

Advances on
Methodological and
Applied Aspects of
Probability and
Statistics

Edited by N. Balakrishnan



CHAPTER 8

UNIMODALITY IN CIRCULAR DATA: A BAYES TEST

SANJIB BASU

Northern Illinois University, DeKalb, IL

S. RAO JAMMALAMADAKA

University of California, Santa Barbara, CA

Abstract: Circular data which represent directions in two dimensions, may be measured as angles. Unimodality, which is often assumed, is a crucial issue since modeling and further inference depend on it. Just as on the real line, descriptive as well as inference tools are different for unimodal data as opposed to multi-modal data. We propose a Bayesian test for unimodality of circular data using mixtures of von-Mises distribution as the alternative. The proposed test is performed and evaluated using Markov Chain Monte-Carlo methodology.

Keywords and phrases: Directional data, von Mises distribution, mixture distribution, Bayes approach

8.1 INTRODUCTION

Suppose we have a set of independent and identically distributed measurements on 2-dimensional directions, say $\alpha_1, \alpha_2, \dots, \alpha_n$. These measurements, called angular or circular data, can be represented as points on the circumference of a circle with unit radius. They may be wind directions, the vanishing angles at the horizon for a group of birds or the times of arrival at a hospital emergency room where the 24 hour cycle is represented

as a circle. Such data may have one or more peaks or show no preferred direction at all, i.e., correspond to an isotropic or uniform distribution. Most circular statistical inference about preferred directions or modes starts after eliminating the last possibility namely that the data has no preferred direction i.e., that it is not uniformly distributed. The next step is to ask if there is just a single mode or if the data is multimodal, which is the subject of this paper.

As an example, consider a meteorologist studying wind directions. Based on past data, (s)he might be interested in knowing if the wind direction is predominantly in one direction or whether it is indeed different say at different times of the day or week. Similarly in calculating the directional spectrum of ocean waves, it is crucial to know whether we are dealing a unimodal or multimodal spectrum.

Circular data involves observations θ which are angles, i.e, $0 \leq \theta < 2\pi$. Such data are inherently periodic, i.e., $\theta \equiv (\theta + 2\pi k)$ for any integer k . This inherent periodicity sets apart circular statistical analysis, from the more common "linear" statistical analysis where one uses methods and models based on the mean, variance, etc. Such models and methods are not appropriate in circular statistics.

In standard(linear) statistics, a univariate density f is unimodal or has a single mode if f is non-decreasing up to a point M and non-increasing thereafter. In circular statistics, however, due to the circular nature and lack of well-defined left and right endpoints (such as $-\infty$ and ∞ in real line), the definition of unimodality also requires an antimode A . We will say that a circular probability density $p(\theta)$ is unimodal with mode at M if there exists an antimode A such that $p(\theta)$ is non-decreasing for $A \leq \theta \leq M$ and is non-increasing for $M \leq \theta \leq A$.

Knowledge of the number of modes of $p(\theta)$ is clearly of importance in circular statistics. For instance, a common example of circular data involves the vanishing directions of pigeons when they are released some distance away from their "home". The underlying scientific question relates to how these birds orient themselves. Are they flying towards their "home-direction"? Unimodality of the density $p(\theta)$ here would imply that pigeons have a preferred vanishing direction and is a hypothesis of considerable scientific interest.

As another example, several stations measure the mean wave direction every hour which corresponds to the dominant energy of the period. The wave directions depend on weather conditions, ocean currents and many other natural factors. The daily variation of the wave directions is an example of circular data on a 24-hour cycle. The hypothesis of unimodality here would imply that there is an overall preferred direction around which

the daily variations of the wave directions are distributed.

In linear statistics, the problem of estimating the number of modes and/or statistical tests for discovering the presence of more than one mode are considered by many authors. The earliest approach involve modeling multimodality through mixtures of distributions, see Wolfe (1970). Later works include several different approaches, density estimation and bump hunting [Good and Gaskins (1981), Silverman (1981)], distance of empirical distribution from the closest unimodal distribution [Hartigan and Hartigan (1985)] and the approach of excess mass functional [Müller and Swatzki (1991)]. Recently, Basu (1995) proposed a Bayesian test for unimodality using the Khintchine representation which states that every unimodal distribution on the real line can be represented as a mixture of uniform distributions.

We address a similar problem here but in the context of circular data. Let $\theta_1, \dots, \theta_n$ be i.i.d. observations from the circular density $p(\theta)$. We want to test $H_0 : p(\theta)$ is a unimodal against the alternative that it is not. We propose a Bayesian test which incorporates observed data and prior information. In this test, we restrict ourselves to the class of models whose density $p(\theta)$ can be represented parametrically as a mixture of two von-Mises distributions. After observing the data $\theta_1, \dots, \theta_n$, the joint prior distribution of the mixing proportion and the location and scale parameters of the two components are updated to their joint posterior distribution. The posterior probability of $p(\theta)$ being unimodal is then compared to the prior probability of unimodality to make a decision between H_0 and H_1 . These probabilities are computed by Markov Chain Monte Carlo sampling. In fact, one of the major strength of the proposed method is the simplicity of the computations involved; they are mostly direct simulations from popular densities which can be routinely implemented.

8.2 EXISTING LITERATURE

Many excellent books discuss statistical analysis of circular data, including Mardia (1972), Batschelet (1981) and Fisher (1993). We refer the reader to these books and the references therein. However, there does not seem to be any work on tests for unimodality of circular data.

