ON FITTING A MIXTURE OF TWO VON MISES DISTRIBUTIONS, WITH APPLICATIONS

by

Qifeng Jiang
B.Sc., University of British Columbia, 2006

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
Spring 2009

APPROVAL

Name: Qifeng Jiang

Degree: Master of Science

Title of Project: On Fitting A Mixture of Two von Mises Distributions, with Applications

Examining Committee: Dr. Derek Bingham
Chair

____

Dr. Michael Stephens,
Simon Fraser University
Senior Supervisor

____

Dr. Richard Lockhart,
Simon Fraser University
Supervisor

____

Dr. Boxin Tang,
Simon Fraser University
External Examiner

Date Approved: ______________________________

Abstract

Circular data refers to data recorded as points on a circle, either denoting directions, or times when the circle acts as a clock. The von Mises distribution is frequently used to analyze circular data sets with a clear peak. When two clear peaks appear on the circle, a mixture of two von Mises distributions is often used to analyze the data. Parameter estimates are produced by using maximum likelihood estimation, and Watson's U^2 is used to test the fit. Two data sets will be discussed in this project: times of Sudden Infant Death Syndrome (SIDS) occurrences and times of Fatal Crash accidents.

Keywords: maximum likelihood estimation, Watson's U^2.

Acknowledgments

When composing this thesis, I have obtained assistance and support from many people, without whom this thesis may not have appeared. First of all, I am most indebted to my supervisor, Professor Michael Stephens, for his expert guidance. He not only guided me to do scientific research, but also made invaluable and inspiring comments which strengthened my interest in statistics. I am grateful for his patient teaching and encouragement, and finally for his careful revising of the drafts of this thesis.

I would also like to give my special thanks to Professor Richard Lockhart, who has provided me with many valuable ideas. Every time when I was stuck with the scientific work, Professor Lockhart opened another window for me.

Finally, my thanks should go to my parents, who have financially supported my 6-year study in Canada and always warmed my heart with care and encouragement when I was in hard time. Their concern and guidance will forever come along with me.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Circular data arise in biology, geography, medicine, astronomy, and many other areas. Each observation of circular data can be shown as a point P on a unit circle with centre O, and its direction OP is measured using its angle in degrees or radians. There are many examples of circular data, for instance, the angles taken by birds released away from home, or the direction of the wind. The circle is also widely used as a 24-hour clock or one year calendar. For example, the times of cars going through a particular crossing can be recorded as circular observations within a 24-hour period. Incidents of disease occurring within one year is also an example of circular data.

The von Mises distribution is used to analyze circular data where there appears to be a peak in the data. However, in some cases, one single von Mises distribution cannot fit the data well. In Figure 1.0.1, the plot shows that there are two modes on the circle. One is around $\frac{\pi}{6}$ radians, and the other is around $\frac{5\pi}{6}$. For this kind of situation, a mixture of two von Mises distributions is a better choice.

Maximum likelihood estimation (MLE) is a standard method to estimate parameters. When the sample size is large, the method of maximum likelihood gives good estimators with minimum variance.

The method of moments is also used to estimate parameters of a distribution. One advantage of the method of moments is that estimators can be found easily and quickly for most distributions. However, for some cases, estimators calculated using the method of moments may not lie inside the parameter space. In this paper, we will estimate parameters of a mixture of two von Mises distributions by MLE, and then compare these estimators with the method of moment estimators.

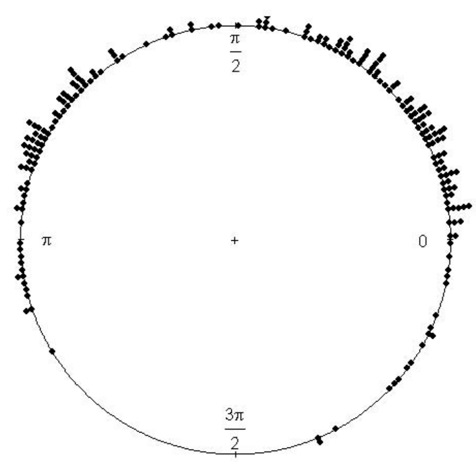

Figure 1.0.1: Data from a mixture of two von Mises distributions

Chapter 2

The von Mises Distribution

In this chapter, we will focus on modeling circular data using a single von Mises distribution. An introduction of the von Mises distribution and its probability density function will be given in Section 2.1. In Section 2.2, we will discuss how to use the method of maximum likelihood to find estimators of the von Mises distribution, and the estimates of the method of moments will be discussed in Section 2.3. An example and a comparison of these two main methods will be given in Section 2.6.

2.1 The von Mises Distribution

Suppose that we have a unit circle with origin O, and $P_i,\ i = 1, \cdots, n$ are points on the circle. The unit vector OP_i gives a direction, with angular co-ordinate θ_i.

The probability density function of the von Mises distribution is:

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

where $I_0(\kappa)$ is the modified Bessel function of order zero and the first kind, and given by:

$$I_0(\kappa) = \frac{1}{2\pi} \int_{\theta=0}^{2\pi} \exp\{\kappa \cos(\theta)\} d\theta.$$

The Bessel function $I_0(\kappa)$ can be also expressed as:

$$I_0(\kappa) = \sum_{r=0}^{\infty} \frac{1}{r!^2} (\frac{\kappa}{2})^{2r}.$$

The distribution is clustered and symmetric around μ, which is the measure of mean direction. Parameter κ is a measure of concentration, and $\frac{1}{\kappa}$ is analogous to the variance. When $\kappa = 0$, the von Mises distribution becomes the uniform distribution. It goes to the point distribution concentrated in the direction μ when κ goes to positive infinity.

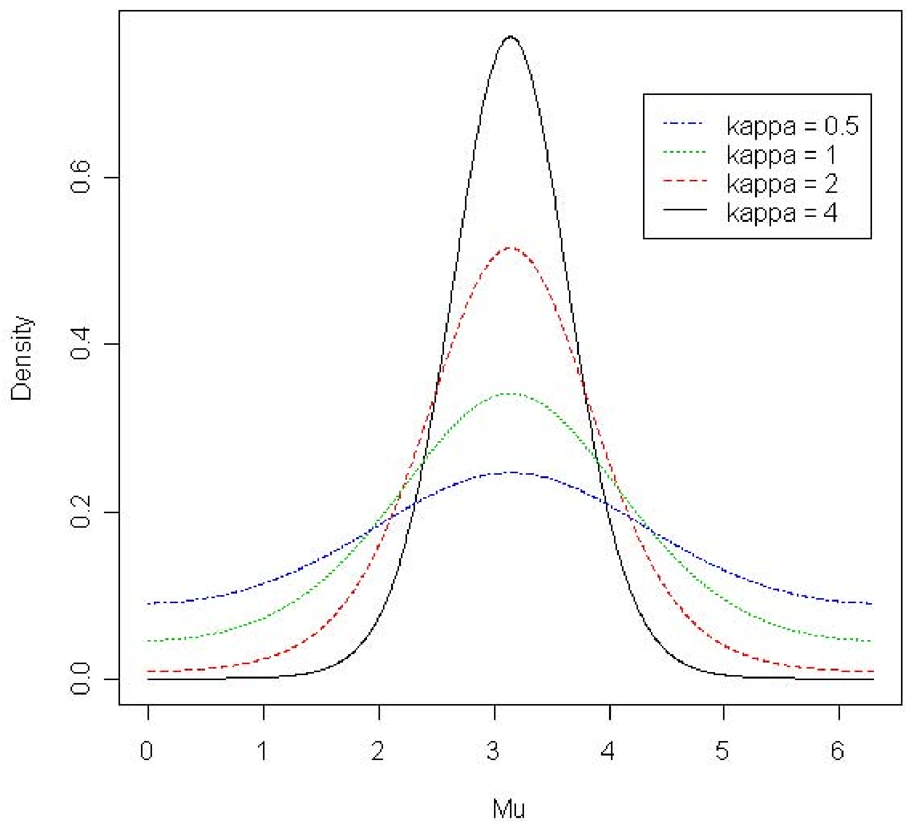

Figure 2.1.1: Probability density functions of von Mises distributions

Figure 2.1.1 shows how the density function changes when the concentration parameter κ is changing.

2.2 Maximum Likelihood Estimation

Let $\theta_1, \cdots, \theta_n$ be independent identically distributed random variables following the von Mises distribution with mean direction μ and concentration parameter κ. The likelihood function will be:

$$L(\mu, \kappa) = \prod_{i=1}^{n} f(\theta_i, \mu, \kappa). \tag{2.2.1}$$

The log-likelihood function becomes:

$$\log L = -n \log I_0(\kappa) - n \log 2\pi + \kappa \sum_{i=1}^{n} \cos(\theta_i - \mu). \tag{2.2.2}$$

Taking the first derivative with respect to μ and κ , we obtain:

$$\frac{\partial \log L}{\partial \mu} = \kappa \sum_{i=1}^{n} \sin(\theta_i - \mu), \tag{2.2.3}$$

and

$$\frac{\partial \log L}{\partial \kappa} = -nA(\kappa) + \sum_{i=1}^{n} \cos(\theta_i - \mu), \tag{2.2.4}$$

where $A(\kappa)$ is the ratio of $I_1(\kappa)$ to $I_0(\kappa)$, and $I_1(\kappa)$ is the modified Bessel function of the first kind and order one, and $I_1(\kappa)$ is the first derivative of $I_0(\kappa)$ with respect to κ.

These must be set equal to zero to obtain ML estimates. Term $\kappa \sum_{i=1}^{n} \sin(\theta_i - \hat{\mu})$ is equal to zero only when $\hat{\mu}$ equals to θ_R, where θ_R is the direction of the resultant; the length of resultant R and θ_R are the solutions of:

$$\textstyle\sum_{i=1}^{n} \cos \theta_i = R \cos \theta_R, \text{ and } \sum_{i=1}^{n} \sin \theta_i = R \sin \theta_R.$$

Therefore, the maximum likelihood estimator $\hat{\mu}$ of μ is θ_R.

Hogg and Craig (1965, pp.229-230) show that R and θ_R are jointly complete sufficient statistics for μ and κ. Furthermore, if κ is given, then the minimal sufficient statistics for μ are $R \cos \theta_R$ and $R \sin \theta_R$.

