

# Semiparametric Bayesian Techniques for Problems in Circular Data

**KAUSHIK GHOSH<sup>1</sup>, S. RAO JAMMALAMADAKA<sup>2</sup> & RAM C. TIWARI<sup>3</sup>,**

<sup>1</sup>*Department of Statistics, George Washington University, USA;* <sup>2</sup>*Department of Statistics and Applied Probability, University of California, Santa Barbara, USA;*

<sup>3</sup>*Department of Mathematics, University of North Carolina, Charlotte, USA*

**ABSTRACT** *In this paper, we consider the problems of prediction and tests of hypotheses for directional data in a semiparametric Bayesian set-up. Observations are assumed to be independently drawn from the von Mises distribution and uncertainty in the location parameter is modelled by a Dirichlet process. For the prediction problem, we present a method to obtain the predictive density of a future observation, and, for the testing problem, we present a method of computing the Bayes factor by obtaining the posterior probabilities of the hypotheses under consideration. The semiparametric model is seen to be flexible and robust against prior misspecifications. While analytical expressions are intractable, the methods are easily implemented using the Gibbs sampler. We illustrate the methods with data from two real-life examples.*

## 1 Introduction

In many natural and physical sciences, the observed data are directions. Such is the case when a meteorologist is recording hourly wind directions, a geologist is studying the direction of sediment deposit from an ancient river or an environmental scientist is studying the direction of pollutant flow from a smokestack. In addition to physical directions, scientists are often interested in data that are periodic with a known period. For example, the variable of interest may be the time of day when an emergency room arrival occurs, or the time of day when a person has the lowest melatonin in his/her body.

In the above examples, the data can be represented as points on the circumference of a unit circle, with one complete cycle ( $= 24$  hours when the variable of interest

*Correspondence:* Kaushik Ghosh, Department of Statistics, George Washington University, Washington, DC 20052, USA. E-mail: ghosh@gwu.edu

is time of day) equivalent to one rotation around the circle. Data represented as points on a circle are sometimes collectively known as circular data. In this article, we deal only with circular data but the methods that we present here can be easily generalized to higher dimensions.

Circular observations can be represented as unit vectors  $(1, x)$  in the polar coordinates, or, alternatively as  $x$ , where  $x$  ( $0 \leq x < 2\pi$ ) is the angle of rotation from a conveniently chosen zero direction (origin) and a sense of positive rotation. However, the choice of origin and sense of positive rotation are arbitrary, making possible multiple representations of the same observation. As an example, mathematicians think of 'East' as 0 and anticlockwise rotation as positive, while geologists follow the convention of 'North' as 0 and clockwise rotation as positive. Thus, an observation of  $\pi$  on the mathematicians' scale would be  $3\pi/2$  on the geologists' scale.

Another problem posed by circular data is the periodicity, and, as a result, large numerical differences need not signify large physical differences. For example, two emergency room arrivals at 23:45 and 00:30 are closer to each other than, say, at 1:00 and 3:00, even though the numerical difference is smaller in the second case.

All these problems render usual statistical techniques useless for circular data, since they do not take into account the latter's special properties. If  $x_1, \dots, x_n$  are  $n$  observed directions, the circular analogue of mean and variance of the observations are given by  $\bar{x}_0$  and  $1 - (R/n)$  respectively, where

$$\left. \begin{aligned} R \cos \bar{x}_0 &= \sum_{i=1}^n \cos x_i \\ R \sin \bar{x}_0 &= \sum_{i=1}^n \sin x_i \end{aligned} \right\} \quad (1)$$

and  $\bar{x}_0$  and  $R$  are the direction and length respectively of the resultant of the unit vectors formed by the  $n$  observed directions. Similar modifications are necessary for almost every 'linear' statistical technique we want to adapt to the circular set-up. For more on the subject of directional data, the reader is referred to Mardia & Jupp (2000), Fisher (1995) and Jammalamadaka & SenGupta (2001).

Often, in the circular set-up, one might be interested in predicting the future behaviour of a system based on past data. For example, a meteorologist who has data on average wind directions from the past month, wants to predict tomorrow's (or next week's) average wind direction. Another example is where hospital administrators have data on the arrival patterns at an emergency room and want to use them to predict future arrival patterns. This is particularly helpful in planning staff and equipment schedules, so that 'rush hours' are well staffed. A third example is in the study of various circadian rhythms in the human body and one might be interested, for example, in the time of peak blood glucose level (to devise an optimal treatment schedule by timing the insulin injection). In the first part of this article, we present a Bayesian approach to finding the predictive density of future observations using a class of Dirichlet process priors (see Ferguson, 1973). This approach allows for flexibility in the prior specification by quantifying belief in the chosen prior.

A different kind of problem in circular data arises when one is interested in testing for the equality of two (or more) preferred directions. For example, physicians might be interested in determining if the average arrival times at an intensive care unit (ICU) are the same for male and female patients. Another example is where geologists are comparing the average direction of flow of an

ancient river during different time periods by looking at the orientation of sedimentary deposits. This is the circular analogue of the usual two-sample problem in the linear case. The frequentist approach is to assume a parametric set-up and derive an optimal test. This, however, requires specification of the probability of Type I error, which is somewhat arbitrary. Moreover, the specified parametric model may not be entirely accurate and the frequentist method does not allow us to incorporate any fuzziness about the model. The alternative is to work in the parametric Bayesian set-up, but this also requires an exact knowledge about the prior distribution of the parameters.

In the second part of this article, we look at this problem from the semi-parametric Bayesian viewpoint, where the parameters are allowed to have a Dirichlet process prior. Like the prediction problem, this allows us flexibility in our model choice and gives a more robust analysis than the parametric Bayesian set-up. We present a technique that uses the data to update prior probabilities to get the posterior probabilities of the null and alternative hypotheses. Using these, we compute the Bayes factor and determine the hypothesis that is supported by the observed data.

A natural model for circular data is the von Mises distribution with the pdf of the observed angle  $x$  given by

$$f(x|\mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp[\kappa \cos(x - \mu)], \quad 0 \leq x < 2\pi, 0 \leq \mu < 2\pi, \kappa > 0$$

This is symmetric, unimodal and is a circular analogue of the normal distribution on the line—hence, is also known as the circular normal distribution. We will denote this distribution by  $VM(\mu, \kappa)$ . The parameter  $\mu$  is a location parameter denoting the mean direction,  $\kappa$  measures the concentration around this mean direction and  $I_0(\kappa)$  is the Bessel function of the first kind of order 0.  $\kappa = 0$  means that there is no preferred direction ( $\mu$  is undefined in this case) and that the observations are uniformly distributed on the unit circle.

For purposes of this article, we will assume that the observed directions are observations from von Mises distributions with unknown  $\mu$ s and  $\kappa$ s. Section 2 discusses the prediction problem in detail and illustrates the whole idea with an example from ICU data. Section 3 discusses the problem of testing of equality of two means, which is illustrated by an example on directions of palaeocurrents. Finally, Section 4 discusses the relative merits and demerits of the proposed procedures and compares them with the traditional methods. Possible future work is also proposed.

