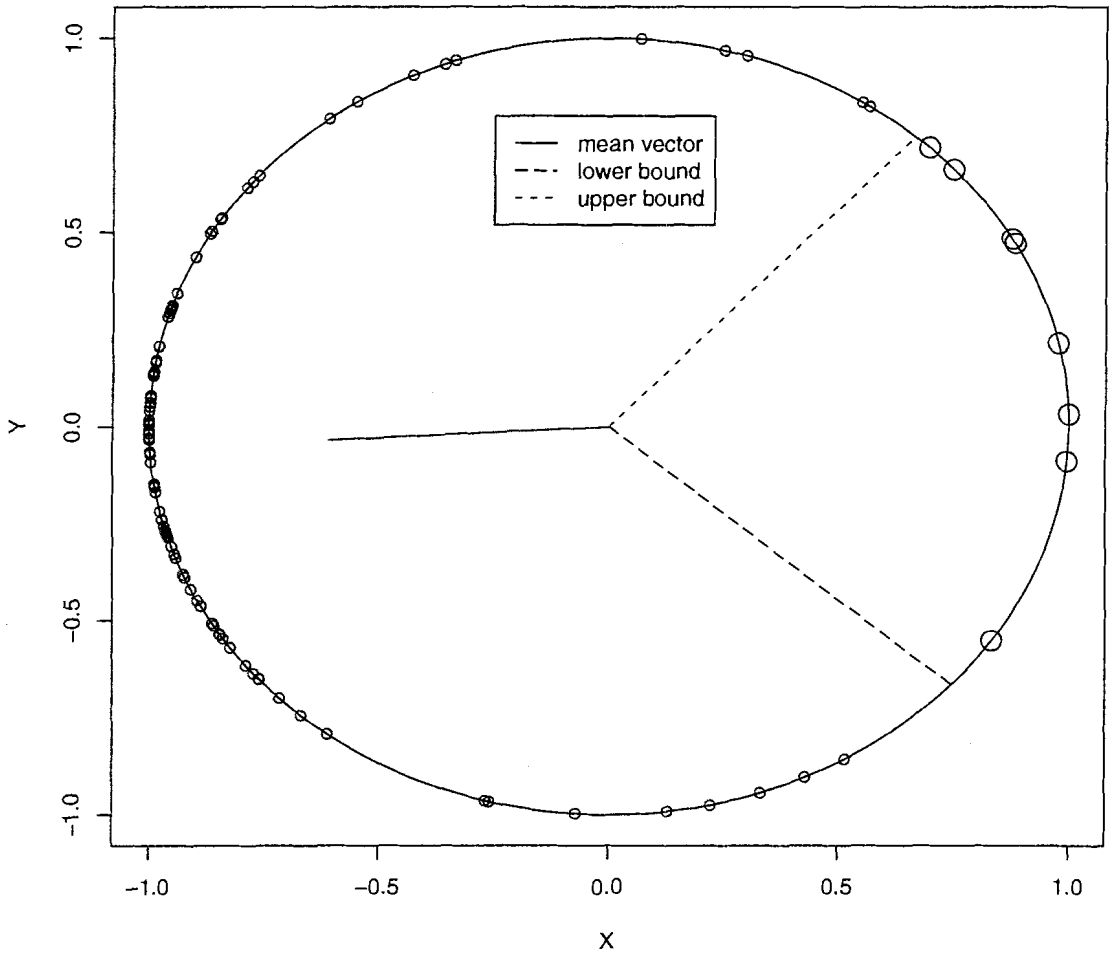
where $U_\kappa(\kappa)$ and $H_\kappa(\kappa)$ are given by

$$U_\kappa(\kappa) = \frac{\partial l(p_0, \mu_0, \kappa)}{\partial \kappa} = p_0 \sum_{i=1}^{n} \frac{f_{VM}(\theta_i; \mu_0, \kappa)}{f(\theta_i; p_0, \mu_0, \kappa)} \{\cos(\theta_i - \mu_0) - A(\kappa)\}$$

and

$$\begin{aligned} H_\kappa(\kappa) = \frac{\partial U_\kappa(\kappa)}{\partial \kappa} = \sum_{i=1}^{n} \Bigg(& p_0 \frac{f_{VM}(\theta_i; \mu_0, \kappa)}{f(\theta_i; p_0, \mu_0, \kappa)} \left[A(\kappa) \left\{ \frac{1}{\kappa} - 2\cos(\theta_i - \mu_0) \right\} - \sin(\theta_i - \mu_0)^2 \right] \\ & - \left(p_0 \frac{f_{VM}(\theta_i; \mu_0, \kappa)}{f(\theta_i; p_0, \mu_0, \kappa)} \right)^2 \{\cos(\theta_i - \mu_0) - A(\kappa)\}^2 \Bigg). \end{aligned}$$

Figure 4.7: Illustration of the calculation of an initial estimate for p



4.7 Algorithm for Mixture Maximum Likelihood Estimation

Now that many of the components are in place from discussion in the previous sections, the calculation of maximum likelihood estimates for the mixture distribution is fairly straightforward. Let $\hat{\phi} = (\hat{p}, \hat{\mu}, \hat{\kappa})$ be the maximum likelihood estimate for $\phi = (p, \mu, \kappa)$. The algorithm for the calculation of $\hat{\phi}$ is given below.

Algorithm

Step 1: Calculate the von Mises maximum likelihood estimates, $\hat{\mu}_{VM}$ and $\hat{\kappa}_{VM}$ as described in Section 4.1.

Step 2: Evaluate

$$\frac{\partial l(p, \hat{\mu}_{VM}, \hat{\kappa}_{VM})}{\partial p} = \sum_{i=1}^{n} \frac{f_{VM}(\theta_i; \hat{\mu}_{VM}, \hat{\kappa}_{VM}) - f_U(\theta_i)}{f(\theta_i; p, \hat{\mu}_{VM}, \hat{\kappa}_{VM})}, \quad \text{at} \quad p = 1.$$

Step 3: If the derivative in Step 2 is positive, then no solution exists to likelihood equations for $0 \leq p \leq 1$ and the restricted range likelihood function is at its maximum when $p = 1$ and the other parameters are at their von Mises MLE values so set $\hat{\boldsymbol{\phi}} = (1, \hat{\mu}_{VM}, \hat{\kappa}_{VM})$.

Step 4: Else set

$$\hat{\boldsymbol{\phi}} = \text{Newton-Raphson}(\hat{\boldsymbol{\phi}}_0, \boldsymbol{U}, \boldsymbol{H}_n),$$

where $\hat{\boldsymbol{\phi}}_0 = (\hat{p_0}, \hat{\mu_0}, \hat{\kappa_0})$ is the set of initial parameter values which were provided in Section 4.6, and $\boldsymbol{U}(\boldsymbol{\phi})$ and $\boldsymbol{H}_n(\boldsymbol{\phi})$ are given in Section 3.2.