In equation 2.2.4, $\sum_{i=1}^{n} \cos(\theta_i - \mu) = R \cos(\theta_R - \mu)$. Therefore, if the right hand side of equation 2.2.4 is set to be zero, we obtain:

$$-nA(\hat{\kappa}) + R \cos(\theta_R - \hat{\mu}) = 0. \tag{2.2.5}$$

This implies:

$$A(\hat{\kappa}) = \bar{R}, \tag{2.2.6}$$

where $\bar{R} = R/n$.

Therefore, the maximum likelihood estimator $\hat{\kappa}$ of κ is equal to $A^{-1}(\bar{R})$. However, the solution of equation 2.2.6 can only be solved numerically. Fisher (1993, p.88) has given a numerical solution for $\hat{\kappa}$ as following:

$$\begin{aligned}
\hat{\kappa} &= 2\bar{R} + \bar{R}^2 + \tfrac{5}{6}\bar{R}^5, && \text{if} \quad R < 0.53;\\
\hat{\kappa} &= -0.4 + 1.39\bar{R} + \tfrac{0.43}{1-\bar{R}}, && \text{if} \quad 053 \leq R < 0.85;\\
\hat{\kappa} &= \tfrac{1}{\bar{R}^3 - 4\bar{R}^2 + 3\bar{R}}, && \text{if} \quad R \geq 0.85.
\end{aligned}$$

For sample size n smaller than or equal to 15, based on the results above, Fisher gave another numerical estimator of κ, calling it $\hat{\kappa}^*$.

$$\begin{aligned}
\hat{\kappa}^* &= \max(\hat{\kappa} - 2(n\hat{\kappa})^{-1}, 0) && \text{if } \hat{\kappa} < 2,\\
\hat{\kappa}^* &= (n-1)^3\hat{\kappa}/(n^3 + n) && \text{if } \hat{\kappa} \geq 2.
\end{aligned}$$

2.3 Method of Moments

The method of moments is as follows. If θ is a random variable following the von Mises distribution with mean direction μ and concentration parameter κ, expectations of θ and θ^2 cannot be solved analytically. Instead, expectations of $\cos\theta$ and $\sin\theta$ are used. Formulas for $E(\cos\theta)$ and $E(\sin\theta)$ are:

$$E(\cos\theta) = A(\kappa)\cos\mu, \text{ and } E(\sin\theta) = A(\kappa)\sin\mu.$$

Estimators of $E(\cos\theta)$ and $E(\sin\theta)$ are $\frac{1}{n}\sum_{i=1}^{n}\cos\theta_i$ and $\frac{1}{n}\sum_{i=1}^{n}\sin\theta_i$. Let $\bar{C} = \frac{1}{n}\sum_{i=1}^{n}\cos\theta_i$, and $\bar{S} = \frac{1}{n}\sum_{i=1}^{n}\sin\theta_i$, by the method of moments, the estimators of μ and κ are the solutions of:

$$\bar{C} = A(\kappa)\cos\mu, \text{ and } \bar{S} = A(\kappa)\sin\mu.$$

These equations are equivalent to the equations given by the method of maximum likelihood.

2.4 Asymptotic Properties

Cox and Hinkley (1974) have shown that the maximum likelihood estimators $\hat{\kappa}$ and $\hat{\mu}$ of κ and μ have the asymptotic distribution:

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

where I denotes the Fisher information matrix:

$$I = \begin{bmatrix} \kappa A(\kappa) & 0 \\ 0 & 1 - A(\kappa)^2 - A(\kappa)/\kappa \end{bmatrix}.$$

This states that for large sample size, the maximum likelihood estimator $\hat{\kappa}$ and $\hat{\mu}$ itself are approximately normally distributed with:

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

Moreover, $\hat{\kappa}$ and $\hat{\mu}$ are approximately independent for large n.

2.5 Goodness of Fit Test

A goodness of fit test describes how well a statistical model can fit a set of observations. There are several ways to test the fit. EDF (empirical distribution function) statistics, which are statistics measuring the difference between the empirical distribution function $F_n(\theta)$ and the cumulative distribution function $F(\theta)$, are some of the most common ways to test the fit (Stephens,1986). In this project, we test the fit using Watson's U^2 statistic, which is designed for the circle because its value does not depend on the origin of θ.

In general, let $\theta_1....\theta_n$ be a random sample drawn from some population. Suppose that we want to test the null hypothesis

$H_0 : \theta_1....\theta_n$ are from some distribution with the cumulative distribution function $F(\theta)$. Watson's U^2 is calculated as follows:

1. For each observation θ_i, calculate its cumulative distribution function $F(\theta_i)$, let $Z_i = F(\theta_i)$. For any unknown parameters, replace them by their maximum likelihood estimates.

2. Sort Z_i into ascending order to obtain $Z_{(i)}$, the order statistics.

3. Obtain U^2 statistics from:

$$U^2 = W^2 - n(\bar{Z} - 0.5)^2$$

where $\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$ and $W^2 = \sum_{i=1}^{n}\{Z_{(i)} - \frac{2i-1}{2n}\}^2 + \frac{1}{12n}$.

If the hypothesis is that $F(\theta)$ is the uniform distribution, the corresponding p-value can be found from the table in Stephens (1970). If the null hypothesis is that $F(\theta)$ is the von Mises distribution, Lockhart and Stephens (1985) give a table of significance points of the asymptotic distribution of U^2 for different values of κ.

If we are testing the fit for other distributions with unknown parameters, for example, a mixture of two von Mises distributions, the p-value can be calculated as follows:

1. Obtain the maximum likelihood estimates for unknown parameters and calculate U^2 as described above, let this value be U_0^2.

2. Generate N_{Boot} bootstrap samples from the distribution specified by $F(\theta)$, using the maximum likelihood estimates from the original sample as parameters.

3. For each bootstrap sample, re-estimate the parameters and obtain U^2, called U_j^2 , where $j = 1, ...N_{Boot}$.

4. Let N_G be the number of U_j^2 greater than U_0^2, and p-value$\approx \frac{N_G}{N_{Boot}}$.

2.6 Example

In this example, ants were placed individually into an arena, and an illuminated black target was placed at 180 degrees. The ants tend to move towards the target, and the orientations of ants were recorded. This data set was given by Fisher (1993, p.243), and are a random sample of size 100 taken from a larger data set (Jander, 1957, Figure 18A). The data are shown in Table 2.6.1. When we make calculations, the data are transformed from degrees to radians.

In Figure 2.6.1, the plot of the ants data is shown, and the density is estimated using a program in R. This is shown in Figure 2.6.2. On the plots, there is one clear mode shown around π. Therefore, we try to fit these data using a single von Mises distribution.

Table 2.6.1: Ants data (in degrees)

330	290	60	200	200	180	280	220	190	180
180	160	280	180	170	190	180	140	150	150
160	200	190	250	180	30	200	180	200	350
200	180	120	200	210	130	30	210	200	230
180	160	210	190	180	230	50	150	210	180
190	210	220	200	60	260	110	180	220	170
10	220	180	210	170	90	160	180	170	200
160	180	120	150	300	190	220	160	70	190
110	270	180	200	180	140	360	150	160	170
140	40	300	80	210	200	170	200	210	190

Table 2.6.2 shows the estimators of μ and κ obtained from maximum likelihood estimates and the method of moments. Figure 2.6.3 gives the fitted density using the estimated parameters calculated from these two methods against the original density estimate.