The circular data literature, related to this article can be broadly divided into three groups. The first group involves tests for randomness against a unimodal alternative. Here the null hypothesis is isotropy, modeled by the uniform distribution on the circle. A common test for this against the von-Mises distribution, is known as the Rayleigh test. This test based on the length of the sample resultant, is known to be uniformly most powerful

invariant test; see, for instance Mardia (1972, pp. 180–182). For other tests which are more nonparametric, see Section 3.1 of Rao (Jammalamadaka) (1984).

Another set of references which could be related to the question we are studying comes from the density estimation point of view. Semiparametric and nonparametric density estimation for circular data are studied by many, for example, see Bai *et al.* (1988). From a density estimate one can determine the number of modes. However, tests of hypotheses are harder to come by since this involves the much harder problem of density estimation.

Other related work is on mixture distributions and estimating the number of mixing components, etc; see Mardia (1972), Bartel (1984). Spurr and Koutbeiy (1991) proposed a stepwise procedure for testing for the number of components in a von-Mises mixture, by first testing for one component against more than one, then two components against more than two and so on. This is a Bootstrap based test and can be computationally intensive. We point out here that a two or more component mixture can still be unimodal and hence the problem we are addressing is clearly distinct from these articles.

We also mention here that Mardia and Spurr (1973) developed a multi-sample test for data drawn from a L -modal population which they model by a mixture of a scaled von-Mises distribution on $[0, 2\pi/L)$, another scaled von-Mises distribution on $[2\pi/L, 4\pi/L)$ and so on. Our approach is also quite distinct from this work.

8.3 MIXTURE OF TWO VON-MISES DISTRIBUTIONS

We are given n i.i.d. circular observations $\theta_1, \dots, \theta_n$ ($0 \leq \theta < 2\pi$) from an unknown circular density $p(\theta)$ and want to test $H_0 : p(\theta)$ is unimodal against $H_1 : p(\theta)$ is not unimodal. We model $p(\theta)$ parametrically as a mixture of two von-Mises distribution,

$$p(\theta) = \pi vm(\theta|\mu_1, \kappa_1) + (1 - \pi) vm(\theta|\mu_2, \kappa_2) \quad (8.3.1)$$

where $vm(\theta|\mu, \kappa) = \exp(\kappa \cos(\theta - \mu)) / \{2\pi I_0(\kappa)\}$, $0 \leq \theta < 2\pi$ denotes the density of a von-Mises distribution. Here $I_0(\kappa)$ is the modified Bessel function of order 0.

The von-Mises distribution $vm(\theta|\mu, \kappa)$ plays a central role in circular statistics, quite similar to that of the normal distribution in linear statistics. The parameter μ ($0 \leq \mu < 2\pi$) is called the mean direction. The von-Mises distribution is symmetric and unimodal about μ . The parameter $\kappa > 0$ is the concentration parameter (similar to a precision parameter) and measures the concentration of mass around μ . Note that while the

mean direction parameter μ is angular (i.e., $0 \leq \mu < 2\pi$), the concentration κ is a positive real parameter, a fact useful our posterior simulations. As $\kappa \rightarrow 0$, the von-Mises distribution converges to the uniform distribution on the circle whereas as $\kappa \rightarrow \infty$, the von-Mises distribution converges to a degenerate distribution at μ . The popularity of von-Mises distribution in circular statistics stems from the fact that closed form results are often available for the sampling distributions of statistics from this model which are almost impossible for most other circular distributions. We refer the reader to books by Mardia (1972), Fisher (1993) for further properties of this distribution.

There are several advantages to modeling $p(\theta)$ parametrically as von-Mises mixture. A 2-component von-Mises mixture allows a wide variety shapes (based on various choices of the parameters $\mu_1, \mu_2, \kappa_1, \kappa_2$ and π) which includes symmetric and asymmetric, as well as both unimodal and bimodal densities. In Figure 8.1 we show three such mixtures to illustrate the different shapes and modality choices that are possible. Secondly, the conjugate prior for the mean direction of a von-Mises distribution is known and is another von-Mises distribution. This structure provides a flexible and at the same time, a mathematically convenient prior structure. Thirdly, if $p(\theta)$ is a 2-component von-Mises mixture, then a complete mathematical characterization is available about when $p(\theta)$ is unimodal and when it is not. This characterization, due to Mardia and Sutton (1975), is described next.

Let $p(\theta) = \pi vm(\theta|\mu_1, \kappa_1) + (1-\pi) vm(\theta|\mu_2, \kappa_2)$. By appropriate choice of the zero direction, one can assume that $\mu_1 = 0$ and $0 \leq \mu_2 \leq \pi$. The characterization is stated in this parametrization.

Case 1. This is a boundary case when $\mu_2 = \pi$, i.e., the two means are at the opposite ends of the circle. Then the density $p(\theta)$ is bimodal if and only if the mixing proportion π satisfies $p_1 \leq \pi \leq p_2$ where $p_1 = \{1 + \kappa^* \exp(\kappa_1 + \kappa_2)\}^{-1}$, $p_2 = \{1 + \kappa^* \exp(-\kappa_1 - \kappa_2)\}^{-1}$, and $\kappa^* = \{\kappa_1 I_0(\kappa_2)\} / \{\kappa_2 I_0(\kappa_1)\}$. In this case, the two modes are at π and 2π .