Step 5: Obtain the estimated variance covariance matrix for $\hat{\boldsymbol{\phi}}$ using,

$$\hat{\text{Var}}(\hat{\boldsymbol{\phi}}) = \left\{ -\boldsymbol{H}_n(\hat{\boldsymbol{\phi}}) \right\}^{-1}.$$

It is important to mention that while the initial values provided in Section 4.6 are usually sufficiently close to the maximum likelihood estimates for the Newton-Raphson algorithm to converge (about 95% of the time for bootstrap samples generated for Examples 1 and 2 from Chapter 2), there can be some exceptions. If convergence cannot be obtained, there are more advanced algorithms such as Powell's Dog Leg algorithm and the Levenberg-Marquardt algorithm that ensure that the step sizes are taken to be sufficiently small so that the sum of the squared components of the score function decreases (typically ensuring that the log-likelihood function increases) with each step. These algorithms can be found in [14], for example.

In some rare cases, the combination of the initial values in Section 4.6 with the more advanced algorithms is still not sufficient to obtain the solution to the likelihood equations in a reasonable number of iterations. Six different sets of initial values that are used to try to obtain solutions to the likelihood equations are listed below:

1. Initial values from Section 4.6.
2. Parameters of the distribution that generated the sample (if known).
3. von Mises maximum likelihood estimates, and $\hat{p} = 1$.
4. Starting with initial values from 3, maximize the likelihood function with respect p, κ, and μ, one parameter at a time, holding the other two parameters fixed. Repeat up to 50 times or until the maximum absolute value of the components of the score function is 0.001.
5. Using the initial value for μ from 4, perform a 100×100 grid search for values of p between 0 and 1, and values of κ between 0 and min(100, $10 \times$ initial value of κ from 4).
6. Starting with initial values from 5, repeat 4 again up to 200 more times.

The three algorithms (Newton-Raphson, Powell's Dog Leg and Levenberg-Marquardt) are applied to each set of initial values.

To ensure that the desired maximum likelihood estimates are obtained several checks are also performed after a convergent result has been returned which have been listed below:

1. $\boldsymbol{H}_n(\hat{\boldsymbol{\phi}})$ is negative definite (all eigenvalues are negative).
2. $0 \leq \hat{p} \leq 1$.
3. if $\hat{\kappa} > 50$ then check to ensure that the log-likelihood function decreases when $\hat{\kappa}$ is increased by 20% and other parameters are held fixed at their returned values (this check was performed to ensure that we are not on one of the ridges discussed about in Section 4.2). Note that this check never failed. In the parametric bootstrap sample taken for Example 1 (described in more detail on the next page), there were 14 out of 10,000 samples that had $\hat{\kappa} > 50$ and all 14 of these cases passed this check. Some of these samples were investigated further and it was verified that there was indeed a significant proportion of data that was very tightly concentrated around the mean angular direction. Therefore the large values for the maximum likelihood estimates of the κ parameter appear to be reasonable for these 14 cases.

If any of these checks fail, then the MLE is not returned and one of the other methods is attempted.

To avoid computational errors several other checks are performed on the parameter estimates generated at each iteration. Let $\hat{\boldsymbol{\phi}}_i = (\hat{p}_i, \hat{\mu}_i, \hat{\kappa}_i)$ be the parameter estimates obtained for the i^{th} iteration. The checks performed on each iteration have been listed below:

1. $\hat{\kappa}_i > 0$
2. $I_0(\hat{\kappa}_i)$ and $I_1(\hat{\kappa}_i)$ are small enough so as not to result in overflow error (or on some software packages, like R for example, are small enough so as not to be set to infinity).
3. the log-likelihood function evaluated at $\hat{\boldsymbol{\phi}}_i$ is negative and does not involve taking the log of a negative number.
4. $\boldsymbol{H}_n(\hat{\boldsymbol{\phi}}_i)$ is not approximately singular (so that $\boldsymbol{H}_n(\hat{\boldsymbol{\phi}}_i)^{-1}$ can be calculated).

If any of these checks fail, then the algorithm in which the check failed is halted, the MLE is not returned and one of the other methods is attempted.

In Chapter 5 a likelihood ratio test is performed on the data from Example 1 in Chapter 2. The null hypothesis is that the data are von Mises distributed and the alternative hypothesis is that the data come from a mixture distribution. To obtain the approximate distribution of the likelihood ratio statistic under the null hypothesis, a parametric bootstrap sample of size 10,000 was taken. Samples were generated using $n = 44, \mu = 199.4°$ and $\kappa = 1.07$, the sample size and parameter estimates for a von Mises fit to Example 1.

Table 4.3 gives a summary of the number of samples out of 10,000 that did not converge to the mixture MLE after the specified set of initial values and algorithms were applied to bootstrap samples generated for Example 1. In the table, the abbreviations NR, PDL, and LM are used to specify the Newton Raphson, Powel's Dog Leg, and Levenberg-Marquardt algorithms, respectively. The initial values in the table are as described in the list on page 42.

Table 4.3: Number of bootstrap samples not converging to MLE for Example 1

Initial Value Set	NR	PDL	LM
1	687	174	164
2	153	144	139
3	137	137	137
4	30	29	29
5	16	8	8
6	1	1	1

Based on the results in the table, after the Newton-Raphson and Powel's Dog Leg algorithms have been applied to the original set of initial values, solutions for the mixture MLE have been found in about 98.3% of the samples. The Levenberg-Marquardt algorithm does not seem to have a significant impact in finding additional solutions. Also, it appears that solutions could be found in at least 99.3% of the samples by only using the Newton-Raphson and Powel's Dog Leg algorithms applied to the initial values sets, 1 and 4.

Chapter 5

Tests of Fit and Model Selection

In this chapter, we discuss how to assess whether or not the mixture model provides a statistically adequate fit to the data and provide a procedure for selecting the most appropriate model within the mixture distribution family (uniform, von Mises, or mixture of uniform and von Mises). An overview of procedures that can be used to select the appropriate model is provided in Section 5.1. In Section 5.2, goodness-of-fit tests based on Watson's U^2 statistic are discussed for the uniform, von Mises and mixture models. Parametric methods for testing for the uniform model against the von Mises model alternative and for testing for von Mises model against the mixture model alternative are given in Section 5.3. The model selection procedures discussed in Section 5.1 will be applied to the examples from Chapters 2 and 3 in Section 5.4.

5.1 Overview of Model Selection Procedures

There are two basic approaches to selecting the appropriate model within a family of distributions. The most commonly used approach starts with the simplest model and gradually adopts increasingly more complex models when there is statistically significant evidence that the simpler models are inadequate. Another approach starts with the most complex model within a family of distributions and gradually adopts the simper models provided there is not statistically significant evidence the more complex model should be kept. The approach of starting with the simplest model, which in our case is the uniform distribution, is more commonly used because it has the appeal of not requiring examination of the more complex models unless the simpler models are found to be inadequate. For this reason we

will present test procedures that start with the uniform model and adopt the more complex von Mises and mixture models when the uniform model model is found to be inadequate. Of course, if examination of the data very clearly reveals that there is a mode in which most of the data is clustered around, then one could consider starting with the von Mises model as the uniform model would be quite clearly inadequate.