Table 2.6.2: Parameter estimates and U^2 for the ants data

	$\hat{\mu}$	$\hat{\kappa}$	U^2
MLE / Method of Moments	3.087	1.558	0.459

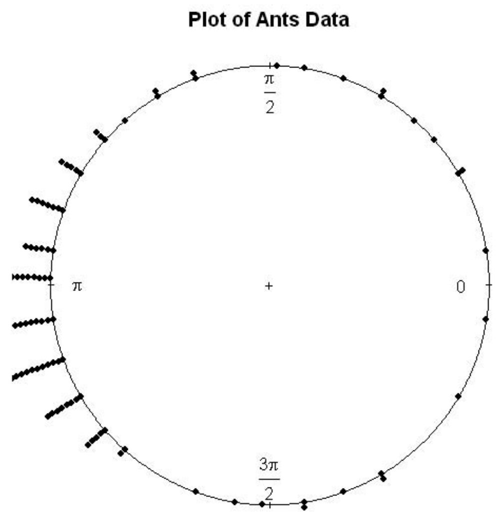

Figure 2.6.1: Plot of ants data

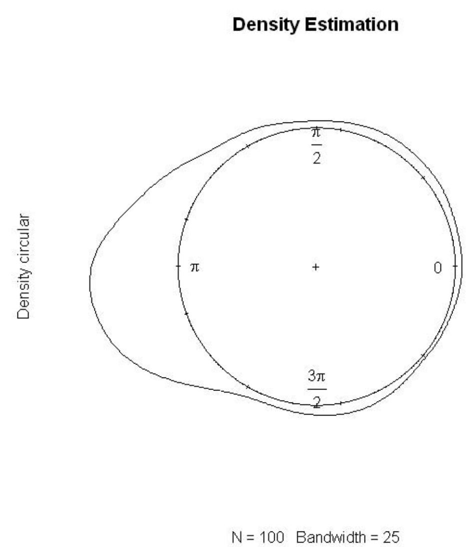

Figure 2.6.2: Non-parametric density estimation of ants data

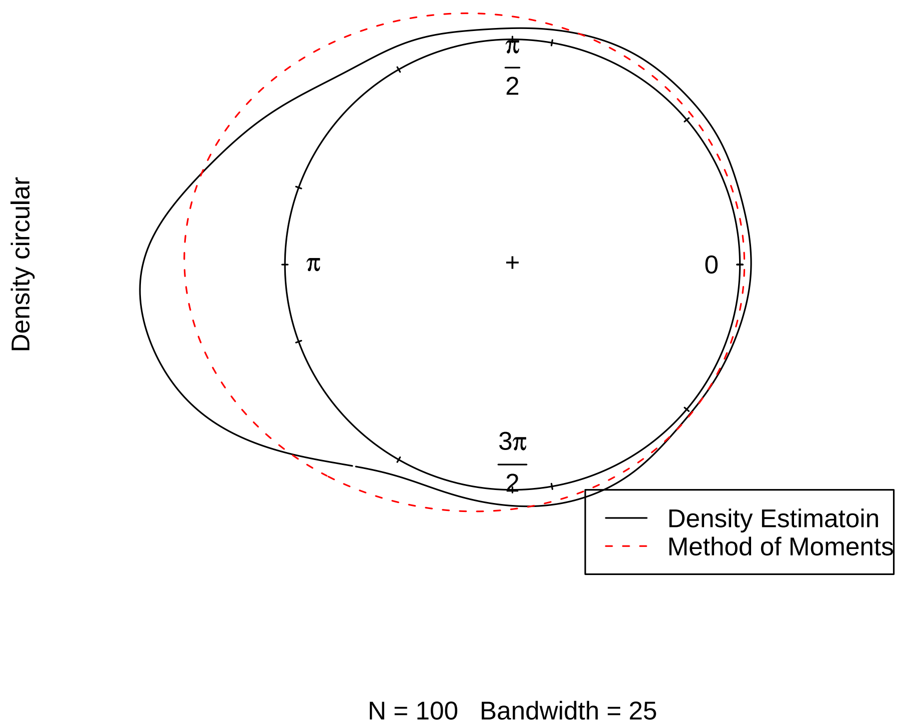

Figure 2.6.3: Comparison of non-parametric density estimation and fitted curves

Chapter 3

A Mixture of Two von Mises Distributions

In Chapter 2, we have discussed how to fit the data with a single von Mises distribution. For some circular data, instead of one mode, two modes are observed. In this situation, fitting the data with a mixture of two von Mises distributions could be a better choice. In Section 3.1, the probability density function of a mixture of two von Mises distributions will be introduced. Estimators of parameters will be discussed from Section 3.2 to Section 3.5.

3.1 Mixture of Two von Mises Distributions

Suppose that a circular random variable θ is from a mixture of two von Mises distributions; then its probability density function is:

$$P f_1(\theta) + (1-P) f_2(\theta), \qquad 0<P<1, \quad 0<\theta<2\pi, \tag{3.1.1}$$

where

$$f_j(\theta) = \frac{1}{2\pi I_0(\kappa_j)} \exp\left(\kappa_j \cos(\theta - \mu_j)\right), \qquad j = 1, 2. \tag{3.1.2}$$

Let $\theta_1 ... \theta_n$ be a sample of independently identically distributed random variables from a mixture of two von Mises distributions; an example of data and the probability density function is shown in Figure 3.1.1. In this example, 200 observations are generated from 3.1.1 with parameters $\mu_1 = \frac{1\pi}{6}$, $\mu_2 = \frac{5\pi}{6}$,$\kappa_1 = 3$,$\kappa_2 = 7$, and $P = \frac{2}{3}$.

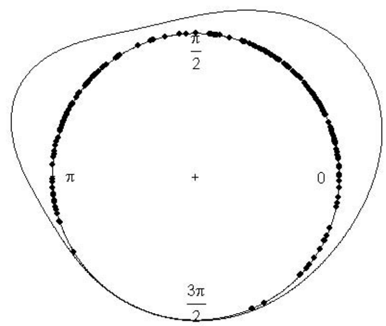

Figure 3.1.1: Plot of 200 observations generated from a mixture of two von Mises distributions

3.2 Maximum Likelihood Estimates

For the above sample, the likelihood function is:

$$L = \prod_{i=1}^{n} [Pf_1(\theta_i) + (1-P)f_2(\theta_i)], \tag{3.2.1}$$

where $f_i(\theta)$ is defined in Section 3.1. The log-likelihood function is:

$$l = \sum_{i=1}^{n} log[Pf_1(\theta_i) + (1-P)f_2(\theta_i)]. \tag{3.2.2}$$

By taking the partial derivatives with respect to each of five parameters, i.e, the score functions, and then setting these functions equal to zero, the maximum likelihood estimators can be found. They are the solutions of the following five equations:

$$\sum_{i=1}^{n} \frac{1}{Pf_1(\theta_i) + (1-P)f_2(\theta_i)}[f_1(\theta_i) - f_2(\theta_i)] = 0,$$

$$\sum_{i=1}^{n} \frac{Pf_1(\theta_i)}{Pf_1(\theta_i) + (1-P)f_2(\theta_i)} \kappa_1 \sin(\theta_i - \mu_1) = 0,$$

$$\sum_{i=1}^{n} \frac{(1-P)f_2(\theta_i)}{Pf_1(\theta_i) + (1-P)f_2(\theta_i)} \kappa_2 \sin(\theta_i - \mu_2) = 0,$$

$$\sum_{i=1}^{n} \frac{Pf_1(\theta_i)}{Pf_1(\theta_i) + (1-P)f_2(\theta_i)} [\cos(\theta_i - \mu_1) - A(\kappa_1)] = 0,$$

$$\sum_{i=1}^{n} \frac{(1-P)f_2(\theta_i)}{Pf_1(\theta_i) + (1-P)f_2(\theta_i)} [\cos(\theta_i - \mu_2) - A(\kappa_2)] = 0.$$

Even though the maximum likelihood equations can be obtained easily, the estimators can not be found analytically. However, the numerical solutions of estimates can be found by some computer languages.

When we are solving the maximum likelihood equations numerically, there is a risk that the estimate of μ_1 or μ_2 may be equal to one of the observations.

Suppose one of the observations is equal to $\hat{\mu}_1$, without loss of generality, we call it θ_1. The likelihood function L^* becomes:

$$L^* = \{\frac{\hat{P}}{2\pi I_0(\hat{\kappa}_1)} \exp(\hat{\kappa}_1) + (1-\hat{P})\exp[\hat{\kappa}_2(\hat{\theta}_1 - \hat{\mu}_2)]\} \prod_{i=2}^{n} \{\hat{P}f_1(\theta_i) + (1-P)f_2(\theta_i)\}. \quad (3.2.3)$$

Let $E_1 = \{\frac{\hat{P}}{2\pi I_0(\hat{\kappa}_1)} \exp(\hat{\kappa}_1) + (1-\hat{P})\exp(\hat{\kappa}_2(\hat{\theta}_1 - \hat{\mu}_2))\}$ be the first part of L^*, and let $E_2 = \prod_{i=2}^{n}\{\hat{P}f_1(\theta_i) + (1-\hat{P})f_2(\theta_i)\}$ be the second part of L^*.

Note that $\prod_{i=2}^{n} f_2(\theta_i)$ is a likelihood function from the second von Mises distribution, and there exists a positive number C , which is independent of $\hat{\kappa}_1$, such that $C < \prod_{i=2}^{n} f_2(\theta_i)$. So, $C^* = (1-\hat{P})C$ is smaller than E_2 , and C^* is independent of $\hat{\kappa}_1$.

The Bessel function $I_0(\hat{\kappa}_1)$ is approximately equal to $\frac{\exp(\hat{\kappa}_1)}{\sqrt{\hat{\kappa}_1}}$ when $\hat{\kappa}_1$ is large. This implies E_1 is approximately equal to $\frac{\hat{P}}{2\pi}\sqrt{\hat{\kappa}_1}$ when $\hat{\kappa}_1$ is large, and the likelihood function $L^* = E_1 * E_2$ is greater than $C^* \frac{\hat{P}}{2\pi}\sqrt{\hat{\kappa}_1}$. The likelihood function becomes unbounded with increasing values of $\hat{\kappa}_1$.

In Section 3.6 and Chapter 4, all the maximum likelihood estimates are obtained numerically using the R program, and comparisons of the maximum likelihood estimates and the method of moments estimates will be given in these sections.

3.3 Method of Moments Estimators

In general mixture model cases, the method of moments is not easily applicable; selecting an appropriate set of moments to estimate five parameters is a problem. For two von Mises mixture, the first two cosine and sine moments are used, but there is no way to construct the fifth equation in a symmetrical manner. However, the method of moments estimators can be found analytically for the special case of $\kappa_1 = \kappa_2$ and $\mu_1 = \mu_2 + \pi$ (detailed information will be given in Section 3.4).

Spurr and Koutbeiy (1991) described five methods to estimate parameters from a mixture of two von Mises distributions including MLE (their method 1). Their method 2 was based on the method of moments. The first four equations were chosen by using the first two sine and cosine moments (equation 3.3.1 and 3.3.2) below, and the derivative of the log likelihood with respect to P was chosen as the fifth equation 3.3.3. Estimates were obtained numerically by minimizing the sum of squares of residuals of these five equations. Their minimum should be zero since we have five unknown parameters and five equations.

$$P\frac{I_j(\kappa_1)}{I_0(\kappa_1)}\cos(j\mu_1) + (1-P)\frac{I_j(\kappa_2)}{I_0(\kappa_2)}\cos(j\mu_2) = \frac{1}{n}\sum_{i=1}^{n}\cos(j\theta_i), \quad j = 1, 2; \tag{3.3.1}$$

$$P\frac{I_j(\kappa_1)}{I_0(\kappa_1)}\sin(j\mu_1) + (1-P)\frac{I_j(\kappa_2)}{I_0(\kappa_2)}\sin(j\mu_2) = \frac{1}{n}\sum_{i=1}^{n}\sin(j\theta_i), \quad j = 1, 2; \tag{3.3.2}$$

$$\sum_{i=1}^{n}\{\frac{1}{2\pi I_0(\kappa_1)}\exp[\kappa_1\cos(\theta_i - \mu_1)] - \frac{1}{2\pi I_0(\kappa_2)}\exp[\kappa_2\cos(\theta_i - \mu_2)]\}\frac{1}{f(\theta_i)} = 0, \tag{3.3.3}$$

where $I_j(\kappa)$ is the modified Bessel function of the first kind and of the jth order, and $I_j(\kappa) = \frac{1}{\pi}\int_0^\pi cos(j\theta)\exp(\kappa\cos\theta)d\theta$.