Case 2. This is the important case when $0 < \mu_2 < \pi$. Then the density $p(\theta)$ is bimodal if and only if $p'_1 \leq \pi \leq p'_2$ and $\sin \mu_2 \leq h(\hat{\theta})$. Here $p'_j = \{1 - \kappa^* / u(\theta_j)\}^{-1}$, $j = 1, 2$; κ^* is as defined above, $u(\theta) = \{\sin(\theta - \mu_2) / \sin \theta\} \exp(\kappa_2 \cos(\theta - \mu_2) - \kappa_1 \cos \theta)$ and $0 < \theta_1 < \theta_2 < \mu_2$ are the two solutions of the equation $h(\theta) = \sin \mu_2$. Finally, $h(\theta) = \sin \theta \sin(\theta - \mu_2) \{\kappa_2 \sin(\theta - \mu_2) - \kappa_1 \sin \theta\}$ and $\hat{\theta}$ maximizes $h(\theta)$ within $0 < \theta < \mu_2$ which further can be obtained as the real root of a cubic equation.

Mardia and Sutton (1975) also provide information on the location of the mode(s) and antimode(s) of the mixture density $p(\theta)$ which are omitted here as they are not of primary importance in our context.

8.4 PRIOR SPECIFICATION

We next specify the prior models for the parameters of the mixture density $p(\theta)$. Note that this density has five parameters, the mixing proportion π , the two mean directions μ_1 and μ_2 and the two concentrations κ_1 and κ_2 . For the mixing proportion π , we assume a Uniform[0, 1] prior which reflects our prior uncertainty about its value.

We next describe the prior distributions for the mean directions μ_i , $i = 1, 2$. Note that the von-Mises distribution can alternatively be written as $vm(\theta|\mu, \kappa) = \{2\pi I_0(\kappa)\}^{-1} \exp(\kappa \eta^T \tilde{l})$ where $\eta^T = (\eta_1, \eta_2) = (\cos \mu, \sin \mu)$ and $\tilde{l}^T = (l_1, l_2) = (\cos \theta, \sin \theta)$. This alternative representation shows the exponential family structure of the von-Mises distribution. Using this structure, Mardia and El-Atoum (1976) showed that a conjugate prior for μ in the $vm(\theta|\mu, \kappa)$ sampling density is another von-Mises distribution, say $\pi(\mu) = vm(\nu, \tau)$.

We use this convenient conjugate structure in our formulation and assume that the two mean directions, μ_j , $j = 1, 2$, have two independent von-Mises prior, $vm(\nu_j, \tau_j)$, $j = 1, 2$. Within the von-Mises parametric structure, the choice of the hyperparameters ν_j, τ_j actually provides considerable flexibility in modeling different prior opinion. Note that as τ_j tends to zero, the $vm(\nu_j, \tau_j)$ prior tends to the uniform distribution on the circle. Thus, one can specify small values for the hyperparameter τ_j to reflect prior ignorance about μ_j .

Finally we specify the priors for the two concentration parameters κ_1 and κ_2 . The concentration parameter κ of a von-Mises distribution does not have a conjugate prior (due to the presence of the modified Bessel function $I_0(\kappa)$ term). However, as we noted before, κ is not restricted to a circular domain and can take any positive real value. A popular prior choice for precision parameter is a Gamma prior. We assume that the two concentration parameters κ_j , $j = 1, 2$ have two independent Gamma(α_j, β_j), $j = 1, 2$ priors. In fact, many other prior choices for κ_j are possible. Basu and Jammalamadaka (1999) describe a broad class of prior choices for κ_j and describe how the unimodality test can be carried out for any prior selection from this broad class.

8.5 PRIOR AND POSTERIOR PROBABILITY OF UNIMODALITY

In the following, we first repeat the complete model structure.

- We observe circular observations $\theta_1, \dots, \theta_n$ i.i.d. from the density $p(\theta) = \pi vm(\theta|\mu_1, \kappa_1) + (1 - \pi) vm(\theta|\mu_2, \kappa_2)$. This results in the likelihood

$$\begin{aligned} L(\pi, \mu_1, \mu_2, \kappa_1, \kappa_2) \\ = \prod_{i=1}^n p(\theta_i) = \prod_{i=1}^n \{ \pi vm(\theta_i|\mu_1, \kappa_1) + (1 - \pi) vm(\theta_i|\mu_2, \kappa_2) \} \end{aligned} \quad (8.5.2)$$

- The mixing proportion π has a Uniform[0, 1] prior distribution.
- The prior for μ_j is $p(\mu_j) = vm(\mu_j, \tau_j)$, $j = 1, 2$.
- The prior for κ_j is $p(\kappa_j) = \text{Gamma}(\alpha_j, \beta_j)$, $j = 1, 2$.

Our proposed Bayesian test of unimodality is performed by comparing the prior probability from this model with the posterior probability of unimodality. The computation for these probabilities are described next.

The prior probability of unimodality is the integral of the joint prior density of $(\pi, \mu_1, \mu_2, \kappa_1, \kappa_2)$ over the region of the joint parameter space (as described in the Mardia and Sutton (1975) result of Section 8.3) on which the mixture density $p(\theta)$ is unimodal. While the joint prior density $p(\pi, \mu_1, \mu_2, \kappa_1, \kappa_2)$ can be easily written down, the form of the unimodality region in the five-dimensional space of $(\pi, \mu_1, \mu_2, \kappa_1, \kappa_2)$ described in the Mardia and Sutton (1975) result is highly complicated and hence the resulting integral is analytically intractable. We instead obtain a Monte Carlo estimate of the prior probability of unimodality as follows. (i) Let $\phi = (\pi, \mu_1, \mu_2, \kappa_1, \kappa_2)$. We generate i.i.d. samples $\{\phi^{(t)} : t = 1, \dots, T_1\}$ from the joint prior distribution of ϕ . (ii) For each generated $\phi^{(t)}$, we examine if the resulting mixture density $p(\theta)$ is unimodal by checking the Mardia-Sutton condition. (iii) Finally, we obtain a simulation-consistent estimator of the prior probability of unimodality as $\{\text{Number of generated } \phi^{(t)} \text{ for which the resulting } p(\theta) \text{ is unimodal}\} / T_1$. Due to the independence structure, simulation from the joint prior can be done component-wise, which only involves random variate generation from some common densities. Further, checking the Mardia-Sutton condition for a given value of ϕ is also relatively straightforward.