There are also two types of tests, goodness-of-fit tests and parametric specific tests, that can be performed and each type lends itself to a different model selection procedure. Goodness-of-fit tests examine differences between the empirical cumulative distribution function and the cumulative distribution function in determining whether or not the fit is adequate. Parametric tests consider a specific alternative in assessing the fit. For example, a test for uniformity($\kappa = 0$) against the von Mises($\kappa > 0$) alternative would be a parametric test.

A goodness-of-fit test based model selection procedure uses only goodness-of-fit tests in determining whether or not more complex models need to be considered. The flowchart in Figure 5.1 illustrates the steps involved in performing a goodness-of-fit test based model selection procedure.

A parametric test based model selection procedure uses parametric tests with specific parametric alternatives at each step in determining which between the simpler null model and the alternative model is more appropriate. One advantage of using parametric tests is that if the specific alternatives are true, then the tests are more powerful than their goodness-of-fit test counterparts. A disadvantage with using parametric tests is that not all alternatives are examined, so if the true population is something quite different than specified by the alternative, then the test may not be very powerful in identifying the inadequacy of a simple model. The flowchart in Figure 5.2 initially uses parametric tests in finding the model that is most appropriate and uses goodness-of-fit tests to confirm that the model is still sufficient when there is no restriction placed on a specific alternative.

Figure 5.1: Flowchart for goodness-of-fit test based model selection procedure

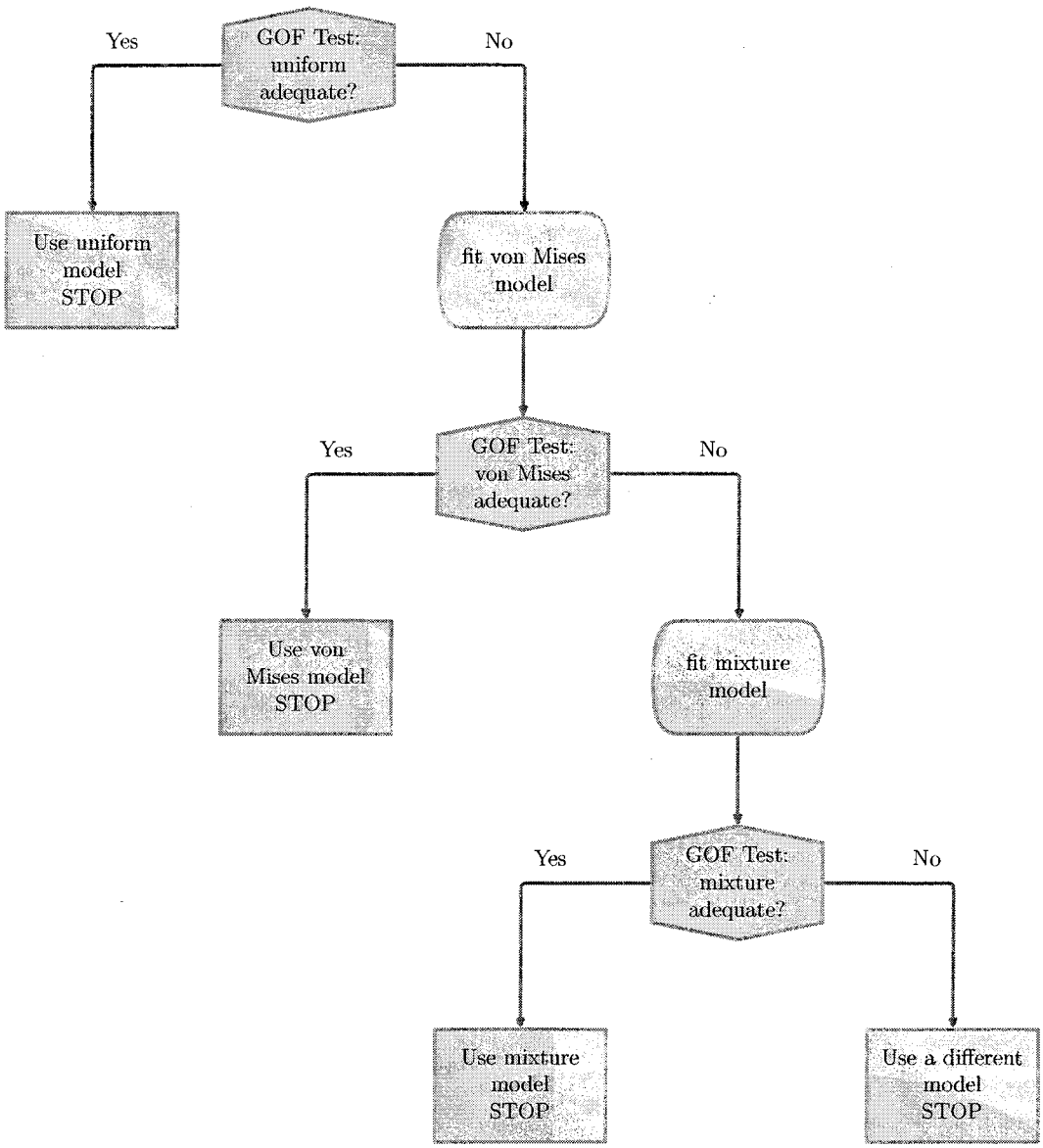
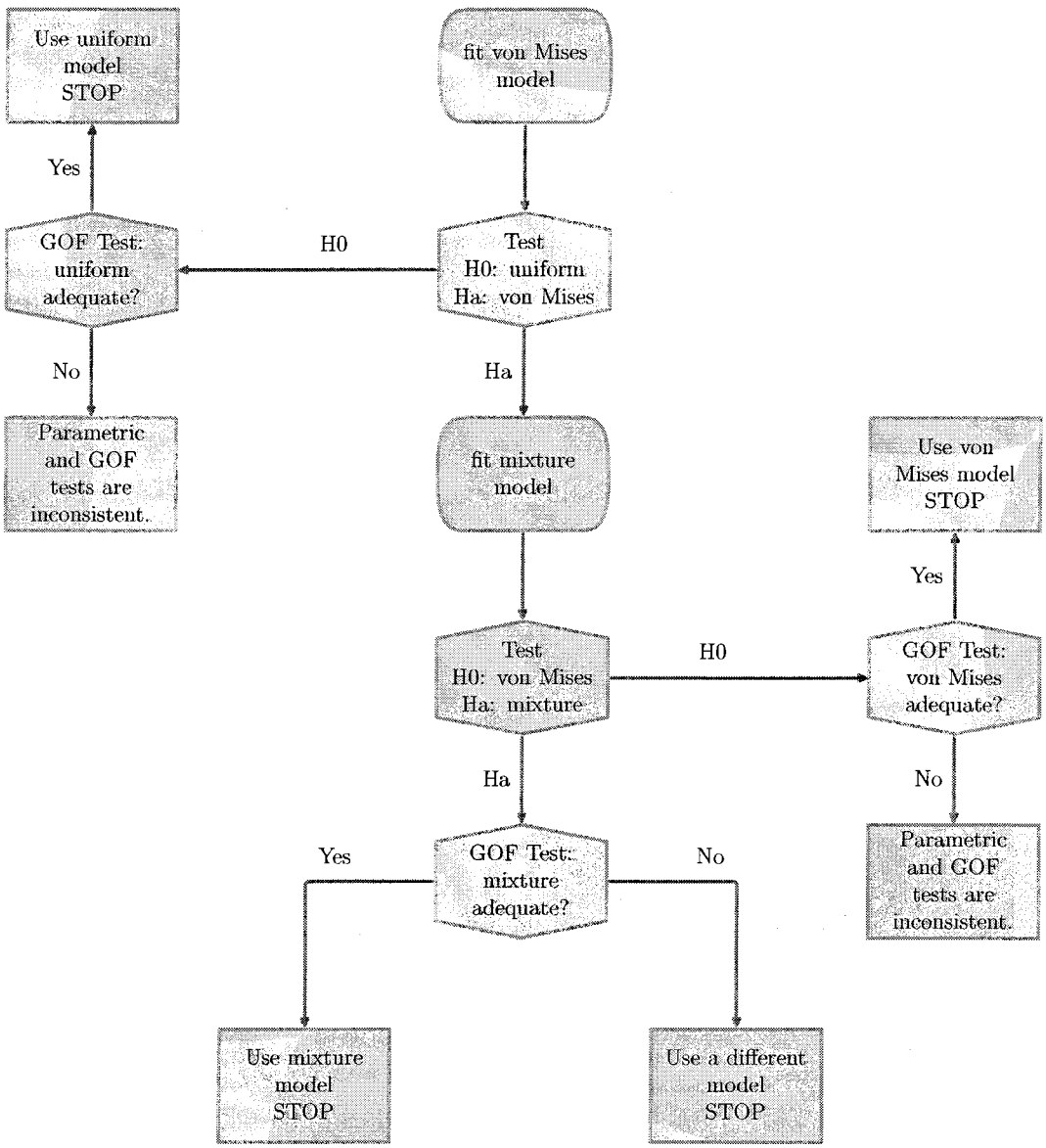