Their method 3 is similar to method 2. Instead of using five equations, six equations, which are the first three sine and cosine moments, were used (equation 3.3.4 and 3.3.5):

$$P\frac{I_j(\kappa_1)}{I_0(\kappa_1)}\cos(j\mu_1) + (1-P)\frac{I_j(\kappa_2)}{I_0(\kappa_2)}\cos(j\mu_2) = \frac{1}{n}\sum_{i=1}^{n}\cos(j\theta_i), \quad j = 1, 2, 3; \tag{3.3.4}$$

$$P\frac{I_j(\kappa_1)}{I_0(\kappa_1)}\sin(j\mu_1) + (1-P)\frac{I_j(\kappa_2)}{I_0(\kappa_2)}\sin(j\mu_2) = \frac{1}{n}\sum_{i=1}^{n}\sin(j\theta_i), \quad j = 1, 2, 3. \tag{3.3.5}$$

3.4 Discussion of Special Case

Suppose we have $\theta_1 ... \theta_n$ independently identically distributed from a mixture of two von Mises distributions with $\kappa_1 = \kappa_2 = \kappa$ and $\mu_1 = \mu_2 + \pi$. The probability density function is:

$$f_s(\theta) = \frac{P}{2\pi I_0(\kappa)} \exp\{\kappa \cos(\theta - \mu_1)\} + \frac{1-P}{2\pi I_0(\kappa)} \exp\{-\kappa \cos(\theta - \mu_1)\}. \tag{3.4.1}$$

If we let $\theta^* = \theta$ (mod π), equation 3.4.1 can be reduced to

$$f_s^*(\theta^*) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos(\theta^* - \mu_1)\}, 0<\theta^*<\pi. \tag{3.4.2}$$

From equation 3.4.2, the random variable θ^* does not depend on parameter P . Mardia (1972, page 128) showed that the moments estimators $\hat{\mu}_1$, $\hat{\kappa}$ of μ and κ are the solutions of the following two equations:

$$\sum_{i=1}^{n} \sin 2(\theta_i^* - \hat{\mu}_1) = 0, \tag{3.4.3}$$

$$\frac{1}{n}\sum_{i=1}^{n} \cos 2(\theta_i^* - \hat{\mu}_1) = \frac{I_2(\hat{\kappa})}{I_0(\hat{\kappa})}. \tag{3.4.4}$$

Mardia also gave the moments estimator $\hat{P}$ of P , which is the solution of

$$(2\hat{P} - 1)A(\hat{\kappa}) = \frac{1}{n}\sum_{i=1}^{n} \cos 2(\theta_i^* - \hat{\mu}_1) = \bar{C} \cos \hat{\mu}_1 + \bar{S} \sin \hat{\mu}_1, \tag{3.4.5}$$

where $\bar{C}$ and $\bar{S}$ are defined in Section 2.3.

The MLE can not be solved analytically even for the special case $\kappa_1 = \kappa_2 = \kappa$ and $\mu_1 = \mu_2 + \pi$. Using a combination of the gradient method and the Newton-Raphson method, the maximum likelihood estimates were obtained by Jones and James (1969).

3.5 Other Methods of Estimation

Spurr and Koutbeiy (1991) described five methods to estimate parameters from a mixture of two von Mises distributions, and we have seen three of them in Section 3.4. Their method 4 minimized the sum of square distances between the empirical characteristic function and the characteristic function of the model, which minimized

$$\sum_{t=1}^{\infty} |\Phi(t) - \tilde{\Phi}_n(t)|^2,$$

where $\tilde{\Phi}_n(t) = \frac{1}{n}\sum_{j=1}^{n} \exp(it\theta_j)$ is the empirical characteristic function, and $\Phi(t)$ is the characteristic function:

$$\Phi(t) = P[\cos(t\mu_1) + i\sin(t\mu_1)]\frac{I_t(\kappa_1)}{I_0(\kappa_1)} + (1-P)[\cos(t\mu_2) + i\sin(t\mu_2)]\frac{I_t(\kappa_2)}{I_0(\kappa_2)}.$$

Their method 5 based on the Cramér-von Mises statistics W^2 (introduced in Section 2.5), which minimized $\sum_{j=1}^{n}[F(\theta_{(j)}) - \frac{j-0.5}{n}]^2$, where $\theta_{(j)}$ is the jth order statistic of the sample.

3.6 Examples

Example 1

Let us first consider the special case when $\kappa_1 = \kappa_2 = \kappa$ and $\mu_1 = \mu_2 + \pi$.

In this example, we will use Gould's turtle data given by Stephens (1969). In Table 3.6.1, orientations of 76 turtles after treatment are listed. It is observed that there are two modes and they are roughly 180 degrees apart. There is one big mode around 60 degrees and a small mode around 240 degrees.

Table 3.6.1: Orientations of 76 turtles after treatment (in degrees)

8	30	48	58	65	83	95	118	223	251
9	34	48	58	68	88	96	138	226	257
13	38	48	61	70	88	98	153	237	268
13	38	48	63	73	88	100	153	238	285
14	40	50	64	78	90	103	155	243	319
18	44	53	64	78	92	106	204	244	343
22	45	56	64	78	92	113	215	250	350
27	47	57	65	83	93				

The estimates from the method of maximum likelihood and the method of moments estimates (MME) for this example are shown in Table 3.6.2. In this example, estimates obtained from both methods are quite close. Watson's U^2 for MLE is equal to 0.0195, and equal to 0.0558 by MME. The corresponding p-values are greater than 0.5 for both methods.

Table 3.6.2: Parameter estimates for turtle data ($\hat{\mu}$ in degrees)

	$\hat{P}$	$\hat{\mu}$	$\hat{\kappa}$	U^2
MLE	0.81	63.1	3.0	0.019
MME	0.85	62.4	3.6	0.056

Example 2

In example two, we will consider a mixture of two von Mises distributions in the general case. We generate 200 observations from a mixture of two von Mises distributions, and the parameters and their estimates are shown in Table 3.6.3.

Table 3.6.3: The maximum likelihood estimates (MLE) and the method of moments estimates (MME)

Parameters	P	μ_1	μ_2	κ_1	κ_2
True value	0.67	0.52	2.62	3	7
MLE	0.63	0.60	2.68	2.84	6.77
MME	0.65	0.63	2.69	2.77	8.98

In Table 3.6.3, the method of moments estimates are obtained by minimizing the sum of squares of residuals from six equations, which are constructed by using the first three moments, i.e. we minimize $S_1^2 + S_2^2$, where:

$$S_1 = \sum_{t=1}^{3}[P\frac{I_1(\kappa_1)}{I_0(\kappa_1)}\cos(t\mu_1) + (1-P)\frac{I_1\kappa_2}{I_0(\kappa_2)}\cos(t\mu_2) - \frac{1}{n}\sum_{i=1}^{n}\cos(t\theta_i)]^2, \tag{3.6.1}$$

$$S_2 = \sum_{t=1}^{3}[P\frac{I_1(\kappa_1)}{I_0(\kappa_1)}\sin(t\mu_1) + (1-P)\frac{I_1(\kappa_2)}{I_0(\kappa_2)}\sin(t\mu_2) - \frac{1}{n}\sum_{i=1}^{n}\sin(t\theta_i)]^2. \tag{3.6.2}$$

In Table 3.6.3, both the maximum likelihood method and the method of moments give estimates close to true values.

For the maximum likelihood method, Watson's $U^2 = 0.0263$. We simulate 5000 bootstrap samples with sample size $n = 200$. Then, 1593 out of 5000 give values greater than 0.0263. Under the hypothesis that sample is from a mixture of two von Mises distributions, we obtain p-value= $1593/5000 = 0.3186$. With type one error chosen as 0.05, we fail to reject the hypothesis. For the method of moments, Watson's $U^2 = 0.0271$, and p-value= $1808/5000 = 0.3616$. Therefore, we fail to reject the hypothesis here too. Both of these two methods give good estimates fitting the data well. The density estimation and fitted density functions for both the maximum likelihood method and the method of moments are shown in Figure 3.6.1.

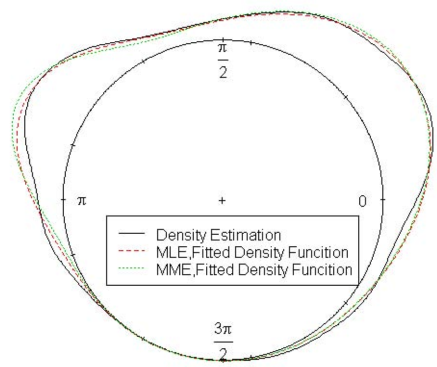

Figure 3.6.1: Non-parametric density estimation and fitted densities for Example 2

Table 3.6.4 and Table 3.6.5 show the mean, variance, and the third central moment of the estimates for both of the method of moments and the method of maximum likelihood. From these two tables, the variances for $\hat{P}$ and $\hat{\mu}$'s are similar; however, for the method of moments, the variances of $\hat{\kappa}$'s are much larger than that from the method of maximum likelihood. Therefore, maximum likelihood method gives more accurate estimates.

Table 3.6.4: Means, variances and third moments for MLE

MLE	Mean	Variance	Third Moment
$\hat{P}$	0.63	0.0016	3.85E-06
$\hat{\mu}_1$	0.60	0.0048	3.49E-05
$\hat{\mu}_2$	2.68	0.0034	-3.38E-05
$\hat{\kappa}_1$	2.94	0.2153	0.0646
$\hat{\kappa}_2$	7.17	3.26	14.09

Table 3.6.5: Means, variances and third moments for MME

MLE	Mean	Variance	Third Moment
$\hat{P}$	0.65	0.0018	1.09E-06
$\hat{\mu}_1$	0.63	0.0060	0.0001
$\hat{\mu}_2$	2.69	0.0031	-2.96E-05
$\hat{\kappa}_1$	2.87	0.2561	0.0690
$\hat{\kappa}_2$	10.89	48.23	1188.49

Example 3

In this example, we will generate a sample with two modes but not from a mixture of two von Mises distributions.

If we try to fit this sample with a mixture of two von Mises distributions, the maximum likelihood estimates are shown in Table 3.6.6.

Table 3.6.6: Maximum likelihood estimates for Example 3

Parameters	P	μ_1	μ_2	κ_1	κ_2
MLE	0.27	0.40	1.72	15.39	0.76

Watson's $U^2 = 0.0667$ in this example. We simulate 5000 bootstrap samples using the estimates above, and calculate Watson's U^2 for each. In these 5000 U^2 values, only six have values greater than 0.0667. Therefore p-value$= 8/5000 = 0.0016$. With type one error 0.05, we reject the hypothesis. The density estimate and the fitted density are shown in Figure 3.6.2.