The other probability required for the assessment of modality is the joint posterior probability of the parameter region on which the mixture density $p(\theta)$ is unimodal. We plan to estimate this probability also as a Monte Carlo average, i.e., once we have samples $\{\phi^{(t)} : t = 1, \dots, T_2\}$ from the joint posterior distribution, we can simply estimate the posterior probability of unimodality by following the method outlined in the previous paragraph.

The joint posterior distribution $p(\pi, \mu_1, \mu_2, \kappa_1, \kappa_2 | \text{data})$ is, however, analytically intractable and hence direct generation $\phi = (\pi, \mu_1, \mu_2, \kappa_1, \kappa_2)$ is very hard. We, instead, take recourse to Gibbs Sampling. We refer the reader to Gelfand and Smith (1990), Casella and George (1992) and the collection of papers by Gilks *et al.* (1995) for the theory, implementation and convergence issues of the Gibbs sampler. The main idea of the Gibbs sampler is to simulate alternately and iteratively for the conditional posterior distributions of each unobservable given the data and other observables.

The details of the Gibbs sampler for our model including the form of the full conditional distributions and how to simulate from this conditional distributions is described in Basu and Jammalamadaka (1999). We note here that latent variables I_1, \dots, I_n are introduced in the implementation of the Gibbs sampler. I_i is an indicator variable denoting the component from which θ_i is coming, i.e., $\theta_i | I_i = 1 \sim vm(\mu_1, \kappa_1)$, $\theta_i | I_i = 2 \sim vm(\mu_2, \kappa_2)$, $i = 1, \dots, n$ and I_1, \dots, I_n are i.i.d. with $P(I_i = 1) = \pi$, $P(I_i = 2) = 1 - \pi$ a priori. For further details of the Gibbs sampler, the reader is referred to Basu and Jammalamadaka (1999).

8.6 THE BAYES FACTOR

Standard Bayesian solution to a hypothesis testing problem involves formulating parametric models for null (H_0) and alternative (H_1) hypotheses and subsequently choosing one over the other in the light of the data and prior opinion. Perhaps the most widely used selection criterion used in this context is the 'Bayes factor of H_1 against H_0 ' formally defined as the ratio $B_{10} = \frac{\text{posterior odds}}{\text{prior odds}} = \frac{P(H_1 | \text{data})P(H_0)}{P(H_0 | \text{data})P(H_1)}$. The Bayes factor is used as a summary of evidence of H_1 against H_0 provided by the data. Thus, operationally, Bayes factor has the same role as that of a P -value in classical hypothesis testing scenario (see the review article by Kass and Raftery (1995) for an illuminating discussion comparing Bayes factor and P -values). From another perspective, the Bayes factor is similar to the likelihood ratio statistics as the former is the ratio of the marginal likelihood under H_1 against that over H_0 . The following Table 8.1 provided by Kass and

Raftery (1995) gives a rounded scale for interpretation of B_{10} and $\log B_{10}$.

TABLE 8.1 Evidence in support of alternative model from Bayes factor

B_{10}	$\log B_{10}$	evidence for M_1
< 1	< 0	negative (supports M_0)
1 to 3	0 to 1	barely worth mentioning
3 to 12	1 to 2.5	positive
12 to 150	2.5 to 5	strong
> 150	> 5	very strong

For our test of $H_0 : p(\theta)$ is unimodal against $H_1 : p(\theta)$ is non-unimodal, we report the estimated Bayes factor B_{10} . This is easily obtained once we estimate $P(H_0)$ = prior probability of unimodality and $P(H_0 | \text{data})$ = posterior probability of unimodality by the Monte Carlo methods mentioned in section 8.5.

8.7 APPLICATION

We consider data collected by Schmidt-Koenig (1963) in an experiment to determine how do birds determine directions and orient themselves. In this experiment, 15 homing pigeons were released about 16.25 kilometers northwest from their loft. The measurements listed in Table 8.2 are their vanishing directions measured in degrees. The direction of the loft is 149° .

TABLE 8.2 Vanishing direction of 15 homing pigeons. The loft direction is 149°

85	135	135	140	145	150	150	150	160	285	200	210	220	225	270
----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

These data have been analyzed by several authors, including Mardia (1972) and Fisher (1993). As can be seen in Table 8.2, most of the observations are concentrated around south (180°) with two observations in the east and west direction. Fisher (1993) reports that a goodness-of-fit test reveals there is some evidence a von-Mises distribution may not be a totally adequate description of the data.

We apply our proposed Bayesian modality test to these circular data. The following priors are used: (i) $\pi \sim \text{Uniform}(0, 1)$, (ii) $\mu_1 \sim \text{vm}(0^\circ, 0.25)$ and $\mu_2 \sim \text{vm}(180^\circ, 0.25)$, and (iii) $\kappa_j \sim \text{Gamma}(1, 5)$, $j = 1, 2$ where $\text{Gamma}(\alpha, \beta)$ has density proportional to $x^{\alpha-1} \exp(-x/\beta)$. These are moderately flat priors and the mean directions for μ_1 and μ_2 are chosen in opposite directions.