Figure 5.2: Flowchart for parametric test based model selection procedure



Some situations can arise in which a parametric test suggests the simpler model is sufficient but a goodness-of-fit test on the simpler model suggests that it is inadequate. Careful consideration needs to be given in these situations. If the goodness-of-fit test suggests only marginally significant evidence that the simpler model is inadequate, then one could consider perhaps still using the simpler model. If, however, there is strong statistical evidence that the simpler model is inadequate, the parametric test may have failed to identify the inadequacy as a result of the alternative model also being a fairly poor choice. One could consider then fitting the most complex model in the family, the mixture distribution. If a goodness-of-fit test also reveals that the mixture distribution is inadequate, then it is probably a good idea to reject the uniform, von Mises and mixture models and consider alternative models outside the scope of this paper.

5.2 Goodness-of-fit Tests

Section 5.2.1 provides on overview of a test procedure that can be applied using Watson's U^2 statistic on any distribution. Details for carrying out a goodness-of-fit test based on Watson's U^2 statistic for the uniform, von Mises, and mixture distributions are given in Sections 5.2.2, 5.2.3, and 5.2.4 respectively.

5.2.1 Overview

A commonly used test statistic in circular data problems is Watson's U^2 statistic, first presented and discussed by Watson [21]. A desirable property of Watson's U^2 statistic is that it is location invariant and thus does not depend on how the starting direction is assigned on the circle.

Let $\theta_1, \ldots, \theta_n$ be a random sample drawn from a population with a specified cumulative distribution function $F(\theta)$ and let $F_n(\theta)$ be the empirical distribution function. After changing the limits of integration appropriately for circular distributions and making some notational changes from Watson's paper, Watson's U^2 statistic can be expressed as

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

Suppose we wish to test the null hypothesis, H_0, that the random sample of θ-values comes from the distribution specified by $F(\theta)$. As outlined by Lockhart and Stephens in [12] but changing the notation slightly, the U^2 statistic is calculated as follows:

1. obtain maximum likelihood estimates for all unknown parameters;
2. for each i in $1, \ldots, n$, calculate $z_i = F(\theta_i)$, where unknown parameters are replaced by their maximum likelihood estimates if necessary;
3. put the z_i in ascending order to obtain $z_{(1)}, \ldots, z_{(n)}$;
4. calculate the U^2 statistic as

$$U^2 = \sum_{i=1}^{n} \left\{ z_{(i)} - (2i-1)/(2n) \right\}^2 - n(\bar{z} - 1/2)^2 + 1/(12n),$$

where $\bar{z} = \sum_{i=1}^{n} z_i/n$ is the sample average of the z values.

P-value calculation when all parameters are known:

If all parameters are known, then Watson's large sample result can be used to obtain a p-value,

$$\text{p-value} = P(U^2 > v) \approx \sum_{k=1}^{\infty} (-1)^{k-1} 2 \exp\left(-2k^2\pi^2 U^2\right).$$

A simple adjustment can be made to Watson's asymptotic result to make the p-value calculate more accurate for finite n as mentioned by Stephens [18]. The adjustment can be made by simply replacing U^2 with $U^* = (U^2 - 0.1/n + 0.1/n^2)(1 + 0.8/n)$ in the above p-value calculation.

P-value calculation when some parameters are unknown:

If some of the parameters are unknown, then obtaining p-values is somewhat more complicated. When we substitute maximum likelihood estimates in place of the true parameter values, values of U^2 typically decrease substantially. A parametric bootstrap sample can be taken to obtain the approximate distribution for U^2 when true parameter values are replaced by maximum likelihood parameter estimates. A method of obtaining p-values using a parametric bootstrap approach is outlined for a general distribution specified by $F(\theta)$ below.

1. obtain maximum likelihood estimates for any unknown parameters and calculate U^2 as described earlier.

2. Let N_{BS} be the specified number of bootstrap samples to take. For each $j \in 1, \ldots, N_{BS}$,

   (a) generate the j^{th} bootstrap sample, $\theta_{1j}, \ldots, \theta_{nj}$, from the distribution specified by $F(\theta)$, using maximum likelihood estimates from the original sample in place of any unknown parameters.

   (b) obtain maximum likelihood parameter estimates for any unknown parameters using the j^{th} bootstrap sample.

   (c) for each i in $1, \ldots, n$, calculate $z_{ij} = F(\theta_{ij})$, where unknown parameters are replaced by the maximum likelihood estimates obtained in (b).

   (d) put the z_{ij} in ascending order to obtain $z_{(1)j}, \ldots, z_{(n)j}$.

   (e) calculate the U^2 statistic for the j^{th} bootstrap sample as

$$U_j^2 = \sum_{i=1}^{n} \left\{ z_{(i)j} - (2i-1)/(2n) \right\}^2 - n(\bar{z}_j - 1/2)^2 + 1/(12n),$$

   where $\bar{z}_j = \sum_{i=1}^{n} z_{ij}/n$ is the sample average of the z_{ij} values for the j^{th} bootstrap sample.