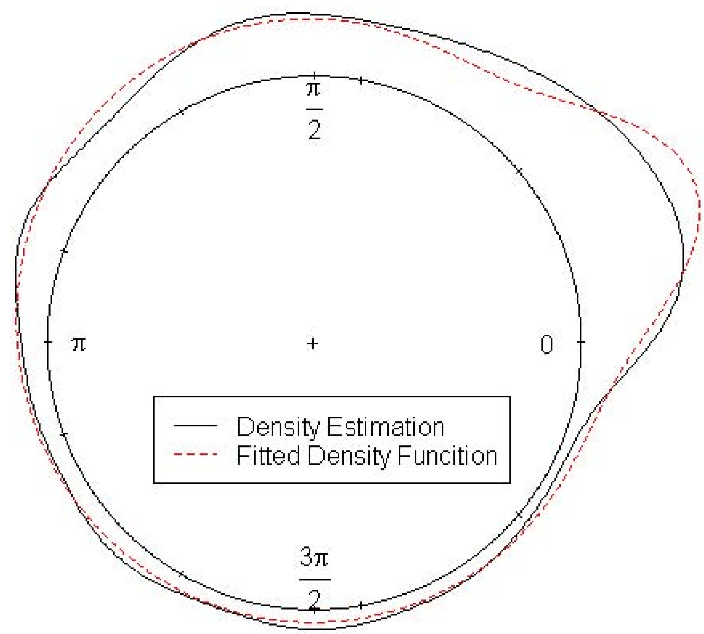

Figure 3.6.2: Non-parametric density estimation and fitted density for Example 3

Chapter 4

SIDS Data and Traffic Crash Data Analysis

In this chapter, we will analyze two data sets by fitting the data to the von Mises distribution and a mixture of two von Mises distributions. We will analyze Mooney's SIDS (Sudden infant death syndrome) data in Section 4.1; and in Section 4.2, we will discuss traffic crash data in the United States.

4.1 SIDS

4.1.1 Introduction

Sudden infant death syndrome (SIDS) is a syndrome marked by sudden and unexplained death of an apparently healthy infant aged one month to one year. The term cot death is often used in the United Kingdom, Australia and New Zealand, while crib death is sometimes used in North America. Mooney, Helms, and Jollife (2003), referred to as MHJ in the future, have collected SIDS data for the UK from 1983 to 1998. In their study, they pointed out that, for some years, there seems to be more than one mode for SIDS data, and a mixture of von Mises distributions should be fitted. The data set records the number of infant death cases by month, and it is shown in Table 4.1.1.

For the SIDS data, we map the time in one year from zero to 2π in radians, and give every month an equal length on the circle. Since the SIDS data given by MHJ are grouped, we vary it by adding random numbers on each cell. The way we add randomness

Table 4.1.1: Number of SIDS cases in UK

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
1983	159	196	147	117	133	71	70	67	94	119	160	190	1523
1984	177	181	191	103	83	94	67	45	83	132	115	150	1421
1985	163	150	132	146	100	79	77	54	92	128	190	193	1504
1986	183	202	173	149	141	89	84	76	97	134	168	222	1718
1987	203	180	181	127	108	120	85	84	88	147	167	242	1732
1988	212	194	175	134	136	102	101	99	114	158	165	188	1778
1989	184	143	180	112	118	76	77	82	99	131	163	161	1526
1990	174	140	141	127	91	87	73	68	104	92	111	163	1371
1991	115	154	133	103	104	81	65	52	70	68	75	91	1111
1992	61	60	45	54	50	43	43	36	37	53	60	42	584
1993	66	53	48	46	46	34	29	35	44	59	43	50	553
1994	65	54	36	38	45	51	35	31	49	45	28	50	527
1995	38	44	57	35	36	39	40	29	35	37	33	49	472
1996	53	49	66	29	38	36	47	28	36	29	40	52	503
1997	48	36	53	47	36	33	40	28	42	36	45	48	492
1998	40	28	25	25	29	32	25	20	26	40	42	63	395

is by generating an amount (equal to the count) of numbers for each month from the uniform distribution. For example, in January 1996, we generate 53 independent identical random variables from the uniform distribution from 0 to $\frac{\pi}{6}$; and in February 1996, we generate 49 uniform random variables from $\frac{\pi}{6}$ to $\frac{\pi}{3}$.

4.1.2 Model Selection

There are two common ways to select models. One way is to start with the simplest model and gradually adopt more complex models if there is significant evidence that the simpler model could not be fitted well. The other way is to start with a complex model and gradually adopt a simpler model.

In this project, we start with the simplest model. Firstly, we test if the data are from the uniform distribution. If the goodness of fit test showed that there was significant evidence that the uniform distribution could not be fitted, we try to fit the von Mises distribution; otherwise, we would stop. If the von Mises distribution could not be fitted, we try to fit a mixture of two von Mises distributions. Figure 4.1.1 shows detailed information of how we do the model selection.

Two remarks should be given here. First of all, if the uniform distribution failed to be rejected, the von Mises distribution or a mixture of two von Mises distributions should also fail to be rejected since the von Mises distribution will become the uniform distribution when κ equals zero. Secondly, we will not fit a mixture of the uniform distribution and the von Mises distribution since it is a special case of a mixture of two von Mises distributions when either $\kappa_1 = 0$ or $\kappa_2 = 0$.

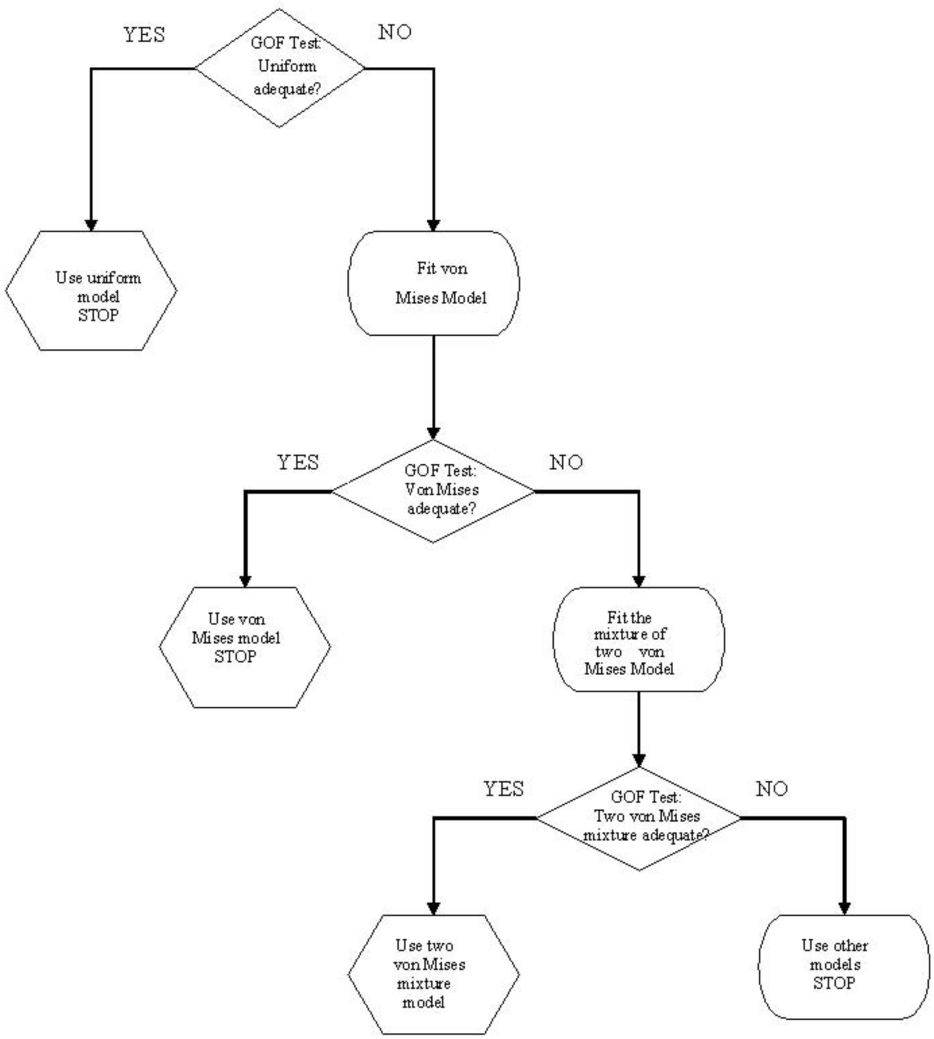

Figure 4.1.1: Model selection procedure

4.1.3 Example from 1996

In this section, we take the SIDS data from 1996 as an example.

In 1996, there were 503 sudden infant death syndrome observations in the UK. The plot of randomly replaced data and their density estimation are shown in Figure 4.1.2.

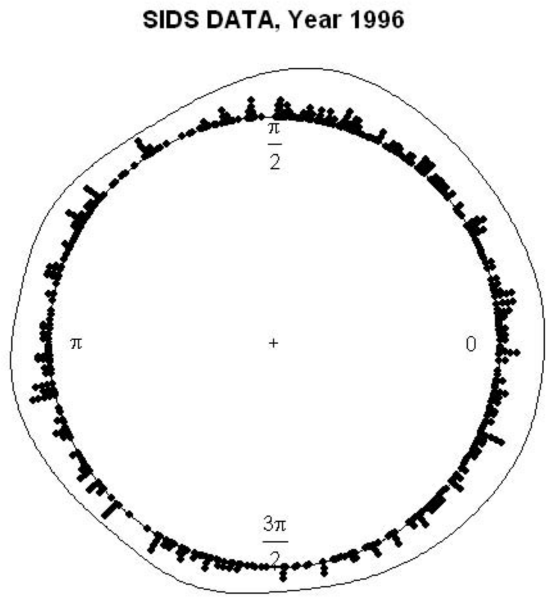

Figure 4.1.2: Plot of SIDS data from 1996

We firstly test uniformity. Under the hypothesis that the data are from the uniform distribution, Watson's $U^2 = 0.395$. The corresponding p-value, obtained from Stephens's table (1970), is smaller than 0.001. Therefore, the hypothesis is rejected.