The prior probability of unimodality is estimated based on 100,000 samples generated from the above prior model and then following the method outlined in Section 8.5. We estimate the prior probability of unimodality to be 0.48867.

The posterior probability is estimated from 50,000 MCMC samples generated from the posterior after an initial burn-in of 10,000 and then once again following the method of section 8.5. Figure 8.2 shows the kernel density estimates of the posterior density for the two mean directions μ_1, μ_2 , the two concentration parameters κ_1, κ_2 and the mixing proportion π . These density estimates are obtained using the CODA software [Best *et al.* (1996)]. The posterior summary estimates (posterior mean, standard deviation and percentiles) of these parameters are shown in Table 8.3. We check convergence of the MCMC sampler using different convergence and stationarity checks available in CODA. The autocorrelation plots at different lags based on the simulated samples of $\kappa_1, \kappa_2, \mu_1, \mu_2$ and π are shown in Figure 8.3. High autocorrelations typically imply slow mixing and slow convergence. In Figure 8.3, the autocorrelations for κ_1 and κ_2 die out quickly. The autocorrelations for μ_1 and μ_2 do not die so quickly whereas π has significant autocorrelations till lag 20.

TABLE 8.3 Estimated posterior mean, standard deviation and percentiles of $\mu_1, \mu_2, \kappa_1, \kappa_2$ and π

	Posterior Mean	Posterior Std. Dev.	Percentiles (2.5%, 25%, 50%, 75%, 97.5%)
μ_1	185	56.4	(61.8, 152, 180, 220, 311)
μ_2	184	51.1	(81.2, 152, 179, 216, 296)
κ_1	3.18	3.24	(0.22, 1.39, 2.22, 3.75, 12.2)
κ_2	3.21	3.20	(0.28, 1.40, 2.24, 3.81, 12.1)
π	0.50	0.28	(0.03, 0.27, 0.50, 0.73, 0.98)

The posterior probability of unimodality from the generated MCMC samples is estimated to be 0.68382. Based on these prior and posterior probabilities of unimodality, the Bayes factor for non-unimodality against unimodality is estimated as $B_{10} = 0.44188$. Thus, the data do not provide almost any evidence against the null hypothesis of unimodality. This is also evident from the posterior density estimates in Figure 8.2. The posterior density estimates of μ_1, μ_2 and κ_1, κ_2 are almost identical, which probably indicate that the two components of the mixture density are close to identical or that the mixture density is just a single von-Mises distribution. If this is true, then the mixing proportion π becomes redundant which could explain the large spread in its posterior density estimate and its autocorrelations staying on for up to lag 20.

We further compute the predictive density of a new circular observation $p(\theta | \text{data})$ conditioned on the 15 vanishing directions already observed. This is obtained as

$$p(\theta | \text{data}) = \int p(\theta | \pi, \mu_1, \mu_2, \kappa_1, \kappa_2) d\pi(\pi, \mu_1, \mu_2, \kappa_1, \kappa_2 | \text{data})$$

where $\pi(\pi, \mu_1, \mu_2, \kappa_1, \kappa_2 | \text{data})$ is the posterior distribution of the parameters. This predictive density is estimated on a grid of θ values where the integral above is estimated by the Monte Carlo average of the generated MCMC samples. This estimated predictive density in Figure 8.4 exhibits a clear unimodal structure and provide further evidence to our test result.

8.8 SOME ISSUES

A. Identifiability: In mixture modeling, identifiability of parameters is typically of concern. To see how identifiability issues may arise in our two-component von-Mises mixture model, consider the likelihood function defined in (8.5.2). It is easy to see from (8.5.2) that $L(\pi, \mu_1, \mu_2, \kappa_1, \kappa_2) = L(1 - \pi, \mu_2, \mu_1, \kappa_2, \kappa_1)$, i.e., $(\pi, \mu_1, \mu_2, \kappa_1, \kappa_2)$ and $(1 - \pi, \mu_2, \mu_1, \kappa_2, \kappa_1)$ provide identical likelihood. In Bayesian analysis, non-identifiability is often avoided by bringing in separation of parameter values in the prior modeling. However, if (μ_1, μ_2) and (κ_1, κ_2) have exchangeable priors and if the prior for π is symmetric around $1/2$, then the prior and hence the posterior also fails to identify between $(\pi, \mu_1, \mu_2, \kappa_1, \kappa_2)$ and $(1 - \pi, \mu_2, \mu_1, \kappa_2, \kappa_1)$. While non-identifiability is not a formal problem in Bayesian inference, it may lead to very slow convergence of the MCMC sampler. The resulting inference could also be troublesome. for example, the posterior distribution of μ_1 may appear to be bimodal due to concentration of mass around the mean directions of both components.

One way to ensure identifiability is to put some prior constraints on the parameter space. For example, a common constraint put in two component mixture is $\mu_1 \leq \mu_2$. In Bayesian analysis, this constraint can be brought in very naturally by simply defining the prior support to be the constrained space. This constraint makes all the parameters identifiable. Bayesian analysis under this constraint can be carried out in a straightforward manner, however it does bring in complications within the MCMC sampler. The full conditional distributions of both μ_1 and μ_2 are now constrained by the other parameter. Robert (1996) discuss the issue of parameterizations and constraints in the context of normal mixture models and suggests the reparametrization $\mu = \mu_1$ and $\lambda = \mu_2 - \mu_1$ where λ is assumed to non-negative a priori. This reparametrization generally achieves more stability

within the MCMC sampler and faster convergence.