## 2 Prediction problem

### 2.1 Introduction and notations

Let  $x_1, \dots, x_n$  be  $n$  circular observations with  $\mu_i$  being the mean direction and  $\kappa_i$  the concentration parameter respectively corresponding to  $x_i$ . We assume that  $x_i | (\mu_i, \kappa_i), i = 1, \dots, n$  are independent random variables distributed as  $x_i | (\mu_i, \kappa_i) \sim VM(\mu_i, \kappa_i), i = 1, \dots, n$ . We also assume that the mean directions  $\mu_1, \dots, \mu_n$  are independent and identically distributed (i.i.d.) with an unknown distribution  $G$  and that the concentration parameters  $\kappa_i$  are i.i.d. from a completely known distribution  $H$  and that the  $\kappa_i$ s are independent of the  $\mu_i$ s. Being unsure about the form of  $G$ , we model  $G$  non-parametrically. There is a prior guess  $G_0$  (which is

completely known) of  $G$  available and the strength of our belief in this guess is quantified by  $M (> 0)$ , also known.  $G_0$  is also sometimes referred to as the ‘baseline prior’ distribution of the  $\mu_i$ s. The above information regarding  $G$  is combined by saying that  $G$  has a Dirichlet process prior (see Ferguson, 1973, p. 214) with parameters  $M$  and  $G_0$ —symbolically,  $G \sim \mathcal{D}(M, G_0)$ .

If  $G \sim \mathcal{D}(M, G_0)$ , we have the following: for any  $k \geq 1$  (integer) and  $B_1, \dots, B_k$  a partition of the sample space  $\Omega$ , the random vector

$$(G(B_1), \dots, G(B_k)) \sim \text{Dir}(MG_0(B_1), \dots, MG_0(B_k))$$

where  $\text{Dir}(\cdot)$  stands for the Dirichlet distribution (see Wilks, 1962, Section 7.7). The Dirichlet process prior  $\mathcal{D}(M, G_0)$  thus gives us a probability measure on the space of all distributions  $G$  that has a location parameter  $G_0$  and dispersion inversely proportional to  $M$ . A high value of  $M$  implies a very strong prior belief that  $G$  is actually  $G_0$ .

Let

$$\mathbf{x}_n = (x_1, \dots, x_n)$$

$$\boldsymbol{\mu}_n = (\mu_1, \dots, \mu_n),$$

$$\boldsymbol{\kappa}_n = (\kappa_1, \dots, \kappa_n)$$

and

$$\boldsymbol{\mu}_n^{(i)} = (\mu_1, \dots, \mu_{i-1}, \mu_{i+1}, \dots, \mu_n)$$

$$\boldsymbol{\kappa}_n^{(i)} = (\kappa_1, \dots, \kappa_{i-1}, \kappa_{i+1}, \dots, \kappa_n)$$

be the parameter vectors obtained by deleting the  $i$ th elements of  $\boldsymbol{\mu}_n$  and  $\boldsymbol{\kappa}_n$  respectively. We will use the notation  $[X|Y]$  to indicate the conditional distribution of  $X$  given  $Y$  (here  $X, Y$  may be vectors) and  $\delta_x(\cdot)$  to denote the degenerate measure that puts a mass of 1 at  $x$ . Finally, where convenient, we will use the generic notation  $dp(\cdot)$  to denote the joint (conditional or unconditional) distribution of the parameters.

## 2.2 Theory

The likelihood function of  $(\boldsymbol{\mu}_n, \boldsymbol{\kappa}_n)$  based on the observed data  $\mathbf{x}_n$  is given by

$$f(\mathbf{x}_n | \boldsymbol{\mu}_n, \boldsymbol{\kappa}_n) = \prod_{i=1}^n f(x_i | \mu_i, \kappa_i)$$

From the Bayes theorem, the posterior of  $(\boldsymbol{\mu}_n, \boldsymbol{\kappa}_n)$  given  $\mathbf{x}_n$  is

$$\begin{aligned} dp(\boldsymbol{\mu}_n, \boldsymbol{\kappa}_n | \mathbf{x}_n) &\propto f(\mathbf{x}_n | \boldsymbol{\mu}_n, \boldsymbol{\kappa}_n) dp(\boldsymbol{\mu}_n, \boldsymbol{\kappa}_n) \\ &= \left[ \prod_{i=1}^n f(x_i | \mu_i, \kappa_i) \right] dp(\boldsymbol{\mu}_n) dp(\boldsymbol{\kappa}_n) \end{aligned} \quad (2)$$

the last equality being due to the independence of  $\boldsymbol{\mu}_n$  and  $\boldsymbol{\kappa}_n$ . Ferguson (1973) showed that if  $G \sim \mathcal{D}(M, G_0)$  and  $\mu | G \sim G$ , then

$$G|\mu \sim \mathcal{D}\left(M+1, \frac{MdG_0 + \delta_\mu}{M+1}\right)$$

Using this argument sequentially, the joint distribution of  $\boldsymbol{\mu}_n$  is (see Blackwell & MacQueen, 1973, Antoniak, 1974)

$$dp(\boldsymbol{\mu}_n) \propto \prod_{i=1}^n \frac{1}{(M+i-1)} \left[ MdG_0(\mu_i) + \sum_{j=1}^{i-1} \delta_{\mu_j}(d\mu_i) \right] \tag{3}$$

From this, we can write the univariate conditional pdfs as

$$dp(\mu_i | \boldsymbol{\mu}_n^{(i)}) \propto \frac{M}{M+n-1} dG_0(\mu_i) + \frac{1}{M+n-1} \sum_{j \neq i}^n \delta_{\mu_j}(d\mu_i) \tag{4}$$

$$i = 1, \dots, n$$

Hence, if  $\mu_{n+1}$  is a future value of the mean direction  $\mu$ , its distribution conditional on the past  $\boldsymbol{\mu}_n$  is given by

$$dp(\mu_{n+1} | \boldsymbol{\mu}_n) \propto \frac{M}{M+n} dG_0(\mu_{n+1}) + \frac{1}{M+n} \sum_{j=1}^n \delta_{\mu_j}(d\mu_{n+1}) \tag{5}$$

The prediction pdf of the unobserved future value  $x_{n+1}$  of  $x$ , given the observed  $\mathbf{x}_n$ , is

$$f(x_{n+1} | \mathbf{x}_n) = \int f(x_{n+1} | \mathbf{x}_n, \boldsymbol{\mu}_n, \boldsymbol{\kappa}_n) dp(\boldsymbol{\mu}_n, \boldsymbol{\kappa}_n | \mathbf{x}_n) \tag{6}$$

Since  $f(x_{n+1} | \mu_{n+1}, \kappa_{n+1}, \mathbf{x}_n, \boldsymbol{\mu}_n, \boldsymbol{\kappa}_n) = f(x_{n+1} | \mu_{n+1}, \kappa_{n+1})$  and  $dp(\mu_{n+1}, \kappa_{n+1} | \mathbf{x}_n, \boldsymbol{\mu}_n, \boldsymbol{\kappa}_n) = dp(\mu_{n+1} | \boldsymbol{\mu}_n) \times dH(\kappa_{n+1})$ , using equation (5), we have

$$\begin{aligned} & f(x_{n+1} | \mathbf{x}_n, \boldsymbol{\mu}_n, \boldsymbol{\kappa}_n) \\ &= \iint f(x_{n+1} | \mu_{n+1}, \kappa_{n+1}, \mathbf{x}_n, \boldsymbol{\mu}_n, \boldsymbol{\kappa}_n) dp(\mu_{n+1}, \kappa_{n+1} | \mathbf{x}_n, \boldsymbol{\mu}_n, \boldsymbol{\kappa}_n) \\ &= \iint f(x_{n+1} | \mu_{n+1}, \kappa_{n+1}) dp(\mu_{n+1} | \boldsymbol{\mu}_n) dH(\kappa_{n+1}) \\ &\propto \iint \exp[\kappa_{n+1} \cos(x_{n+1} - \mu_{n+1})] \tag{7} \\ &\times \left[ \frac{M}{M+n} dG_0(\mu_{n+1}) + \frac{1}{M+n} \sum_{j=1}^n \delta_{\mu_j}(d\mu_{n+1}) \right] dH(\kappa_{n+1}) \\ &= \frac{M}{M+n} \iint \exp[\kappa_{n+1} \cos(x_{n+1} - \mu_{n+1})] dG_0(\mu_{n+1}) dH(\kappa_{n+1}) \\ &+ \frac{1}{M+n} \sum_{j=1}^n \int \exp[\kappa_{n+1} \cos(x_{n+1} - \mu_j)] dH(\kappa_{n+1}) \end{aligned}$$

Note that the second term in equation (7) can be obtained from a linear combination of  $VM(\mu_j, \kappa)$  densities after averaging over the prior distribution of  $\kappa$ .