We then try to fit the von Mises distribution, and the results are shown in Table 4.1.2.

Table 4.1.2: SIDS data from 1996 by the von Mises distribution

	μ	κ	U^2	AIC
MLE	0.763	0.230	0.0608	1839.751

The maximum likelihood estimates for μ is equal to 0.763, and equals 0.230 for κ. Watson's $U^2 = 0.0608$ under the hypothesis that the data are from the von Mises distribution, and the corresponding p-value obtained from the table given by Lockhart and Stephens (1985), is between 0.05 to 0.10. From Figure 4.1.2, there seems to be no clear peak around 0.763. Considering all the results, we reject the single von Mises distribution.

Now, we try to fit a mixture of two von Mises distributions. The results are shown in Table 4.1.3.

Table 4.1.3: SIDS data from 1996 by a mixture of two von Mises distributions

	$\hat{P}$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	AIC
MLE	0.926	0.522	3.259	0.384	7.436	1838.367

The MLE results show that about 93% of the data are from the first von Mises distribution with a peak around 30 days, and this peak is quite flat since its corresponding κ is very small (equal to 0.384). About 7% of the data are from the other von Mises distribution with a peak around 180 days.

Watson's $U^2 = 0.0315$ for the mixture model, and we simulate 5000 bootstrap samples. More than 800 out of 5000 give U^2 values greater than 0.0315. We obtain p-value greater than 0.15. The null hypothesis fails to be rejected. In other words, a mixture of two von Mises distribution is an acceptable model for the data.

Akaike information criterion (AIC) is calculated as follows: $AIC = 2N_p - 2\log(L)$, where N_p is the number of parameters. Several statistical models can be ranked according to their AIC for a given data set. The one having the lowest AIC is the best. For the single von Mises model, we obtained an AIC value equal to 1839.751, and equal to 1838.367 for mixture model. The mixture model has a slightly smaller AIC. However, the goodness of fit test shows the mixture model is better than the single von Mises.

Table 4.1.4: Normalized SIDS data from 1996 by a mixture of two von Mises distributions

Parameters	P	μ_1	μ_2	κ_1	κ_2
MLE	0.944	0.538	3.268	0.360	21.726

If we normalize the days by using 31 days in January and 29 days in February and so on in 1996, the new estimates for the five parameters are shown in Table 4.1.4. All the

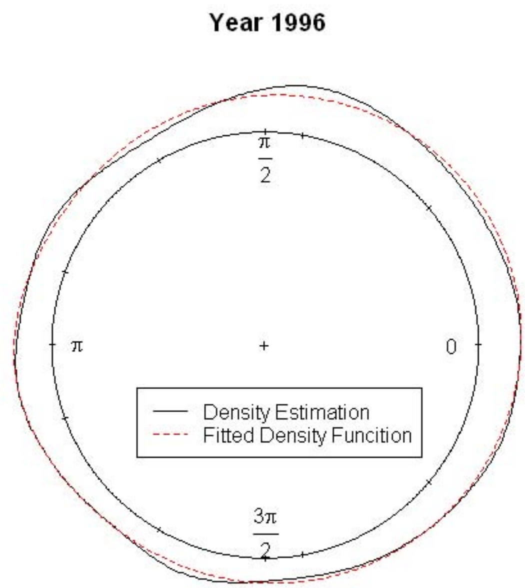

Figure 4.1.3: SIDS data, 1996; fitted density function against non-parametric density estimation

estimates of the five parameters except κ_2 are close to those given as above, when the months were divided into $\pi/6$ segments equally around the circle.

4.1.4 Analysis of SIDS Data: Summary

In this section, we analyze the full SIDS data set from 1983 to 1991, and separately, from 1992 to 1998. This is because there is a huge change for the SIDS monthly counts in 1992. We combine the monthly counts of years before 1992 together, and also for the counts of years after 1992.

We try to fit the data of each year to the uniform distribution. Under the hypothesis that our observations are from the uniform distribution, we obtain significantly large Watson's U^2 for all years. The hypothesis is rejected for the type one error chosen as 0.05. We next fit the data using the von Mises distribution, and the results are shown in Table 4.1.5.

Table 4.1.5: Results of SIDS data by fitting the von Mises distribution

Year	$\hat{\mu}$	$\hat{\kappa}$	U^2	p-value	Reject H_0
1983	0.35	0.45	0.341	<0.05	Yes
1984	0.49	0.51	0.103	<0.05	Yes
1985	0.15	0.45	0.690	<0.05	Yes
1986	0.41	0.45	0.478	<0.05	Yes
1987	0.26	0.45	0.405	<0.05	Yes
1988	0.28	0.35	0.031	>0.1	No
1989	0.27	0.39	0.068	<0.05	Yes
1990	0.43	0.39	0.026	>0.1	No
1991	1.01	0.42	0.121	<0.05	Yes
1992	0.50	0.16	0.073	<0.05	Yes
1993	0.20	0.24	0.084	<0.05	Yes
1994	0.51	0.15	0.211	<0.05	Yes
1995	0.92	0.15	0.049	>0.1	No
1996	0.76	0.23	0.061	(0.05, 0.1)	Further Discussion
1997	0.51	0.18	0.076	<0.05	Yes
1998	5.92	0.35	0.383	<0.05	Yes
Before 1992	0.38	0.40	0.620	<0.05	Yes
After 1992 (include)	0.34	0.17	0.057	(0.05, 0.1)	Further Discussion

Based on the goodness of fit test, we obtain a reasonably small Watson's U^2 under the hypothesis that the data are from the von Mises distribution in years 1988, 1990, and 1995. We fail to reject the von Mises distribution in these years. For 1996 and the mixture of data after 1992, we obtain a p-value between 0.05 and 0.1. The von Mises distribution may not give a good fit. We will also fit a mixture of two von Mises distributions for these two data sets and hope to see better results. The p-values for all the other years are small, and the hypothesis are rejected at $\alpha = 0.05$.

Table 4.1.6 shows the estimates of parameters and the p-value when we fit a mixture of two von Mises distributions. The p-value is calculated by the same process in Section 2.5 with 5000 bootstraps.

Regarding the data of years which have rejected the von Mises distribution, only the data of 1989 and the combination of data before 1992 reject a mixture of two von Mises distribution with α =0.05. For all other years, a mixture of two von Mises distributions give good fits. For most years where a mixture of two von Mises distributions fits well,

Table 4.1.6: Results of SIDS data by fitting a mixture of two von Mises distributions

Year	$\hat{P}$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	p-value	Reject H_0
1983	0.96	0.20	2.30	0.51	19.41	0.17	No
1984	0.96	0.64	5.11	0.54	27.24	0.08	Further Discussion
1985	0.56	1.59	5.70	0.68	1.52	0.38	No
1986	0.83	0.99	5.76	0.43	2.02	0.34	No
1987	0.87	0.71	5.85	0.36	2.72	0.20	No
1989	0.36	1.41	5.77	0.86	0.59	0.01	Yes
1991	0.98	1.04	4.50	0.47	9.95	0.21	No
1992	0.95	1.03	5.10	0.20	8.20	0.76	No
1993	0.94	0.63	4.84	0.30	11.28	0.42	No
1994	0.12	0.26	3.06	7.53	0.13	0.35	No
1996	0.93	0.52	3.26	0.38	7.44	0.18	No
1997	0.94	1.22	5.90	0.15	8.26	0.23	No
1998	0.46	2.72	5.89	0.76	1.56	0.83	No
Before 1992	0.75	1.06	5.68	0.45	1.15	0.02	Yes
After 1992(include)	0.97	0.25	3.13	0.24	8.54	0.29	No

there is one peak around 1 to 1.5 (February to March), and the other peak around 5.5 to 6 (November to December). Final comment: the SIDS data is a time series, and might be analyzed using time series methods. The advantage using a circle is that peaks may occur at different times in different years, and this may not be recognized by standard time series analysis. Also, the calendar year could be broken at say, June or July, in order to determine the peaks, because these months are far from the peaks.

4.2 Traffic Crash Data Analysis

In this section, we will analyze the traffic crash data in the United States in 2007. The data are obtained from the Fatality Analysis Reporting System Encyclopedia webpage (FARS Encyclopedia, 2008). The database records the traffic crash information of each state in the US from 1994 to 2007, and the crash time are accurate to seconds. In this project, we will choose some fatal crash data from three states in the US in 2007, namely Washington, New York, and the District of Columbia.

We map each crash time in 24-hour periods onto a unit circle from 0 to 2π. Every 15 degrees on the circle denotes 1 hour in real time. Midnight (12:00 a.m.) is mapped to 0 on the circle.

4.2.1 Fatal Crash Data Analysis: Washington State

In Washington State, 528 fatal crashes (after deleting missing or incomplete data) are recorded in 2007, and the distribution of observations and their density estimation are shown in Figure 4.2.1.

There are two peaks in this data set: a smaller one around $\frac{\pi}{6}$, and a larger one around $\frac{3\pi}{2}$. We will fit a mixture of two von Mises distributions for this data set.

Table 4.2.1: MLE for fatal crash data, Washington State

Parameters	P	μ_1	μ_2	κ_1	κ_2
MLE	0.044	0.54	4.70	134.66	0.48

Table 4.2.1 shows the estimates of fitting the data using a mixture of two von Mises distributions, and there is one mode at 0.54, which is about 2:00 a.m.. The proportion of the first von Mises distribution is small, only 4.4%. The variation for the first mode is small since the estimate of κ is large, namely 134.66. Therefore, the standard deviation for the first mode is about $(\hat{\kappa})^{-0.5} = 0.086$, which is about 20 minutes. The second mode is at 4.70 (about 6:00 p.m.). The proportion of the second mode is large (95.6%). The variation of the second mode is large too since κ_2 is very small.

Goodness of fit testing gives Watson's $U^2 = 0.0278$. We simulate 5000 bootstrap samples using the estimates in Table 4.2.1 and calculate Watson's U^2 for each. There are more than 2000 U^2 values from bootstrap samples greater than 0.0278. Therefore, our p-value is greater than 0.4. A mixture of two von Mises distributions does give a good fit for this data set.