Let N_l be the number of bootstrap samples that have a U^2 statistic that less than the U^2 statistic for the original sample. An approximate p-value for the test specified by H_0 is then given by

if $N_l = 0$ then

$$\text{p-value} \approx 1 - \frac{0.5}{N_{BS}},$$

else if $N_l < N_{BS}$ then

$$\text{p-value} \approx 1 - \frac{N_l}{N_{BS}},$$

else

$$\text{p-value} \approx \frac{0.5}{N_{BS}},$$

where N_{BS} is the total number of bootstrap samples.

5.2.2 Uniform Goodness-of-fit Test

The uniform goodness-of-fit test can be carried out as outlined in Section 5.2.1, replacing $F(\theta)$ with $F_u(\theta) = \theta/2\pi$. Since the uniform distribution has no parameters, there is no need for a parametric bootstrap.

5.2.3 von Mises Goodness-of-fit Test

The von Mises goodness-of-fit test can be carried out as outlined in Section 5.2.1, replacing $F(\theta)$ with

$$F_{VM}(\theta; \mu, \kappa) = \int_0^\theta f_{VM}(\phi; \mu, \kappa) d\phi,$$

where $f_{VM}(\theta; \mu, \kappa)$ is defined in Section 2.1. Typically μ and κ will not be known and need to be estimated. Thus, in the typical case, the parametric bootstrap mentioned in Section 5.2.1 will be required to obtain a p-value.

An alternative method that uses large sample theory to obtain the asymptotic distribution of Watson's U^2 statistic has been developed by Lockhart and Stephens. The theoretical details along with a table containing asymptotic critical points for various significance levels and all known/unknown parameter combinations are given by Lockhart and Stephens in [12].

For moderate to large sample sizes (ie $n \geq 50$) both the parametric bootstrap approach and the asymptotic distribution approach used by Lockhart & Stephens produce similar results and either of these methods should be sufficiently accurate for all practical purposes.

5.2.4 Mixture Goodness-of-fit Test

The mixture goodness-of-fit test can be carried out as outlined in Section 5.2.1, replacing $F(\theta)$ with

$$F(\theta; p, \mu, \kappa) = \int_0^\theta f(\phi; p, \mu, \kappa) d\phi,$$

where $f(\theta; p, \mu, \kappa)$ is defined in Section 3.1. Typically p, μ, and κ will not be known and need to be estimated. Thus, in the typical case, the parametric bootstrap mentioned in Section 5.2.1 will be required to obtain a p-value.

5.3 Parametric Tests

In this section two parametric tests are provided. Tests of uniformity again the von Mises alternative are discussed in Section 5.3.1 and a test of the von Mises family against the mixture alternative is given in Section 5.3.2.

5.3.1 Tests of Uniformity Against the von Mises Alternative

Using the Neyman-Pearson lemma, the likelihood ratio test (also referred to as Rayleigh's test as will be explained later) can easily be shown to be the uniformly most powerful invariant test of uniformity against the von Mises alternative as mentioned in [1](p. 348), for example. Consequently this test has been widely discussed by many authors. A discussion of the likelihood ratio test of uniformity against the von Mises and other alternatives is given by Stephens in [20] (pp. 347-349) and a more detailed discussion of tests of uniformity against unimodal and bimodal von Mises alternatives is given by Stephens [19].

We summarize some of the details for the likelihood ratio test. Let $\theta_1, \ldots, \theta_n$ be a random sample of circular data points. Some details required in explaining the likelihood ratio test below have already been provided in Chapter 2. The von Mises density, along with a description of its parameters μ and κ is given in Section 2.1. The von Mises log-likelihood function, $l_{VM}(\mu, \kappa)$, and maximum likelihood estimates $\hat{\mu}$ and $\hat{\kappa}$ are provided in Section 2.2.

A likelihood ratio test procedure can be performed to test the null hypothesis

$$H_0 : \text{the points are uniformly distributed around the circle}(\kappa = 0),$$

against the alternative hypothesis,

$$H_A : \text{the points are von Mises distributed around the circle}(\kappa > 0).$$

This leads to the likelihood ratio test statistic

$$LR = 2\left\{l_{VM}(\hat{\mu}, \hat{\kappa}) - l_U\right\} = 2\left[\hat{\kappa}R - n\log\left\{I_0(\hat{\kappa})\right\}\right],$$

where $l_U = -n\log(2\pi)$ is the uniform log-likelihood, the resultant length, R, and the zeroth order modified bessel function of the first kind, $I_0(\kappa)$, are defined in Section 2.2.

On first examination it may appear as though we are performing a one tailed rather than a two-tailed test since the null hypothesis, $\kappa = 0$ appears to be on the boundary of the parameter space rather than an interior point. However, we restrict values of κ to being nonnegative merely for convenience since the parameters $\kappa^* = -\kappa$ and $\mu^* = \mu + \pi$ specify the same von Mises distribution as the parameters κ and μ. Consider re-parameterizing the von Mises distribution with parameters $\phi_1 = \kappa\cos(\mu)$ and $\phi_2 = \kappa\sin(\mu)$. The null hypothesis for testing for uniformity against the von Mises alternative then becomes

$$H_0 : \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \mathbf{0}.$$

which can be seen to be an interior point in the parameter space. Therefore standard theory for likelihood ratio tests can be used to show that the likelihood ratio test statistic has an asymptotic χ^2_2 distribution or equivalently has an asymptotic exponential distribution with mean parameter 2. An approximate p-value can be obtained using

$$\text{p-value} \approx \exp(-LR/2).$$

Rayleigh developed a commonly used test of H_0 against H_A using the test statistic R/n. The likelihood ratio statistic can be expressed as a monotone function of R/n and therefore Rayleigh's test statistic produces a test that is equivalent to the likelihood ratio test. More details for Rayleigh's test can be found in Batschelet [2].

Rayleigh [17] showed that the statistic $2R^2/n$ is asymptotically χ^2_2 distributed and in large samples,

$$\text{p-value} = \text{Prob}(R \geq r) \approx \exp(-r^2/n).$$

Both Rayleigh's and the likelihood ratio statistic approximate p-value calculations are reasonably accurate for large sample sizes ($n \geq 50$). In the extreme tail of the distribution though neither of these approximations will be very accurate.

Many approximations have been proposed by various authors for smaller sample sizes. Stephens [19] uses the first four moments of R^2 to fit Pearson curves which yield very accurate approximations for the distribution of R/n. Critical values for the R/n statistic for various sample sizes and significance levels that were obtained by this method are provided in [19] (Table 2, p. 285). A table of the approximate sample sizes required to correctly identify when the data come for a von Mises rather than circular uniform distribution for various desired detection probabilities and values of κ is also provided in [19] (Table 4, p. 288). Stephens' approximation is sufficiently accurate even for small sample sizes for all practical purposes.