Hence, from equation (6), using the law of large numbers, an estimator of the predictive density of  $[x_{n+1} | \mathbf{x}_n]$  can be obtained by averaging equation (7) evaluated at various values of  $(\boldsymbol{\mu}_n, \boldsymbol{\kappa}_n)$  sampled from the posterior (2). Evaluation of equation (7) is straightforward if we choose  $G_0(\cdot)$  to be a von Mises or a cardioid distribution and  $H(\cdot)$  to be discrete (say, discrete uniform). Even with other choices of  $G_0(\cdot)$ , the evaluation is not too difficult, since equation (7) just requires a one-dimensional numerical integration at most. The procedure of sampling  $(\boldsymbol{\mu}_n, \boldsymbol{\kappa}_n)$  from equation (2) is described in the next subsection.

From equation (7), the suggested choice of  $M$  is given by the relation

$$p^* = \frac{M}{M+n} \quad \text{or equivalently,} \quad M = n \left( \frac{p^*}{1-p^*} \right)$$

where  $p^*$  is the proportion of belief we have in  $G_0$  being the true prior guess at  $G$ . For example, if we are 50% sure that the prior is  $G_0$ , we suggest a choice of  $M = n$ .  $M$  is sometimes interpreted as the ‘prior sample size’ (see Ferguson, 1973) although this interpretation has been disputed by Sethuraman & Tiwari (1982). Other methods of choice of  $M$  include putting a prior on  $M$  (see Escobar & West, 1995, Leite *et al.*, 1998) and using the maximum likelihood estimation (see Korwar & Hollander, 1973).

### 2.3 Computations

Sampling from the posterior (2) is done using the Gibbs sampler, a Markov Chain Monte Carlo (MCMC) procedure as outlined in Escobar & West (1995). This requires all the univariate conditional pdfs of  $[\mu_i | \boldsymbol{\mu}_n^{(i)}, \boldsymbol{\kappa}_n, \mathbf{x}_n]$  and  $[\kappa_i | \boldsymbol{\kappa}_n^{(i)}, \boldsymbol{\mu}_n, \mathbf{x}_n]$ . From equations (2) and (3), it is easily checked that

$$\begin{aligned} dp(\mu_i | \boldsymbol{\mu}_n^{(i)}, \boldsymbol{\kappa}_n, \mathbf{x}_n) &\propto M f(x_i | \mu_i, \kappa_i) dG_0(\mu_i) + \sum_{\substack{j=1 \\ j \neq i}}^n f(x_i | \mu_i, \kappa_i) \delta_{\mu_j}(\mathbf{d}\mu_i) \\ &= q_{i,0} dG_{i,0}(\mathbf{d}\mu_i) + \sum_{\substack{j=1 \\ j \neq i}}^n q_{i,j} \delta_{\mu_j}(\mathbf{d}\mu_i) \end{aligned} \quad (8)$$

Here,

$$dG_{i,0}(\mu_i) \propto f(x_i | \mu_i, \kappa_i) dG_0(\mu_i)$$

is the posterior of  $\mu_i$  obtained by updating the ‘baseline prior’  $dG_0$  using the likelihood of the  $i$ th observation  $f(x_i | \mu_i, \kappa_i)$  and the Bayes theorem, and is sometimes referred to as the ‘baseline posterior’.  $q_{i,j}$ s are constants with the following property:

$$\left. \begin{aligned} q_{i,0} &= C_i M h(x_i, \kappa_i), \\ q_{i,j} &= C_i f(x_i | \mu_j, \kappa_i) \quad (j \neq i) \end{aligned} \right\} \quad (9)$$

where

$$h(x_i, \kappa_i) = \int f(x_i | \mu_i, \kappa_i) dG_0(\mu_i)$$

and the constant  $C_i$  is chosen such that  $q_{i,0} + \sum_{j=1}^n (j \neq i) q_{i,j} = 1$ .

The univariate conditional posteriors for  $\kappa_i$  are given by

$$dp(\kappa_i | \mathbf{\kappa}_n^{(i)}, \boldsymbol{\mu}_n, \mathbf{x}_n) \propto f(x_i | \mu_i, \kappa_i) dH(\kappa_i) \tag{10}$$

The traditional Gibbs sampler works as follows: starting from an arbitrary initial value, the components of  $(\boldsymbol{\mu}_n, \mathbf{\kappa}_n)$  are sequentially updated according to the univariate conditional distributions in equations (8) and (10). The process is repeated until convergence to obtain the desired posterior distribution of  $(\boldsymbol{\mu}_n, \mathbf{\kappa}_n) | \mathbf{x}_n$ .

We will, however, slightly modify the computations by exploiting the special properties of the distributions involved. This, as we show later, will result in speedier convergence.

*2.3.1 Configuration* Since the Dirichlet process prior  $\mathcal{D}(M, G_0)$  is discrete with probability 1 (see Ferguson, 1973; Blackwell, 1973; Basu & Tiwari, 1982), with positive probability, not all  $\mu_1, \dots, \mu_n$  are distinct. Let there be  $n^*$  distinct  $\mu$ s given by  $\mu_1^*, \dots, \mu_{n^*}^*$  with corresponding frequencies  $w_1, \dots, w_{n^*}$  where  $w_1 + \dots + w_{n^*} = n$ . Following Escobar & West (1995), define the set of indices  $\mathcal{S} = \{S_1, \dots, S_n\}$  by  $S_k = j$  if  $\mu_k = \mu_j^*$  ( $1 \leq k \leq n, 1 \leq j \leq n^*$ ).  $\mathcal{S}$  is called the configuration of  $\boldsymbol{\mu}_n$ .  $\mathcal{S}$  determines a classification of  $\boldsymbol{\mu}_n$  into  $n^*$  distinct groups or clusters. Then equation (7) becomes

$$\begin{aligned} & f(x_{n+1} | \mathbf{x}_n, \mathcal{S}, \boldsymbol{\mu}_n, \mathbf{\kappa}_n) \\ & \propto \frac{M}{M+n} \int \int \exp[\kappa_{n+1} \cos(x_{n+1} - \mu_{n+1})] dG_0(\mu_{n+1}) dH(\kappa_{n+1}) \tag{11} \\ & + \frac{1}{M+n} \sum_{j=1}^{n^*} w_j \int \exp[\kappa_{n+1} \cos(x_{n+1} - \mu_j^*)] dH(\kappa_{n+1}) \end{aligned}$$

Similarly, for any  $i$  ( $1 \leq i \leq n$ ), we will have a cluster structure  $\mathcal{S}^{(i)}$  for  $\boldsymbol{\mu}_n^{(i)}$ . Let there be  $n^{(i)}$  ( $\leq n-1$ ) distinct  $\mu$ s among the  $\boldsymbol{\mu}_n^{(i)}$ . We denote them by  $\mu_1^{(i)}, \dots, \mu_{n^{(i)}}^{(i)}$  with the corresponding frequencies being  $w_1^{(i)}, \dots, w_{n^{(i)}}^{(i)}$  where  $w_1^{(i)} + \dots + w_{n^{(i)}}^{(i)} = n-1$ . We can then rewrite equation (8) in terms of  $\mu_j^{(i)}$  as follows:

$$dp(\mu_i | \boldsymbol{\mu}_n^{(i)}, \mathcal{S}^{(i)}, \mathbf{x}_n) = q_{i,0} dG_{i,0}(\mu_i) + \sum_{j=1}^{n^{(i)}} w_j^{(i)} q_j^{(i)} \delta_{\mu_j^{(i)}}(d\mu_i) \tag{12}$$

where  $q_j^{(i)} \propto f(x_i | \mu_j^{(i)}, \kappa_i)$  and  $q_{i,0} + \sum_{j=1}^{n^{(i)}} w_j^{(i)} q_j^{(i)} = 1$ . Using equation (12) in place of (8) reduces the amount of necessary computations, speeding up the algorithm.