B. Choice of two component mixture. We model the sampling distribution as a two component mixture of von-Mises distribution. This model allows substantial flexibility as the mixing proportion π , the two mean directions μ_1, μ_2 and the two concentration parameters κ_1, κ_2 are allowed to vary thus resulting in different shapes and scales of the mixture density. However, a two-component mixture can at most produce a bimodal density. Thus, if data generated from a tri-modal or multi-modal distribution is fed into our model, it is not obvious how our proposed test will behave. Secondly, the unimodal or non-unimodal densities that can be obtained within our model are only those which can be characterized as two-component mixtures of von-Mises distributions. We thus do not have extensive flexibility on the functional form of the density.

The problem of more than two modes can be addressed by considering a k -component mixture density model: $p(\theta) = \sum_{j=1}^k \pi_j vm(\theta|\mu_j, \kappa_j)$ where

$\sum_{j=1}^k \pi_j = 1$. The analysis for such a model can be performed in an analogous manner with some minor modifications in the full conditional distributions of the parameters. However, the identifiability issues discussed above becomes more severe and convergence issues in the MCMC sampler becomes more critical. Another problem is how to determine the value of k . One can put a hierarchical structure to the problem by assuming a prior distribution supported on positive integer values for k . This, however, makes the problem very hard as it now becomes a variable dimension problem and one may need to use the reversible jump algorithm to move from one dimension to another within the MCMC sampler. Green and Richardson (1997) recently addressed this variable dimension problem in the context of normal mixtures.

Mixtures of more than two components allows somewhat more flexibility in the functional form of the mixture density. Further flexibility can be obtained by semiparametric modeling. For example, in real line, all univariate unimodal distributions can be characterized as mixtures of uniform distributions (known as the Khintchine representation). This mixture representation is often used in modeling univariate unimodal distribution, one then assumes a prior on the mixing distribution of the uniforms. A similar representation also exists for unimodal distributions on the circle, they can also be written as mixtures of uniform distributions [see Fang *et al.* (1989)]. We are currently developing a semiparametric test for unimodality of circular data based on this representation.

REFERENCES

- Bai, Z. D., Rao, C. R. and Zhao, L. C. (1988). Kernel estimators of density function of directional data, *Journal of Multivariate Analysis*, **27**, 24-39.
- Bartels, R. (1984). Estimation in a bidirectional mixture of von-Mises distributions, *Biometrics*, **40**, 777-784.
- Basu, S. (1995). Bayesian tests for unimodality, *Technical Report*.
- Basu, S. and Jammalamadaka, S. Rao (1999). Parametric and semiparametric Bayesian tests for unimodality in directional data, Manuscript under preparation.
- Batschelet, E. (1981). *Circular Statistics in Biology*, Academic Press, London.
- Best, N., Cowles, M. K. and Vines, K. (1996). *CODA: Convergence Diagnosis and Output Analysis Software for Gibbs Sampling Output*, Version 0.30.
- Fang, K-T., Kotz, S. and Ng, K-W. (1989). *Symmetric Multivariate and Related Distributions*, Chapman and Hall, London.
- Fisher, N. I. (1993). *Statistical Analysis of Circular Data*, Cambridge University Press, Cambridge.
- Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (Eds) (1995). *Markov Chain Monte Carlo in Practice*, Chapman and Hall, New York.
- Good, I. J. and Gaskins, R. A. (1980). Density estimation and bump-hunting by the penalized likelihood method exemplified by scattering and meteorite data, *Journal of the American Statistical Association*, **75**, 42-73.
- Green, P. and Richardson, S. (1997). On Bayesian analysis of mixtures with an unknown number of components, *Journal of the Royal Statistical Society, Series B*, **59**, 731-792.
- Hartigan, J. A. and Hartigan, P. M. (1985). The dip test for unimodality, *Annals of Statistics*, **13**, 70-84.
- Kass, R. E. and Raftery, A. E. (1995). Bayes factors, *Journal of the American Statistical Association*, **90**, 773-795.

- Mardia, K. V. (1972). *Statistical Analysis of Directional Data*, Academic Press, London.
- Mardia, K. V. and El-Atoum, S. A. M. (1976). Bayesian inference for the von Mises-Fisher distribution, *Biometrika*, **63**, 203–205.
- Mardia, K. V. and Spurr, B. D. (1973) Multisample tests for multimodal and axial circular populations, *Journal of the Royal Statistical Society, Series B*, **35**, 422–436.
- Mardia, K. V. and Sutton, T. W. (1975). On the modes of mixture of two von-Mises distributions, *Biometrika*, **62**, 699–701.
- Müller, D. W. and Sawitzki, G. (1991). Excess mass estimates and test for multimodality, *Journal of the American Statistical Association*, **86**, 738–746.
- Rao, J. S. (1984). Nonparametric methods in directional data, In *Handbook of Statistics, Volume 4* (Eds., P. R. Krishnaiah and P. K. Sen), pp. 755–770, Elsevier Science Publishers, Amsterdam.
- Robert, C. P. (1996). Mixtures of distributions: inference and estimation, In *Markov Chain Monte Carlo in practice* (Eds., W. Gilks, S. Richardson and D. Spiegelhalter), pp. 441–461, Chapman and Hall, London.
- Silverman, B. W. (1981). Using kernel density estimates to investigate multimodality, *Journal of the Royal Statistical Society, Series B*, **43**, 97–99.
- Spurr, B. D. and Koutbeiy, M. A. (1991). A comparison of various methods for estimating the parameters in mixtures of von-Mises distributions, *Communications in Statistics—Simulation and Computation*, **20**, 725–741.
- Wolfe, J. H. (1970). Pattern clustering by multivariate mixture analysis, *Multivariate Behavioural Research*, **5**, 329–350.