Exact Distribution of R

Pearson initially posed the problem of finding the distribution of R as a random walk problem in [16]. Shortly after, Kluyver [11] solved the problem and obtained an expression for the exact cumulative distribution of R,

$$\text{Prob}(R < r) = r \int_0^\infty \{J_0(x)\}^n J_1(rx) dx, \tag{5.1}$$

where $J_0(x)$ and $J_1(x)$ are the usual Bessel functions,

$$J_k(x) = \frac{1}{2\pi} \int_0^{2\pi} \cos\{k\theta - x\sin(\theta)\}\, d\theta.$$

An exact p-value is therefore given by

$$\text{p-value} = \text{Prob}(R \geq r) = 1 - \text{Prob}(R < r) = 1 - r \int_0^\infty \{J_0(x)\}^n J_1(rx) dx.$$

Durand and Greenwood [5] use quadrature to obtain numerical results for equation 5.1, accurate to 5 decimal places, and provide a table for samples of size 6 through 24 and various

values of r on page 235. With improved computer technology and using adaptive quadrature with double precision for all numerical operations, it is possible to obtain p-values accurate to about 10 decimal places for sample sizes less than or equal to 100. Due to propagation of round off errors though, it is still not feasible to obtain accurate p-values in the extreme tail of the distribution, particularly when the sample size is large. For all practical purposes, though, accuracy of p-values beyond the 3^{rd} decimal point is of no real importance.

Comparison of the χ^2 approximations

For $n \geq 50$, p-values can be approximately obtained using the either the χ^2 approximations for the likelihood ratio statistic or the $2R^2/n$ statistic. A brief investigation was done in an attempt to determine which of these approximations is better over different regions of the distributions of R.

Figure 5.3 compares the exact p-values with the p-values obtained from the 2 different approximations for $n = 50$ and over the most critical part of the distribution (p-values between 0.01 and 0.05) where the cutoff point is typically made in deciding whether or not to reject the uniform null hypothesis. As can be seen in the figure, the exact p-value is always between the two approximations. The approximation based on the $2R^2/n$ statistic is overly conservative and does not reject often enough while the reverse is true of the likelihood ratio statistic. The approximation based on the $2R^2/n$ statistic appears slightly more accurate for most of the critical region and is noticeably more accurate for p-values around 0.05. For large samples ($n \geq 50$) though, both of these approximations appear to be sufficiently accurate.

Many samples exist that have very strong evidence of belonging to a von Mises rather than uniform population. In these samples the p-values can often be very small. As mentioned earlier, exact p-values are difficult to obtain when they are in the extreme tail due to propagation of round of errors. Accurate p-values in the extreme tail are not particularly important though because p-values in the extreme tail are an indication that either a highly unlikely sample was obtained or that the null hypothesis is false. Ultimately the accuracy of the p-value in the extreme tail really has no bearing on whether or not the null hypothesis of uniformity is rejected since p-values less than 0.01 result in rejecting the null hypothesis by almost all standards and typically the null hypothesis is rejected when the p-value is less than 0.05.

Even though the accuracy of p-values in the extreme tail is not particularly important, it may still be somewhat useful to have at least some idea as to their accuracy in the extreme tail since p-values are frequently reported. Wilson [22] comments on not being able to find an accurate approximation for the probability of a large resultant in a random walk in the literature. In an attempt to obtain a rough idea of the accuracy of the p-values obtained using the two χ^2 different approximations, p-values obtained using the two different approximations are compared with exact p-values of 10^{-6} for samples of size $n = 50 \ldots 200$ in Figure 5.4. Once again, the exact p-value appears to be bounded by the two approximations. As can be seen in Figure 5.4, the approximation based on the likelihood ratio statistic is much closer to the exact p-value and hence appears to be the more accurate approximation in the extreme tail of the distribution.

Figure 5.3: Comparison of χ^2 approximations in the critical region of the distribution ($n = 50$)

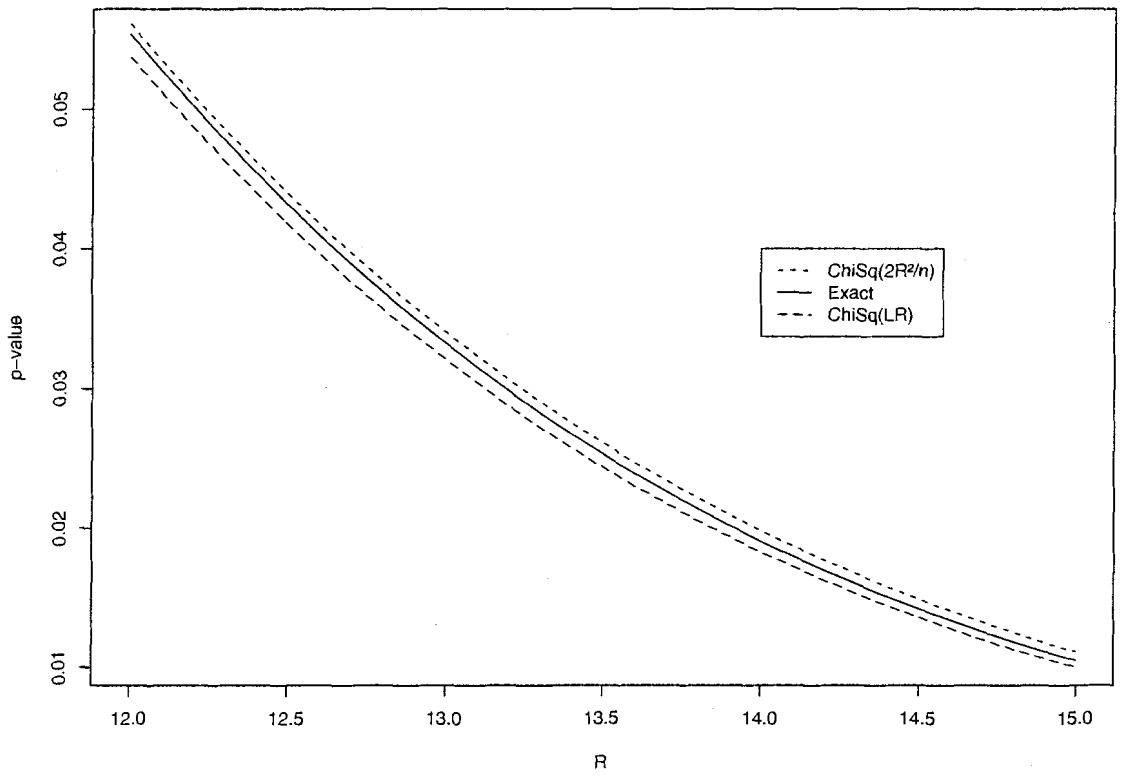
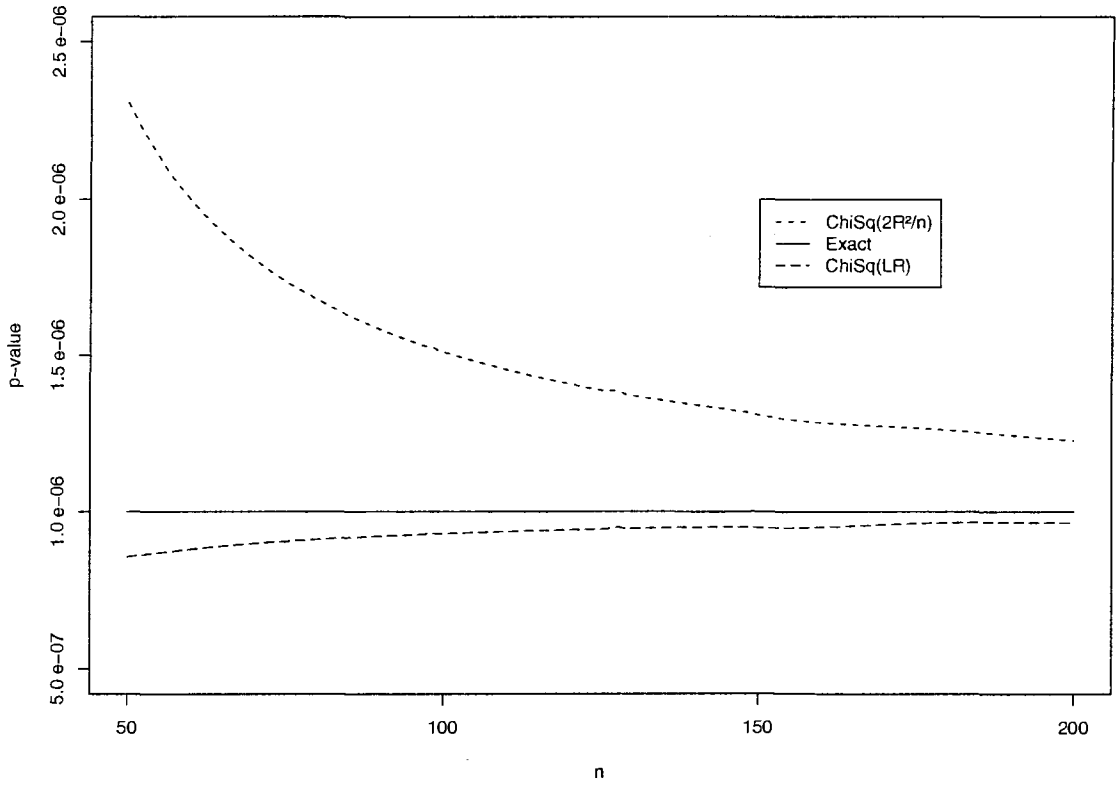