Thus, generating an observation from univariate conditional (12) is equivalent to generating an observation from a mixture distribution which picks an observation from the ‘baseline posterior’  $dG_{i,0}(\cdot)$  with probability proportional to  $q_{i,0}$  or picks one of the other distinct  $\mu_j^{(i)}$ s with probability proportional to  $w_j^{(i)} q_j^{(i)}$ .

MacEachern (1994) suggests a remixing algorithm to prevent the possibility of the above MCMC procedure getting stuck in a few clusters and, as a result, not generating any new value of  $\mu_j^{(i)}$ . This may happen when the value of  $q_{i,0}$  becomes small compared with the sum  $\sum_{j=1}^{n^{(i)}} w_j^{(i)} q_j^{(i)}$ .

Let  $\mathcal{J}_j = \{i : S_i = j\}$  be the set of indices of the observations in group  $j, j = 1, \dots, n^*$ . Let  $\mathbf{x}_{(j)} = \{x_i : S_i = j\}$  be the corresponding group of observations. Once the set  $\mathcal{S}$

is known, the posterior analysis of  $\mu_j^*$ s becomes a collection of  $n^*$  independent analyses for  $j = 1, \dots, n^*$ . In particular,

$$\begin{aligned} dp(\mu_j^* | \mathbf{x}, \mathcal{S}, n^*, \mathbf{\kappa}_n) &\equiv dp(\mu_j^* | \mathbf{x}_{(j)}, \mathcal{S}, n^*, \mathbf{\kappa}_{(j)}) \\ &\propto \prod_{i \in \mathcal{J}_j} f(x_i | \mu_j^*, \kappa_i) dG_0(\mu_j^*), \quad j = 1, \dots, n^* \end{aligned} \quad (13)$$

This is just the posterior of  $\mu_j^*$  given  $x_i (i \in \mathcal{J}_j)$  sampled from  $f(\cdot | \mu_j^*)$ .

We thus proceed with Gibbs sampling by iterating through the following two steps in order—sampling at each stage is based on current values of the conditioning variables and the process is repeated until convergence (i.e. when steady state is reached):

- (1) Sample from  $[\kappa_i | \mathbf{\kappa}_n^{(i)}, \boldsymbol{\mu}_n, \mathbf{x}_n]$  for each  $i = 1, \dots, n$  using methods in equation (10).
- (2) Sample from  $[\mu_i | \boldsymbol{\mu}_n^{(i)}, \mathbf{\kappa}_n, \mathbf{x}_n]$  for each  $i = 1, \dots, n$  using methods outlined in equation (12). After completing the cycle, this will result in a new configuration  $\mathcal{S}$  and a corresponding value of  $n^*$ .
- (3) Given  $n^*$  and  $\mathcal{S}$ , generate a new set of parameters  $\boldsymbol{\mu}^*$  by sampling each new  $\mu_j^*$  from the relevant component posterior in equation (13).

All the previous calculations are easily modified to accommodate the situations where all the observations have the same concentration parameter (either random or fixed).

**2.3.2 Special case** For the special case of von Mises priors, we now give the expressions for  $q_{i,0}$ ,  $q_{i,j}$ ,  $dG_{i,0}$  and  $dp$  in equation (12). We will use the notation  $V(x, \mu, \kappa)$  to denote the density at  $x$  of  $VM(\mu, \kappa)$ .

Assuming  $G_0 \sim VM(\mu_0, \kappa_0)$ , we have

$$\begin{aligned} q_{i,0} &\propto \frac{M}{2\pi} \frac{I_0(\tilde{\kappa}_i)}{I_0(\kappa_i) \times I_0(\kappa_0)} \\ q_j^{(i)} &\propto V(x_i, \mu_j^{(i)}, \kappa_i) \\ dG_{i,0}(\mu_i) &= V(\mu_i, \tilde{x}_i, \tilde{\kappa}_i), \\ dp(\mu_j^* | \mathbf{x}_n, \mathcal{S}, n^*, \mathbf{\kappa}_n) &= V(\mu_j^*, \tilde{x}_j, \tilde{\kappa}_j) \end{aligned}$$

where

$$\begin{aligned} \tilde{\kappa}_i \cos \tilde{x}_i &= \kappa_i \cos x_i + \kappa_0 \cos \mu_0 \\ \tilde{\kappa}_i \sin \tilde{x}_i &= \kappa_i \sin x_i + \kappa_0 \sin \mu_0 \\ \tilde{\kappa}_j \cos \tilde{x}_j &= \sum_{i \in \mathcal{J}_j} \kappa_i \cos x_i + \kappa_0 \cos \mu_0 \\ \text{and } \tilde{\kappa}_j \sin \tilde{x}_j &= \sum_{i \in \mathcal{J}_j} \kappa_i \sin x_i + \kappa_0 \sin \mu_0 \end{aligned}$$

Hence,

$$f(x_{n+1} | \mathbf{x}_n, \boldsymbol{\mu}_n, \boldsymbol{\kappa}_n, \mathcal{S}) = \frac{1}{M+n} \int \left[ \frac{M}{2\pi} \times \frac{I_0(\tilde{\kappa})}{I_0(\kappa) \times I_0(\kappa_0)} + \sum_{j=1}^{n^*} w_j \times V(x_{n+1}, \mu_j^*, \kappa) \right] dH(\kappa) \tag{14}$$

where

$$\tilde{\kappa}^2 = \kappa^2 + \kappa_0^2 + 2\kappa\kappa_0 \cos(x_{n+1} - \mu_0)$$

Putting  $\kappa_0 = 0$  in the above gives a uniform baseline prior as a further subcase.

### 2.4 An example

We applied the above method to data on arrival times at an ICU in the UK, presented in Cox & Lewis (1966). The full data shown in Fig. 1 consist of arrival times (on a 24-hour clock) of 254 patients over a period of about 12 months from February 1963 to March 1964. The original data were converted to a circular scale by representing 0:00 hrs (12 o'clock midnight) as 0 radians and with usual clockwise rotation being positive. The figure was constructed using CircStats (see Lund, 2000).

The arrival times appear to have a peak at around 5:30 pm and hit a nadir in the early morning hours of around 3:00–7:00 am, when people are the least active. This suggests that perhaps a lot of the arrivals are due to traffic-related incidents, which peak during the evening rush-hours.