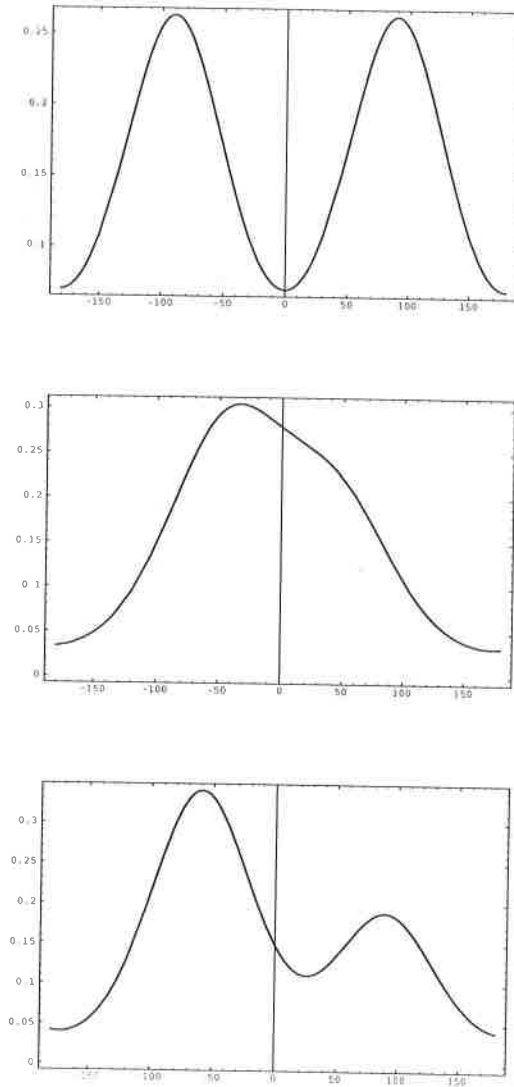


FIGURE 8.1 Three von-Mises mixtures: Top $= 0.5 \text{ vm}(\theta | -90^\circ, 2) + 0.5 \text{ vm}(\theta | 90^\circ, 2)$, Middle $= 0.6 \text{ vm}(\theta | -45^\circ, 1.5) + 0.4 \text{ vm}(\theta | 45^\circ, 1.5)$, Bottom $= 0.65 \text{ vm}(\theta | -60^\circ, 2) + 0.35 \text{ vm}(\theta | 90^\circ, 2)$

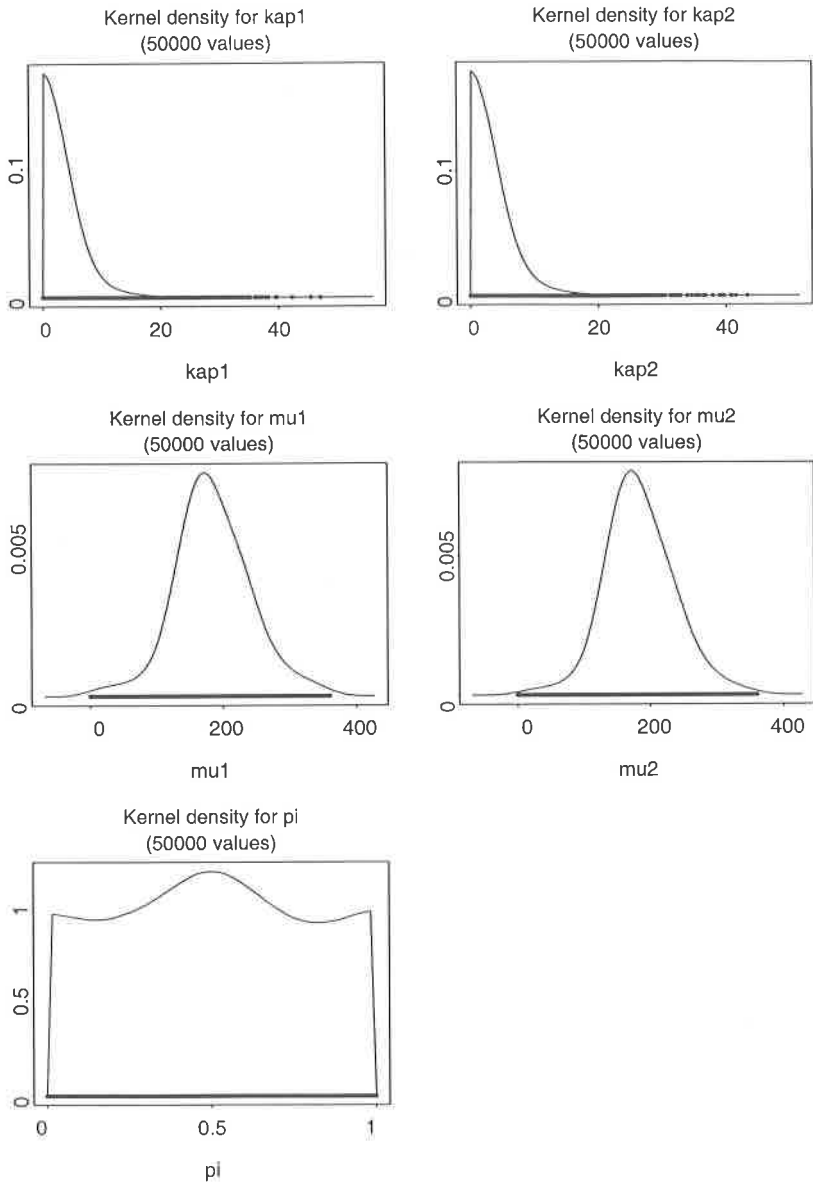


FIGURE 8.2 Kernel estimates of the posterior density of the parameters $\kappa_1, \kappa_2, \mu_1, \mu_2, \pi$