Figure 5.4: Comparison of χ^2 approximations in the extreme tail of the distribution



5.3.2 Test of the von Mises Family Against the Mixture Alternative

Let $\theta_1, \ldots, \theta_n$ be a random sample of circular data points. Some details required in explaining the likelihood ratio test below have already been provided in chapters 2 and 3. The von Mises density, along with a description of its parameters μ and κ, is given in Section 2.1. The von Mises log-likelihood function, $l_{VM}(\mu, \kappa)$, and maximum likelihood estimates $\hat{\mu}$ and $\hat{\kappa}$ are provided in Section 2.2. To avoid confusion with the mixture maximum likelihood estimates, the von Mises maximum likelihood estimates for μ and κ will be referred to as $\hat{\mu}_{VM}$ and $\hat{\kappa}_{VM}$ throughout the rest of this section. The mixture density, along with a description

of its parameters p, μ and κ, is given in Section 3.1. The mixture log-likelihood function, $l(p, \mu, \kappa)$, and maximum likelihood estimates, $\hat{p}$, $\hat{\mu}$ and $\hat{\kappa}$ are provided in Section 3.2.

A likelihood ratio test procedure can be performed to test the null hypothesis,

$$H_0 : \text{the points are von Mises distributed around the circle}(p = 1),$$

against the alternative hypothesis,

$$H_A : \text{the points are a mixture of uniform and von Mises distrubted data}(p < 1).$$

By the Neyman-Pearson lemma, the likelihood ratio test is the uniformly most powerful invariant test of H_0 against H_A. The likelihood ratio test statistic is given by.

$$LR = 2\left\{l(\hat{p}, \hat{\mu}, \hat{\kappa}) - l_{VM}(\mu_{\hat{VM}}, \kappa_{\hat{VM}})\right\}.$$

It is important to mention the null hypothesis is on the boundary of the parameter space(ie we restrict values of p to be no greater than 1). If there were no restriction placed on p, then the asymptotic distribution of LR under the null hypothesis would be χ^2_1. Recall from Chapter 4 that we set the mixture MLE equal to the von Mises MLE when the partial derivative of the mixture log-likelihood with respect to the parameter p evaluated at the von Mises MLE is greater than 0. In these situations, the likelihood ratio statistic will be equal to 0 and in large samples, under the null hypothesis, this will occur with an approximate probability of 1/2. Thus, in large samples, the p-value for the likelihood ratio test can be calculated using

$$\text{p-value} \approx \frac{1 - F_{\chi^2_1}(LR)}{2},$$

where $F_{\chi^2_1}(x)$ is the cumulative distribution function of the χ^2_1 distribution.

For samples of size $n \geq 50$ using the χ^2_1 approximate distribution for the likelihood ratio statistic yields sufficiently accurate results for all practical purposes. For smaller samples, it is a good idea to take a bootstrap sample from a von Mises population with parameters set to the von Mises MLE to obtain the approximate distribution of the likelihood ratio statistic under the null hypothesis. An approximate p-value based on the likelihood ratio statistic from the original sample can then be obtained similarly to what was described in Section 5.2.1 for p-value calculations when parameters are unknown but replacing Watson's U^2 statistic with the likelihood ratio statistic.

5.4 Examples

5.4.1 Tests of Fit for Example 1

The two different model selection procedures discussed in Section 5.1 were applied to the sandstone rock data from example 1 in Section 2.5. All tests were performed at a significance level of 0.05. Tables 5.1 and 5.2 provide p-values for the tests used by the two different procedures.

Table 5.1: P-values of goodness-of-fit tests for Example 1

Model	U^2 statistic	p-value
uniform	0.546	3.6×10^{-5}
von Mises	0.047	0.30

Bootstrap samples of size 10,000 were taken for performing the von Mises goodness-of-fit test.

Table 5.2: P-values of parametric tests for Example 1

Null Model	R	LR statistic	p-value
uniform	20.7	20.7	3.7×10^{-5}
von Mises		0.727	0.23

The p-value for the uniform null model against the von Mises alternative is exact.

The p-value for the von Mises null model against the mixture alternative is obtained using a parametric bootstrap sample of size 10,000 to approximate the distribution of the likelihood ratio statistic. The likelihood ratio statistic could not be obtained for one of the bootstrap samples but this has no impact on the p-value reported to 2 decimal places. The p-value obtained using the the large sample χ^2 approximation is 0.20 and is reasonably consistent to the p-value obtained from taking a parametric bootstrap. Since the sample size is somewhat small, $n = 44$, the p-value obtained using the parametric bootstrap is thought to be more accurate.

Both the goodness-of-fit test based and parametric test based model selection procedures find the von Mises model to be the most appropriate model when the tests are performed at the 0.05 significance level. In both cases, the uniform model is very clearly rejected.

5.4.2 Tests of Fit for Example 2

The two different model selection procedures discussed in Section 5.1 were applied to the ant data from example 2 that was fit to the von Mises model in Section 2.6 and was fit to the mixture model in Section 3.4. Tables 5.3 and 5.4 provide p-values for the tests used by the two different procedures.