We used the first 60 observations (in the order of occurrence) to illustrate the proposed methods. Assuming a common von Mises distribution of the arrival times, the maximum likelihood estimates (MLEs) of the mean direction and

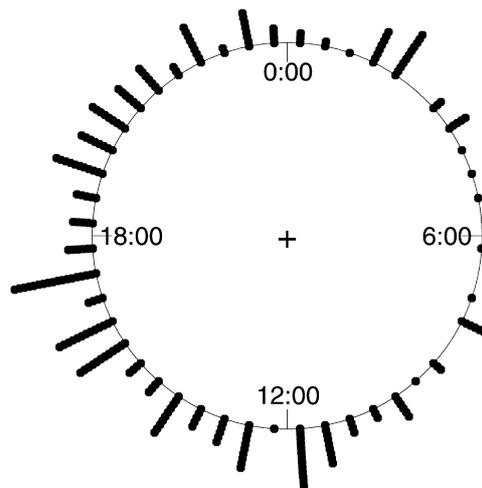


FIG. 1. Circular dotplot of the 254 arrival times at the Intensive Care Unit. Data taken from Cox & Lewis (1966).

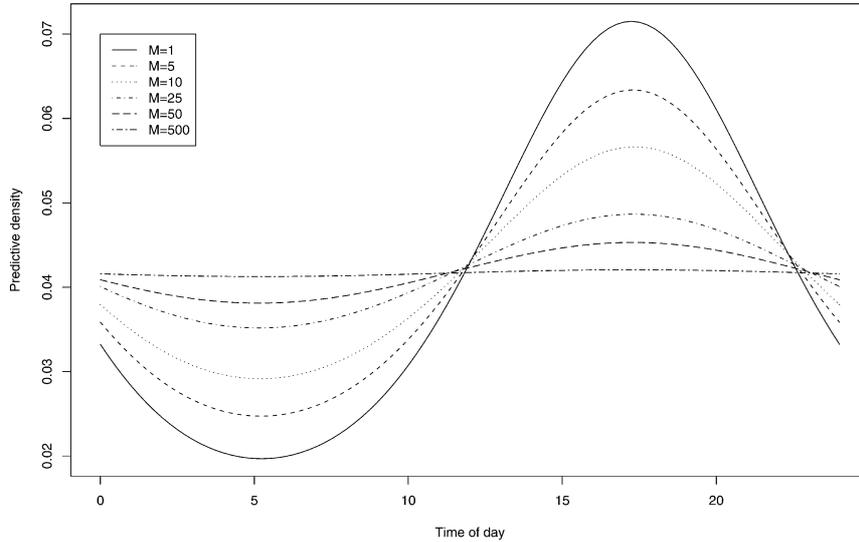


FIG. 2. Effect of change in belief in baseline prior on the predictive density; uniform baseline prior.

concentration parameter are  $\hat{\mu} = 4.55$  radians (equivalent to 5:23 pm) and  $\hat{\kappa} = 0.7299$  respectively. Thus, a frequentist approach to this prediction problem would result in a von Mises distribution with  $\mu = 4.55$  radians and  $\kappa = 0.7299$ . Note, however, that the assumption of a common von Mises distribution is possibly not valid, due to the pronounced multimodality of the data.

We used three methods to get a predictive density : (i) assuming the  $\kappa$ s are all equal and known, (ii) assuming  $\kappa$ s are all equal and random and, (iii) assuming  $\kappa$ s are unequal and random. In all the cases, we assumed a von Mises baseline prior for the  $\mu$ s with  $\mu_0 = \hat{\mu}$  and  $\kappa_0 = \hat{\kappa}$ . For (i), we chose  $\kappa = \hat{\kappa}$ . For (ii) and (iii), we assumed that the common distribution of  $\kappa$  is discrete uniform with support on  $k$  equidistant points (including 0) such that the prior mean equals  $\hat{\kappa}$ . Similar analyses were run with a uniform baseline prior for  $\mu$ , achieved by taking  $\kappa_0 = 0$ .

The following features of the predictive density were evident:

- Increasing the value of  $M$  led to a flatter density (see Figs 2 and 3). This is explained by the fact that higher  $M$  implies more faith in baseline prior, resulting in more  $\mu$ s drawn from the baseline posterior. This gives a ‘wider variety’ of  $\mu$ s to average upon, as opposed to only point masses when  $M$  is close to 0.
- A switch from equal to unequal  $\kappa$ s introduced more uncertainty in the model, which is reflected in a flatter density (see Fig. 4). However, if the  $\kappa$ s were known and equal, the predictive density is between the above graphs in Fig. 4. This is because, although there is less randomness (because the  $\kappa$ s are known and equal), there is less dependence among the original data.
- There is negligible effect of changing the number of points in the discrete uniform prior for  $\kappa$  (see Fig. 5).
- The effect of deleting a few observations was also found to be negligible, provided that the effect of the sample was small compared with the prior (i.e.  $M/(M+n)$  is large). Hence, the proposed analysis is robust to extreme data.

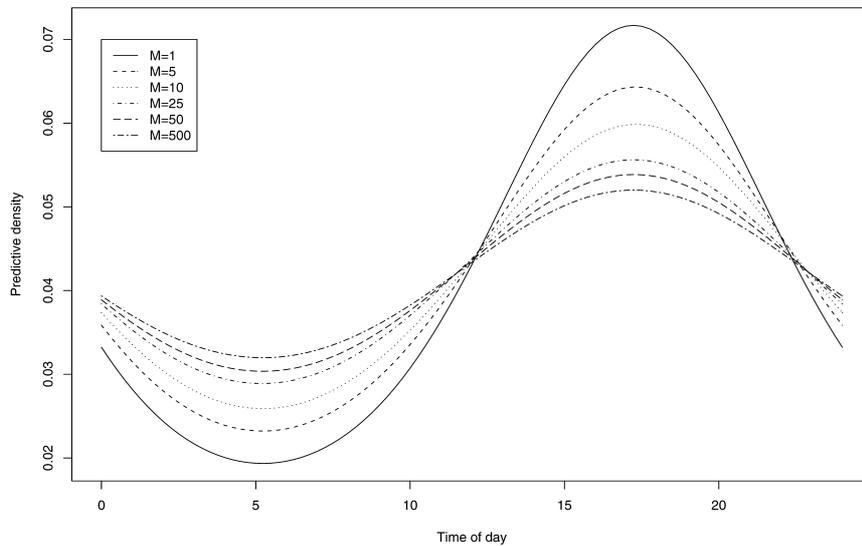


FIG. 3. Effect of change in belief in baseline prior on the predictive density; von Mises baseline prior.

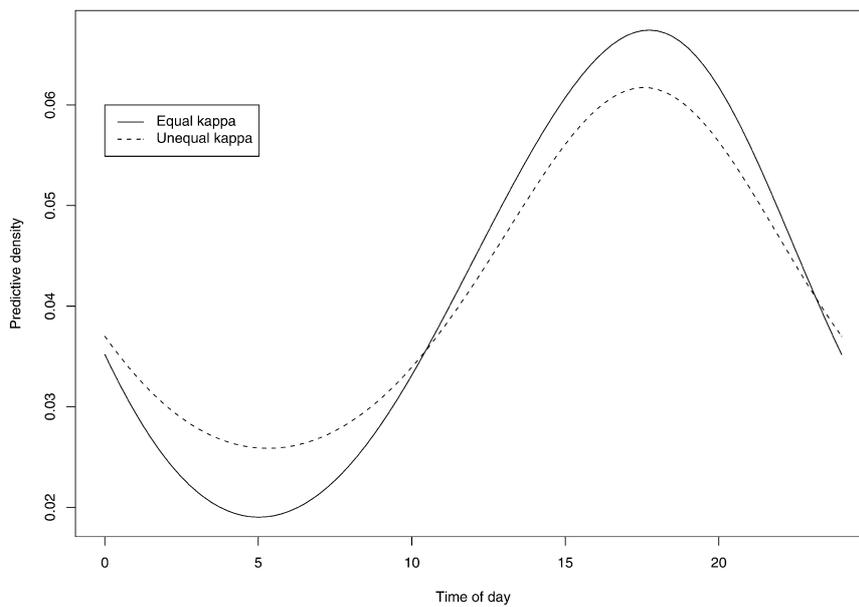


FIG. 4. Effect of unequal  $\kappa$ s.