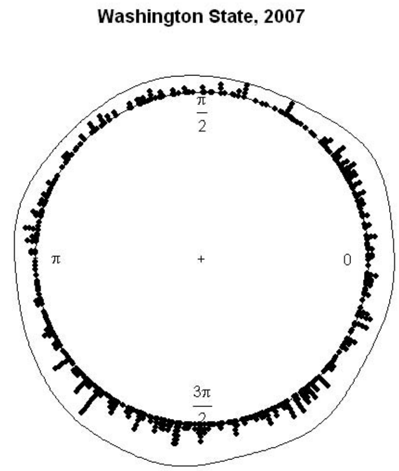

Figure 4.2.1: Fatal crash data of Washington State in 2007; data plot

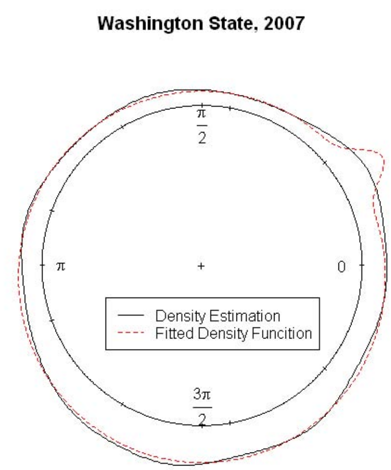

Figure 4.2.2: Fatal crash data of Washington State in 2007; fitted densities

4.2.2 Fatal Crash Data Analysis: District of Columbia

For the fatal crash data set of the District of Columbia in 2007, there are only 36 observations. The data distribution is shown in Figure 4.2.3, and there is one clear peak around $\frac{7\pi}{4}$ (9:00 p.m.).

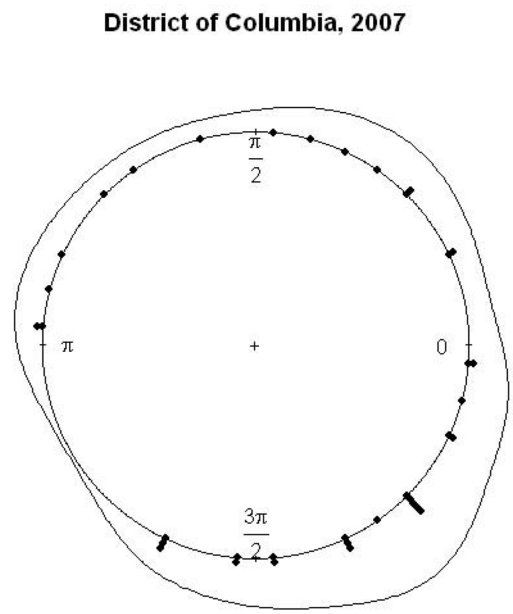

Figure 4.2.3: Fatal crash data of the District of Columbia in 2007; data plot

The estimates of parameters are shown in Table 4.2.2 when we fit a mixture of two von Mises distributions. MLE of μ_1 equals to 1.29 (about 5:00 a.m.) and MLE of μ_2 equals to 5.24 (about 8:00 p.m.).

Table 4.2.2: MLE for fatal crash data, District of Columbia

Parameters	P	μ_1	μ_2	κ_1	κ_2
MLE	0.56	1.29	5.24	0.62	3.68

The test result gives Watson's $U^2 = 0.0184$, and the p-value is greater than 0.5.

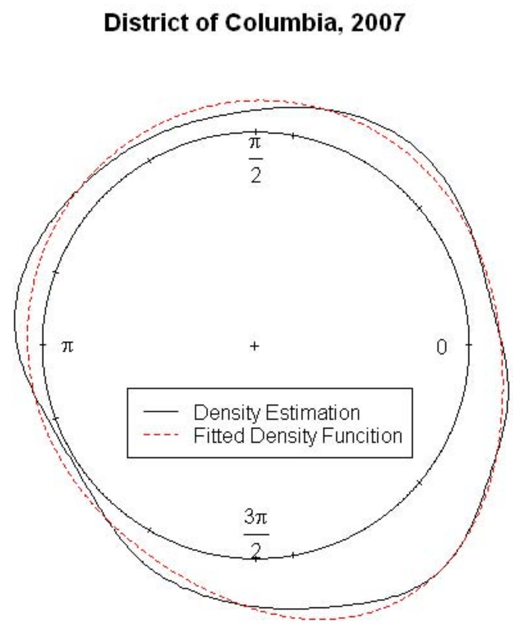

Figure 4.2.4: Fatal crash data of District of Columbia in 2007; fitted densities

4.2.3 Fatal Crash Data Analysis: New York State

Now we come to the fatal crash data set of New York State in 2007. After deleting missing values and incomplete records, there are 1236 observations in this data set. Their density estimation is given in Figure 4.2.5. There is only one clear peak, which appeared around $\frac{5\pi}{4}$ (3:00 p.m.). We firstly fit a single von Mises distribution.

Table 4.2.3: MLE for fatal crash data, New York State

	$\hat{\mu}$	$\hat{\kappa}$	U^2
Values	4.61	0.26	0.056

Table 4.2.3 shows that if a single von Mises distribution is fitted, the mode is at 4.61 (about 5:30 p.m.). Watson's $U^2 = 0.0561$, and the corresponding p-value is about 0.09. It fails to reject the hypothesis with type one error chosen as 0.05.

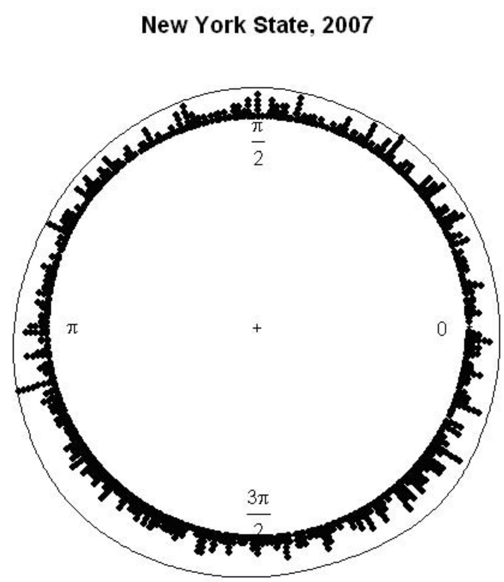

Figure 4.2.5: Fatal crash data of New York State in 2007; data plot

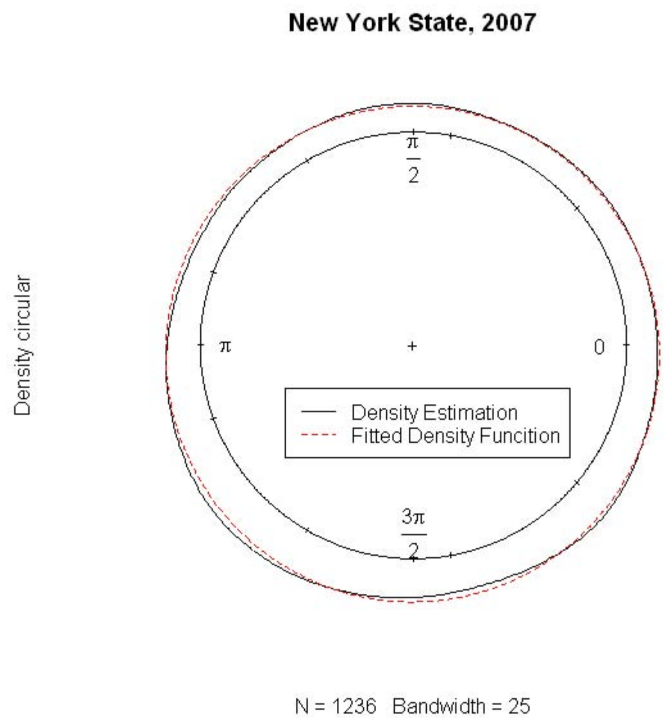

Figure 4.2.6: Fatal crash data of New York State in 2007; fitted densities

Chapter 5

Computational Details

5.1 Initial Values

In this project, all the computations are done by the R program. We obtain the maximum likelihood estimates by minimizing the negative log likelihood function and by using the in-built function "nlminb" in R. Initial values are required when we call the function "nlminb". Usually, if observations do come from a mixture of two von Mises distributions with two clear modes, the same estimates will be obtained even though we start at different initial values. However, if there are not two clear modes or there are more than one local maximum for the likelihood function, different initial values may give different results. In this project, all the initial values for κ's are chosen as 4, and $\frac{1}{2}$ for P. For the initial values of μ, we choose the values where the peak(s) roughly are located by looking at the plots of data or the density estimation plots.

Table 5.1.1 shows the results of maximum likelihood estimates starting with different initial values of μ, using the data in Example 2 in Chapter 3. All the different initial values give almost the same results (there are slight differences if we keep more than 5 decimal digits, and the differences may be caused by computational error).

For some data sets, the estimates do not converge to the same values if we pick different initial values. There are two main cases. In the first case, the estimates may go outside the parameter space for some initial values. In the second case, different initial values may give more than one result, and all these results are inside the parameter space. In other words, there are more than one local maxima for the likelihood function. We can treat the first case as a kind of error, and it can be easily checked. The second case cannot

Table 5.1.1: Comparing different initial values (Example 2 in Chapter 3)

Initial Values $\kappa_1 = \kappa_2 = 4, P = 0.5$	$\hat{P}$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\kappa}_1$	$\hat{\kappa}_2$
$\mu_1 = \pi/6, \mu_2 = 5\pi/6$	0.630178	0.596674	2.679197	2.838340	6.765724
$\mu_1 = \pi/6, \mu_2 = \pi/3$	0.630178	0.596674	2.679197	2.838341	6.765737
$\mu_1 = \pi/6, \mu_2 = \pi$	0.630178	0.596674	2.679197	2.838340	6.765724
$\mu_1 = \pi/6, \mu_2 = 3\pi/2$	0.630178	0.596674	2.679197	2.838341	6.765737
$\mu_1 = \pi/2, \mu_2 = \pi$	0.630178	0.596674	2.679197	2.838340	6.765724
$\mu_1 = \pi/2, \mu_2 = 3\pi/2$	0.630178	0.596674	2.679197	2.838341	6.765737
$\mu_1 = \pi, \mu_2 = 3\pi/2$	0.630178	0.596674	2.679197	2.838340	6.765724
$\mu_1 = 3\pi/2, \mu_2 = 7\pi/4$	0.630178	0.596674	2.679197	2.838341	6.765737

be checked easily, and we cannot guarantee that there is a unique local maxima for the likelihood function for a new data set.