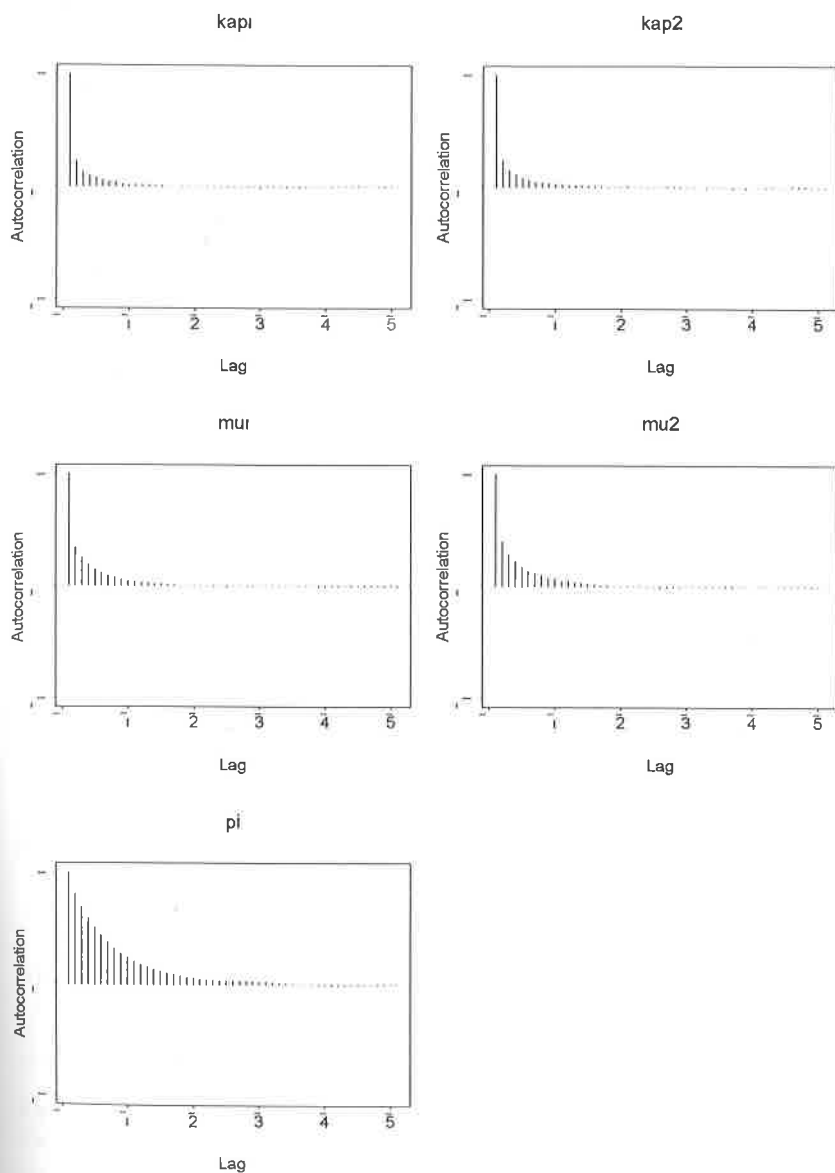


FIGURE 8.3 Autocorrelation plot at different lags for the five parameters: $\kappa_1, \kappa_2, \mu_1, \mu_2, \pi$

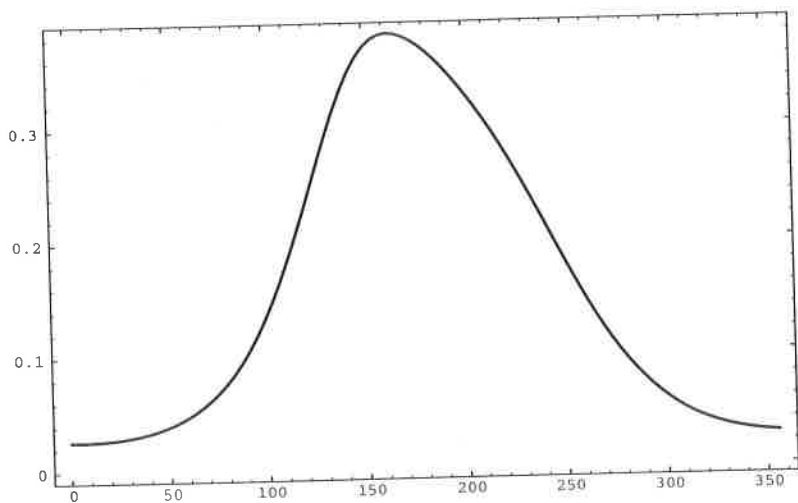


FIGURE 8.4 Predictive density of a new circular observation for the pigeon data