All the above calculations were done using C routines. The uniform random number generators and sorting routines were based on Press *et al.*(1997). Generation of von Mises variables was performed using the algorithm due to Best & Fisher (1979). From empirical studies using different starting values, we found the Gibbs sampler quickly reached steady state. We chose the first 10,000 simulations as our burn-in and the results of further 10 000 simulations were used for the predictive density calculations.

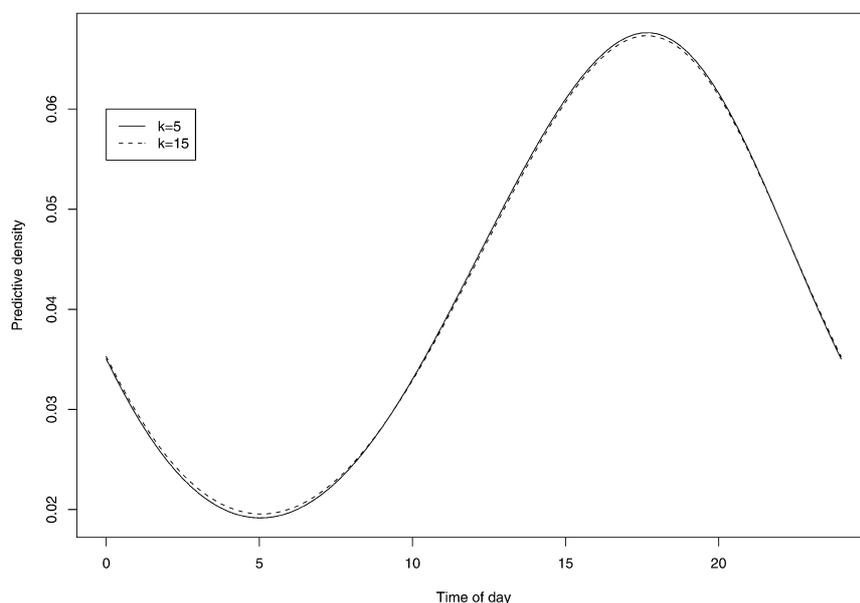


FIG. 5. Effect of change in number of points ( $k$ ) in the discrete uniform prior of  $\kappa$ .

### 3 Two-sample test

#### 3.1 Introduction and notations

Suppose we have (circular) data from two sources, and we want to test whether their mean directions are the same. For example, we have data from two hospitals on their emergency room arrivals and would like to see whether the average arrival time is the same in both cases. For simplicity, we assume that the two data sets are coming from von Mises populations with the same concentration parameter within a group. The problem then is one of testing whether the mean directions of the von Mises populations are the same. The traditional frequentist approach uses the Neyman–Pearson type of argument to come up with large sample tests (such as the two-sample  $t$ -test) that control the Type I error while minimizing the Type II error. We will be using a semiparametric Bayesian approach in this article. Formally, let

$$\mathbf{x}_n = (x_1, \dots, x_n)$$

and

$$\mathbf{y}_m = (y_1, \dots, y_m)$$

be observed data on directions from two populations. We assume that

$$x_i | (\mu_1, \kappa_1) \stackrel{\text{i.i.d.}}{\sim} VM(\mu_1, \kappa_1), (i = 1, \dots, n)$$

and

$$y_j | (\mu_2, \kappa_2) \stackrel{\text{i.i.d.}}{\sim} VM(\mu_2, \kappa_2), (j = 1, \dots, m)$$

Let us write  $\boldsymbol{\theta} = (\mu_1, \mu_2, \kappa_1, \kappa_2)$ ,

$$\Theta = \{\boldsymbol{\theta}: 0 \leq \mu_1 < 2\pi, 0 \leq \mu_2 < 2\pi, \kappa_1 > 0, \kappa_2 > 0\}$$

$$\Theta_0 = \{\boldsymbol{\theta} \in \Theta: \mu_1 = \mu_2\}$$

and  $\Theta_1 = \{\boldsymbol{\theta} \in \Theta: \mu_1 \neq \mu_2\}$

Then  $\Theta_0 \cup \Theta_1 = \Theta$  and that  $\Theta_0 \cap \Theta_1 = \emptyset$ . We want to test  $H_0: \boldsymbol{\theta} \in \Theta_0$  against the alternative  $H_1: \boldsymbol{\theta} \in \Theta_1$ .

Let  $p_i = P(H_i)$  be the prior probability that  $H_i$  is true ( $i = 0, 1$ ). In addition, let  $\boldsymbol{\theta} | H_i \sim g_i(\cdot)$  be the distribution of  $\boldsymbol{\theta}$  when  $H_i$  is true ( $i = 0, 1$ ). The likelihood of  $\boldsymbol{\theta}$  is

$$L(\boldsymbol{\theta} | \mathbf{x}_n, \mathbf{y}_m) = \left[ \prod_{i=1}^n f(x_i | \mu_1, \kappa_1) \right] \left[ \prod_{j=1}^m f(y_j | \mu_2, \kappa_2) \right]$$

By Bayes theorem,

$$\begin{aligned} P(H_i | \mathbf{x}_n, \mathbf{y}_m) &= \frac{\int_{\Theta} P(\mathbf{x}_n, \mathbf{y}_m | H_i, \boldsymbol{\theta}) P(\boldsymbol{\theta} | H_i) P(H_i) \, d\boldsymbol{\theta}}{\sum_{k=0}^1 \int_{\Theta} P(\mathbf{x}_n, \mathbf{y}_m | H_k, \boldsymbol{\theta}) P(\boldsymbol{\theta} | H_k) P(H_k) \, d\boldsymbol{\theta}} \\ &= \frac{\int_{\Theta_i} P(\mathbf{x}_n, \mathbf{y}_m | H_i, \boldsymbol{\theta}) g_i(\boldsymbol{\theta}) p_i \, d\boldsymbol{\theta}}{\sum_{k=0}^1 \int_{\Theta_k} P(\mathbf{x}_n, \mathbf{y}_m | H_k, \boldsymbol{\theta}) P(\boldsymbol{\theta} | H_k) P(H_k) \, d\boldsymbol{\theta}} \\ &= \frac{p_i \int_{\Theta_i} L(\boldsymbol{\theta} | \mathbf{x}_n, \mathbf{y}_m) g_i(\boldsymbol{\theta}) \, d\boldsymbol{\theta}}{\sum_{k=0}^1 p_k \int_{\Theta_k} L(\boldsymbol{\theta} | \mathbf{x}_n, \mathbf{y}_m) g_k(\boldsymbol{\theta}) \, d\boldsymbol{\theta}} \end{aligned}$$

for  $i = 0, 1$  are the posterior probabilities of the two hypotheses. We will select that hypothesis to be true whose posterior probability is the highest (reflected in the Bayes factor). Unfortunately, exact calculations in the above are difficult to perform. Hence, we resort to simulated values of the posterior probabilities. Once we simulate a large number (say  $N$ ) of observations  $\boldsymbol{\theta}$  from  $\{\boldsymbol{\theta} | \mathbf{x}_n, \mathbf{y}_m\}$ , our estimate of the posterior probability of a hypothesis  $H_r$  is

$$P(H_r | \mathbf{x}_n, \mathbf{y}_m) = \frac{1}{N} (\text{number of times } H_r \text{ occurs among the } N \text{ sampled } \boldsymbol{\theta}\text{s})$$