Table 5.1.2 shows an example of different initial values giving different results. This example is based on the SIDS data from 1984. There are at least two local maxima of the likelihood function for this data set. In Chapter 4, we choose the second one ($\mu_1 = 0.64, \mu_2 = 5.11$) to fit the data because it has a larger likelihood function and smaller Watson's U^2.

Table 5.1.2: Comparing different initial values (SIDS data, 1984)

Initial Values $\kappa_1 = \kappa_2 = 4, P = 0.5$	$\hat{P}$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	Log Likelihood	U^2
$\mu_1 = \pi/3, \mu_2 = \pi$	0.98	0.45	2.96	0.54	15.63	-2523	0.0985
$\mu_1 = \pi/3, \mu_2 = 7\pi/4$	0.96	0.64	5.11	0.54	27.24	-2518	0.0491

5.2 Large κ

For some data sets, when a mixture of two von Mises distributions is fitted, one κ estimate will be large; this κ belongs to the von Mises distribution with the smaller proportion.

Figure 5.2.1 shows the distribution of fatal crash data of Washington State in 2007 that was discussed in Chapter 4. For this data set, we can see that there is one large peak around $\frac{3\pi}{2}$, and about 96% of the data are from this von Mises distribution. There is another cluster of observations around $\frac{\pi}{6}$. This cluster is small (only about 30 observations) but dense. Therefore, we will obtain a large κ for the second peak. When a mixture of two von Mises distributions is fitted, a large κ will often appear with small proportion for one of the two von Mises distributions.

Moreover, if we obtain a large estimate of one κ, the likelihood function will not change much when the large κ increases. Figure 5.2.2 shows the relationship of κ_1 and the likelihood function of fatal crash data of Washington State in 2007.

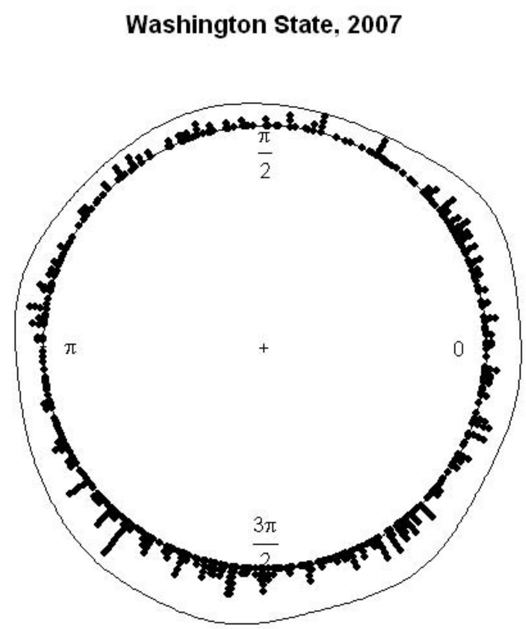

Figure 5.2.1: Fatal crash data of Washington State in 2007; data plot

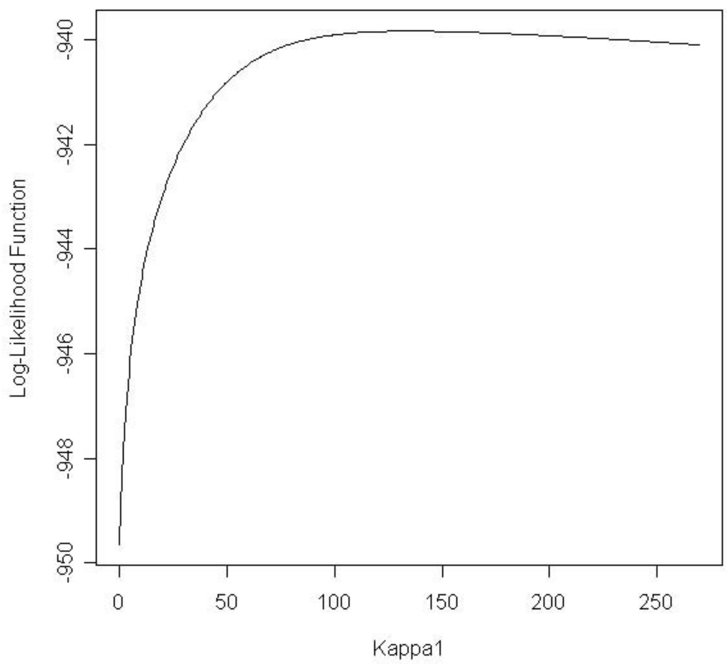

Figure 5.2.2: Relationship between κ_1 and likelihood function, Washington State data, 2007

5.3 Properties of Estimates

Table 5.3.1 shows the means, variances, and the third moments of the estimates from the data set of Example 2 in Chapter 3. The means, variances, and the third moments are calculated numerically from 5000 simulation samples. With the first three moments, we can fit the approximate distribution of a parameter in the form $a + b\chi^2_P$, and solve for a, b, and P.

Table 5.3.1: Sample estimate and three moments of simulated samples. (Example 2 in Chapter 3)

Parameter	Estimate	Mean	Variance	Third Moment
P	0.63	0.63	0.0016	3.85E-06
μ_1	0.60	0.60	0.0048	3.49E-05
μ_2	2.68	2.68	0.0034	-3.38E-05
κ_1	2.84	2.94	0.2153	0.0646
κ_2	6.77	7.17	3.26	14.09

From Table 5.3.1, the sample means of 5000 simulated samples are close to the original estimates; the variances and the third moments of the simulated estimates are small except for κ_2, the one from the von Mises distribution with smaller proportion.
Table 5.3.2 shows the moments for the fatal crash data of Washington State in 2007.

Table 5.3.2: Sample estimate and three moments of simulated samples, fatal crash data of Washington State in 2007

Parameter	Estimate	Mean	Variance	Third Moment
P	0.044	0.049	0.0003	7.99E-06
μ_1	0.54	0.54	0.0021	0.0005
μ_2	4.70	4.68	0.0204	0.0003
κ_1	134.66	166.70	10868.70	1086297.14
κ_2	0.48	0.49	0.0051	0.0001

Table 5.3.2 also shows that the variances and the third moments of the estimates of P, μ_1, μ_2, and κ_2 are reasonably small. The variation of κ_1 is quite large, while the proportion of first von Mises distribution is small.

Chapter 6

Conclusions

In Chapter 3, we have discussed how to fit a mixture of two von Mises distributions. The maximum likelihood estimates of the five parameters can be found only numerically. The method of moments estimators can be found analytically only under the assumption that the two μ's are opposite and the two κ's are equal. When we are finding the maximum likelihood estimates, if any observation is equal to either of the $\hat{\mu}$'s, the likelihood function will go to infinity. However, this risk will become trivial if we keep enough digits during calculation.

For most data sets with distinct peaks, different initial values will converge to the same estimates. However, for some data sets, different initial values may give different estimates since there is more than one local maxima for the likelihood function. In this project, if more than one maximum is found, we will pick the one giving a larger likelihood function or smaller Watson's U^2.

The estimates of P and μ's are usually accurate, and the variances and the third moments of these estimates are usually small. However, the variances of κ's can be large sometimes, especially for the κ from the von Mises distribution with a small proportion.

For the SIDS data, we first fit a single von Mises distribution and test the fit using Watson's U^2. If the single von Mises is rejected at the 5% level, we try to fit a mixture of two von Mises distributions, and test the fit again. Mooney, Helms, and Jollife attempted to fit the double von Mises distribution because they thought a second pick may exist; however, they ran into numerical problems. The data for some of the years can be fitted using a single von Mises distribution, but the data for most years should be fitted using a mixture of two von Mises distributions. There are also a few years that cannot be fitted using either model; for example, the data for 1989. For the fatal crash data set, we selected three states to discuss: Washington, New York, and the District of Columbia. A mixture of two von Mises distributions can be fitted for the data of Washington State and the District of Columbia. For New York State, a single von Mises model is enough.

The purpose of this project was to show how a mixture can be fitted using continuous data. The motivation was provided by several papers analyzing SIDS data and a report on U.S. crash data.

Bibliography

[1] Cox, D.R. and Hinkley, D.V. (1974). "Theoretical Statistics". Chapman and Hall, London.

[2] D'Agostino, R.B. and Stephens, M.A. (1986). "Goodness-of-fit Techniques". Marcel Dekker, INC. New York.

[3] Fatality Analysis Reporting System Encyclopedia webpage. (March, 2009). http://www-fars.nhtsa.dot.gov/Main/index.aspx

[4] Fisher, N.I. (1993). "Statistical Analysis of Circular Data". Cambridge University Press. Cambridge.

[5] Hogg, R.V. and Craig, A.T. (1965). "Introduction to Mathematical Statistics". 2nd edn. Macmillan. New York.

[6] Jander, R. (1957) Die optische Richtangsorientierung der roten Waldameise (Formica rufa. L.). Z. vergl. Physiologie, 40: 162-238.

[7] Jones, T.A. and James, W.R. (1969). Analysis of bimodal orientation data. Math.Geol., 1: 129-135.

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

[9] Mardia, K.V. (1972). "Statistics of Directional Data". Academic Press, INC. London and New York.

[10] Mooney, J.A., Helms, P.J., and Jollife, I.T. (2002). Fitting Mixtures of von Mises Distributions: a case study involving sudden infant death syndrome. Computational Statistics & Data Analysis, 41: 505-513.

[11] Stephens, M.A. (1969). Techniques for directional data. Tech. Report No. 150. Dept. of Statist., Stanford Univ.

[12] Stephens, M.A. (1970). Use of the Kolmogorov-Smirnov, Cramer-von Mises and Related Statistics without Extensive Tables. J. Roy. Statist. Soc., B, 32: 115-122.

[13] Spurr, B.D. and Koutbeiy, M.A. (1991). A comparison of various methods for estimating the parameters in mixtures of von Mises distributions. Commun. Statist.-Simul. Comput., 20: 725-742.