Table 5.3: P-values of goodness-of-fit tests for Example 2

Model	U^2 statistic	p-value
uniform	2.243	8.4×10^{-20}
von Mises	0.288	5.0×10^{-5}
mixture	0.019	0.87

Bootstrap samples of size 10,000 each were taken for performing the von Mises and mixture goodness-of-fit tests.

For the von Mises goodness-of-fit test, all 10,000 bootstrap samples had a smaller value for Watson's U^2 statistic than was obtained from the original sample. Therefore the actual p-value may well be less than 5.0×10^{-5}.

Table 5.4: P-values of parametric tests for Example 2

Null Model	R	LR statistic	p-value
uniform	60.9	83.0	9.4×10^{-19}
von Mises		25.5	2.2×10^{-7}

The p-values for uniform null model against the von Mises alternative and the von Mises null model against the mixture alternative are obtained using the χ^2 approximation of the likelihood ratio statistic.

Both the goodness-of-fit test based and parametric based model selection procedures find the mixture model to be the most appropriate model when the tests are performed at

the 0.05 significance level. In both cases, the uniform and von Mises models are very clearly rejected.

Chapter 6

Future Work

Three areas for future work are discussed in this chapter. A discussion of a future Monte Carlo study to verify the power and accuracy of the tests is in Section 6.1. Extensions of the mixture model to grouped data and to spherical data are discussed in sections 6.2 and 6.3, respectively.

6.1 Monte Carlo study

The accuracy and power of the tests for the mixture distribution given in Chapter 5 have not been studied yet. Specifically the U^2 goodness-of-fit test for the mixture distribution discussed in section 5.2.4 and the likelihood ratio test for testing for von Misesness against the mixture alternative discussed in 5.3.2 need to be studied further. A future Monte Carlo study is planned where we will:

1. verify the accuracy of the tests for various parameter values for the mixture distribution and sample sizes
2. estimate and compare the power of the goodness-of-fit and parametric tests for various distributions

6.2 Extending mixture distribution to grouped data

The mixture distribution and associated tests used in this project were all based on continuous data. In some situations, data can be grouped and it is only known what interval

each of the points is in as in example 3 in section 4.3.1. A proper analysis a grouped data requires that the likelihood function be based on the probabilities of each of the points being in their specific intervals. The maximum likelihood estimates would then be obtained using the grouped rather than continuous likelihood function. The tests of fit used for continuous data in chapter 5 also could be updated to facilitate the analysis of grouped data problems. The goodness-of-fit test for discrete uniformity based on Watson's U^2 statistic has already been developed by by Choulakian, Lockhart and Stephens in [3].

6.3 Extending mixture distribution to spherical data

The von Mises distribution is the special circular(1-dimensional) case of the more general N-dimensional von Mises-Fisher distribution. The same concept of the mixture distribution that was introduced for circular data in this project can be extended to the sphere. For spherical data, a model could be considered for analysing spherical(2-dimensional) von Mises-Fisher distributed data that is mixed with points uniformly distributed around the sphere.

Appendix A

Newton-Raphson Algorithm

The Newton-Raphson algorithms requires an initial value, x_0, derivative function, U, and hessian matrix, H, as inputs and attempts to find a solution, $\hat{x}$, to $U(x) = 0$. Let M be the maximum number of iterations and *tol*, the tolerance, be the maximum allowable absolute value of the components of $U(x)$. A basic Newton-Raphson algorithm is provided below.

Newton-Raphson(x_0,U,H)

Step 1: Set $i = 0$
Step 2: While $(i < M)$ and $(\max\{\text{abs}(U(x_i))\} > tol)$ repeat Steps 3 and 4
Step 3: Set $i = i + 1$
Step 4: Set $x_i = x_{i-1} - H(x_{i-1})^{-1}U(x_{i-1})$
Step 5: If $(\max\{\text{abs}(U(x_i))\} \leq tol)$, then set $\hat{x} = x_i$
Step 6: Else algorithm has exceeded maximum number of iterations.

Values of $M = 50$, and $tol = 10^{-8}$ were used as defaults for the Newton-Raphson algorithm.

Bibliography

[1] B. Ajne (1968). A simple test for uniformity of a circular distribution. *The Canadian Journal of Statistics*, 55:343–354.

[2] Batschelet, E. (1981). *Circular Statistics in Biology.* Academic Press Inc.: London.

[3] Choulakian, V., Lockhart, R. A. & Stephens, M. A. (1994). Cramér von mises statistics for discrete distributions. *The Canadian Journal of Statistics*, 22:125–137.

[4] Cox, D. R. & Hinkley, D. V. (1974). *Theoretical Statistics.* Chapman & Hall: London.

[5] Durant, D., Greenwood, J. A. (1955). The distribution of lenght and components of the sum of n random unit vectors. *Ann. Math. Statist.*, 26:233–246.

[6] Edwards, J. H. (1961). The recognition and estimation of cyclic trends. *Ann. Hum. Gen.*, 25:83–86.

[7] Fisher, N. I. (1993). *Statistical Analysis of Circular Data.* Cambridge University Press: Cambridge.

[8] Jammalamadaka, S. R. & SenGupta, A. (2001). *Topics in Circular Statistics.* World Scientific: New Jersey.

[9] Jander, R. (1957). Die opstische Richtangsorientierung der roten Waldameise. *(Formica rufa. L.) Z. vergl. Physiologie*, 40:162–238.

[10] Kiersch, G. A. (1950). Small scale structures and other features of Navajo sandstone, northern part of Rafael Swell, Utah. *Am. Assoc. Pet. Geol.*, 34:923–942.

[11] Kluyver, J. C. (1906). A local probability problem. *Nederl. Akad. Wetnsch, Proc.*, 8:341–350.

[12] Lockhart, R. A. & Stephens, M. A. (1985). Tests of fit for the von Mises distribution. *Biometrika*, 72:647–652.

[13] Mardia, K. V. & Jupp, P. E. (2000). *Directional Statistics.* John Wiley & Sons: Chichester.

[14] Madsen, K., Nielson, H. B. & Tingleff, O. (2004). *Methods for Non-Linear Least Squares Problems.* Technical University of Denmark, `http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/3215/pdf/imm3215.pdf`.

[15] Pearson, E. S. & Stephens, M. A. (1972). *Biometrika Tables for Statisticians Sections VIII and IX*, volume 2. Cambridge University Press: Cambridge.

[16] Pearson, K. (1905). The problem of the random walk. *Nature*, 72:294–302.

[17] Rayleigh, Lord, (Strutt, J. W.) (1919). On the problem of random vibrations and random flights in one, two and three dimensions. *Philosophical Magazine*, 37:321–347.

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

[19] Stephens, M. A. (1969). Tests for randomness of directions against two circular alternatives. *J. Amer. Statist. Assoc.*, 64:280–289.

[20] Stephens, M. A. (1986). Chapter 8: Tests for the uniform distribution, in *Goodness-of-fit Techniques,* Eds R. A. D'Agostino & M. A. Stephens, pp. 331-366. Marcel Dekker, Inc.: New York.

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

[22] Wilson, A. J. C. (1980). How probable is a large resultant in a random walk? *Technometrics*, 22:629–630.