### 3.2 Theory and computations

We assume that  $\mu_1, \mu_2 \lambda G$  and that  $G \sim \mathcal{D}(M, G_0)$ . Further, assume as before that  $\kappa_1, \kappa_2 \stackrel{\text{i.i.d.}}{\sim} H$  and that the  $\mu$ s and  $\kappa$ s are independent. Using interpretations as in the previous section,  $G_0$  is the baseline prior and  $M$  quantifies our belief that  $G_0$  is

actually the value of  $G$ . In fact, it is easily shown that  $p_0 = 1/(M + 1)$  and  $p_1 = M/(M + 1)$  are the prior probabilities of the two hypotheses, thus relating the choice of  $M$  to our prior beliefs in the two hypotheses (see Gopalan & Berry, 1998). Using the results given earlier, we have

$$\begin{aligned} dG(\mu_1 | \mu_2) &\propto \frac{M}{M+2-1} dG_0(\mu_1) + \frac{1}{M+2-1} \delta_{\mu_2}(d\mu_1) \\ dG(\mu_2 | \mu_1) &\propto \frac{M}{M+2-1} dG_0(\mu_2) + \frac{1}{M+2-1} \delta_{\mu_1}(d\mu_2) \end{aligned}$$

Let  $\boldsymbol{\kappa}_2 = (\kappa_1, \kappa_2)$  and  $\boldsymbol{\mu}_2 = (\mu_1, \mu_2)$ . Then,

$$\begin{aligned} dG(\mu_1 | \mu_2, \boldsymbol{\kappa}_2, \mathbf{x}_n, \mathbf{y}_m) &\sim q_{1,0} dG_{1,0}(\mu_1) + q_{1,2} \delta_{\mu_2}(d\mu_1) \\ dG(\mu_2 | \mu_1, \boldsymbol{\kappa}_2, \mathbf{x}_n, \mathbf{y}_m) &\sim q_{2,0} dG_{2,0}(\mu_2) + q_{2,1} \delta_{\mu_1}(d\mu_2) \end{aligned}$$

where

$$\begin{aligned} q_{1,0} &= c_1 M h(\mathbf{x}_n, \kappa_1), \quad q_{1,2} = c_1 f(\mathbf{x}_n | \mu_2, \kappa_1) \\ q_{2,0} &= c_2 M h(\mathbf{y}_m, \kappa_2), \quad q_{2,1} = c_2 f(\mathbf{y}_m | \mu_1, \kappa_2) \end{aligned}$$

The constants  $c_1$  and  $c_2$  are chosen such that

$$q_{1,0} + q_{1,2} = 1 = q_{2,0} + q_{2,1}$$

$$h_1(\mathbf{x}_n, \kappa_1) = \int f(\mathbf{x}_n | \mu_1, \kappa_1) dG_0(\mu_1)$$

$$h_2(\mathbf{y}_m, \kappa_2) = \int f(\mathbf{y}_m | \mu_2, \kappa_2) dG_0(\mu_2)$$

and  $dG_{1,0}(\cdot)$ ,  $dG_{2,0}(\cdot)$  are the posterior pdfs of  $\mu_1$  and  $\mu_2$  respectively with respect to the baseline prior  $dG_0(\cdot)$ :

$$\begin{aligned} dG_{1,0}(\mu_1) &\propto f(\mathbf{x}_n | \mu_1, \kappa_1) dG_0(\mu_1) \\ dG_{2,0}(\mu_2) &\propto f(\mathbf{y}_m | \mu_2, \kappa_2) dG_0(\mu_2) \end{aligned}$$

The updating equations for  $\boldsymbol{\kappa}_2$  are given by

$$\begin{aligned} dp(\kappa_1 | \kappa_2, \boldsymbol{\mu}_2, \mathbf{x}_n, \mathbf{y}_m) &\propto \prod_{i=1}^n f(x_i | \mu_1, \kappa_1) dH(\kappa_1), \\ dp(\kappa_2 | \kappa_1, \boldsymbol{\mu}_2, \mathbf{x}_n, \mathbf{y}_m) &\propto \prod_{j=1}^m f(y_j | \mu_2, \kappa_2) dH(\kappa_2) \end{aligned}$$

All the above are implemented in a Gibbs sampler as before, along with the remixing algorithm. The updating equations are easily modified to accommodate a common  $\boldsymbol{\kappa}$  that can be random or fixed.

3.2.1 *Special case* Assuming  $G_0 \sim VM(\mu_0, \kappa_0)$ , we have

$$h_1(\mathbf{x}_n, \kappa_1) = \frac{I_0(\tilde{\kappa}_1)}{I_0(\kappa_0)[2\pi I_0(\kappa_1)]^n}$$

$$h_2(\mathbf{y}_m, \kappa_2) = \frac{I_0(\tilde{\kappa}_2)}{I_0(\kappa_0)[2\pi I_0(\kappa_2)]^m}$$

$$dG_{1,0}(\mu_1) = V(\mu_1, \tilde{x}, \tilde{\kappa}_1)$$

$$dG_{2,0}(\mu_2) = V(\mu_2, \tilde{y}, \tilde{\kappa}_2)$$

where

$$\tilde{\kappa}_1 \cos \tilde{x} = \kappa_1 \sum_{i=1}^n \cos x_i + \kappa_0 \cos \mu_0$$

$$\tilde{\kappa}_1 \sin \tilde{x} = \kappa_1 \sum_{i=1}^n \sin x_i + \kappa_0 \sin \mu_0$$

$$\tilde{\kappa}_2 \cos \tilde{y} = \kappa_2 \sum_{i=1}^m \cos y_i + \kappa_0 \cos \mu_0$$

$$\tilde{\kappa}_2 \sin \tilde{y} = \kappa_2 \sum_{i=1}^m \sin y_i + \kappa_0 \sin \mu_0$$

Putting  $\kappa_0 = 0$  gives the uniform baseline prior.

### 3.3 Simulation results

We applied the above method on data of cross-bed azimuths of palaeocurrents taken from Fisher & Powell (1989). The data consist of palaeocurrents measured at two sites in the Belford Anticline at New South Wales, Australia. Palaeocurrent analysis is used to determine the direction of flows of ancient rivers. The goal here is to detect if the two sites have a common mean direction.

As before, we analysed the data under 3 main assumptions: (i)  $\kappa$ s are known and equal, (ii)  $\kappa$ s are random and equal, and (iii)  $\kappa$ s are random and unequal. In all the cases, we ran separate analyses using von Mises and uniform baselines for  $\mu$ .

The von Mises populations were assumed to have a common concentration of 1.0 (MLE based on the observed data) for (i); for (ii) and (iii) a discrete uniform prior was used for  $\kappa$ .

Calculations were performed both for uniform and von Mises baseline priors. In the latter case, the mean direction of the baseline was chosen to be the (sample) mean direction of the combined data. 10 000 iterations were used to let the Gibbs sampler stabilize and a further 10 000 iterations were used to calculate the Bayes factor. Table 1 gives the Bayes factors for testing the equality of the two mean directions.

As expected, we found that introduction of the randomness in the  $\kappa$ s resulted in smaller Bayes factors for the von Mises baseline but larger ones for a uniform baseline. The calculations show that even a moderate baseline prior (uniform or von Mises) probability of  $H_0$  is amplified by the data to show more evidence in

TABLE 1. Bayes factors under uniform and von Mises priors.

$M$	$\kappa$ known		$\kappa$ random		$\kappa_1, \kappa_2$ random	
	Uniform	von Mises	Uniform	von Mises	Uniform	von Mises
99.00	8.5386	14.6364	9.7673	3.6971	0.8890	4.9916
49.00	7.6998	12.3266	8.5793	3.5977	0.9847	4.7399
32.33	7.8124	12.1907	8.4247	3.7530	1.0103	4.6825
24.00	7.7460	10.9905	8.2711	3.5514	1.0600	4.4158
19.00	7.1636	10.1322	7.6704	3.5760	1.0570	4.2843
15.67	7.0947	10.0712	7.4986	3.4857	1.0160	4.1620
13.28	6.4994	9.5498	7.2931	3.4406	0.9663	4.1679
11.50	6.8063	9.3598	7.2816	3.4370	1.0491	3.9570
10.11	6.5026	9.1042	7.1316	3.4046	1.0049	3.9477
9.00	6.3610	8.4182	6.8200	3.3001	1.0458	3.9478
5.67	6.1266	7.7742	6.3926	3.0580	0.9703	3.5116
4.00	5.6479	7.3282	5.9010	2.8306	0.9505	3.2267
3.00	5.3940	6.8977	5.7032	2.7088	1.0069	2.9654
2.33	5.2326	6.4883	5.4705	2.5987	1.0303	2.7668
1.86	5.0161	6.3385	4.9506	2.4759	0.9817	2.7160
1.50	4.8586	6.5645	4.8131	2.3760	1.0265	2.6040
1.22	4.8404	6.3929	4.7720	2.2919	1.0122	2.4437
1.00	5.1387	5.9348	4.8548	2.2300	1.0052	2.3212
0.82	4.7363	5.9548	4.8089	2.1144	0.9714	2.3226
0.67	4.7271	5.9870	4.4146	2.1134	0.9921	2.2458
0.54	4.5800	6.2007	4.7718	2.0171	1.0401	2.1325
0.43	4.4582	5.7114	4.1162	2.0274	0.9863	2.0274
0.33	4.6123	5.9089	4.0354	1.9420	0.9744	2.0682
0.25	4.6713	6.0000	4.3968	1.8561	1.0301	1.9201
0.18	4.1068	5.5346	3.9467	1.8851	0.9694	1.9652
0.11	4.2633	5.7678	4.4614	1.8146	0.9614	1.8801
0.05	4.6466	5.0081	4.3702	1.7018	1.0817	1.8271
0.01	5.0404	7.2049	4.7999	2.1858	0.9519	1.7620

support of  $H_0$  than of  $H_1$  (Bayes factor  $> 1$ ). Note that in the case of uniform baseline when the two concentrations are random, data do not provide evidence for one hypothesis over the other. We thus conclude that the palaeocurrents at the two sites have the same average direction.

#### 4 Conclusion

We have looked at prediction and testing problems for directional data using Bayesian semiparametric techniques. The methods, although computationally intensive, are easy to implement. The standard frequentist and parametric Bayesian approaches are shown to be special cases of these methods.

The prediction procedure can be easily modified to incorporate uncertainty in the precision parameter ( $M$ ), by incorporating a prior distribution for  $M$ . Possible uncertainty in  $H$  can be modelled by another Dirichlet process prior:  $H \sim \mathcal{D}(M_H, H_0)$ . The assumption of independence of  $\mu$ s and  $\kappa$ s can also be relaxed.

Instead of assuming a common distribution for all the  $\mu$ s, we might employ a mixture distribution:  $\mu_j \sim \varepsilon G + (1 - \varepsilon)G_j$ , where  $G_1, \dots, G_n \sim \mathcal{D}(M, G)$ .  $\varepsilon$  can be assumed to have a Beta distribution, adding to another level in the hierarchy.

The testing procedure is easily generalized to several hypotheses. It is, however, apparent from Table 1 that the Bayes Factor computations are unstable when using

the ‘hit-and-miss’ method, especially in the case of extreme  $M$  values. An alternative approach due to Chib (1995) is worth investigating.

Finally, it would be of interest to see how the above methods can be modified to include additional information through covariates (say, a predictive density of ICU arrivals during the summer as opposed to winter, males versus females, etc).

## REFERENCES

- ANTONIAK, C. E. (1974) Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems, *The Annals of Statistics*, 2(6), 1152–1174.
- BASU, D. & TIWARI, R. C. (1982) A note on the Dirichlet process. In: *Statistics and Probability: Essays in Honor of C. R. Rao*, pp.89–103 (New York, North-Holland).
- BEST, D. J. & FISHER, N. I. (1979) Efficient simulation of the von Mises distribution, *Applied Statistics*, 28, 152–157.
- BLACKWELL, D. (1973) Discreteness of Ferguson selections, *Annals of Statistics*, 1, 356–358.
- BLACKWELL, D. & MACQUEEN, J. B. (1973) Ferguson distributions via Polya urn schemes, *Annals of Statistics*, 1, 353–355.
- CHIB, S. (1995) Marginal likelihood from the Gibbs output, *Journal of the American Statistical Association*, 90, 1313–1321.
- COX, D. R. & LEWIS, P. A. W. (1966) *The Statistical Analysis of Series of Events* (London, Methuen).
- ESCOBAR, M. D. & WEST, M. (1995) Bayesian density estimation and inference using mixtures, *Journal of the American Statistical Association*, 90(430), 577–588.
- FERGUSON, T. S. (1973) A Bayesian analysis of some nonparametric problems, *The Annals of Statistics*, 1(2), 209–230.
- FISHER, N. I. (1995) *Statistical Analysis of Circular Data* (Cambridge University Press).
- FISHER, N. I. & POWELL, C. M. (1989) Statistical analysis of two-dimensional palaeocurrent data: methods and examples, *Australian Journal of Earth Sciences*, 36, 91–107.
- GOPALAN, R. & BERRY, D. A. (1998) Bayesian multiple comparisons using Dirichlet process priors, *Journal of the American Statistical Association*, 93(443), 1130–1139.
- JAMMALAMADAKA, S. & SENGUPTA, A. (2001) *Topics in Circular Statistics* (Singapore, World Scientific Press).
- KORWAR, R. M. & HOLLANDER, M. (1973) Contributions to the theory of Dirichlet processes, *Annals of Statistics*, 1, 705–711.
- LEITE, J. G., SALINAS-TORRES, V. H., TIWARI, R. C. & ZALKIKAR, J. N. (1998) Bayes estimation of Dirichlet processes, *Sankhya*, 1, 102–118.
- LUND, U. (2000) CircStats 1.0, a collection of S-Plus routines for analyzing circular data.
- MACEachern, S. N. (1994) Estimating normal means with a conjugate style Dirichlet process prior, *Communications in Statistics: Simulation and Computation*, 23, 727–741.
- MARDIA, K. V. & JUPP, P. E. (2000) *Directional Statistics* (UK, Wiley).
- PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T. & FLANNERY, B. P. (1997) *Numerical Recipes in C*, 2nd edn (Cambridge University Press).
- SETHURAMAN, J. & TIWARI, R. C. (1982) *Convergence of Dirichlet measures and the interpretation of their parameter*, pp. 305–315 (New York, Academic Press).
- WILKS, S. S. (1962) *Mathematical Statistics* (New York, Wiley).

