



Working Paper 08-13
Statistic and Econometric Series 05
March 2008

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34-91) 6249849

A SEMI-PARAMETRIC MODEL FOR CIRCULAR DATA BASED ON MIXTURES OF BETA DISTRIBUTIONS *

Jose Antonio Carnicero ¹, Michael Peter Wiper ²

Abstract

This paper introduces a new, semi-parametric model for circular data, based on mixtures of shifted, scaled, beta (SSB) densities. This model is more general than the Bernstein polynomial density model which is well known to provide good approximations to any density with finite support and it is shown that, as for the Bernstein polynomial model, the trigonometric moments of the SSB mixture model can all be derived.

Two methods of fitting the SSB mixture model are considered. Firstly, a classical, maximum likelihood approach for fitting mixtures of a given number of SSB components is introduced. The Bayesian information criterion is then used for model selection. Secondly, a Bayesian approach using Gibbs sampling is considered. In this case, the number of mixture components is selected via an appropriate deviance information criterion.

Both approaches are illustrated with real data sets and the results are compared with those obtained using Bernstein polynomials and mixtures of von Mises distributions.

Keywords: circular data; shifted, scaled, beta distribution; mixture models; Bernstein polynomials
JEL Classification: C14

¹ Corresponding autor: Jose Antonio Carnicero Carreño, Universidad Carlos III Departamento de Estadística, C/ Madrid 126. 28903 Getafe (Madrid), España, e-mail: jcarnice@estecon.uc3m.es,

² Corresponding autor: Michael P. Wiper, Universidad Carlos III, Departamento de Estadística, C/ Madrid 126. 28903 Getafe (Madrid), España, e-mail: mwiper@est-econ.uc3m.es

A semi-parametric model for circular data based on mixtures of beta distributions

J.A. Carnicero and M.P. Wiper

April 3, 2008

Abstract

This paper introduces a new, semi-parametric model for circular data, based on mixtures of shifted, scaled, beta (SSB) densities. This model is more general than the Bernstein polynomial density model which is well known to provide good approximations to any density with finite support and it is shown that, as for the Bernstein polynomial model, the trigonometric moments of the SSB mixture model can all be derived.

Two methods of fitting the SSB mixture model are considered. Firstly, a classical, maximum likelihood approach for fitting mixtures of a given number of SSB components is introduced. The Bayesian information criterion is then used for model selection. Secondly, a Bayesian approach using Gibbs sampling is considered. In this case, the number of mixture components is selected via an appropriate deviance information criterion.

Both approaches are illustrated with real data sets and the results are compared with those obtained using Bernstein polynomials and mixtures of von Mises distributions.

Keywords: circular data; shifted, scaled, beta distribution; mixture models; Bernstein polynomials.

AMS classification: 62H11.

1 Introduction

Problems where the data are angular directions occur in many different scientific fields such as biology (direction of movement of migrating animals),

meteorology (wind directions) and geology (directions of joints and faults). Also, phenomena that are periodic in time such as times of hospital admittance for births or the times when crimes are committed may also be converted to angular data via a simple transformation modulo some period. Data of this type are commonly known as circular data and are usually represented as points on the circumference of an unit circle or as angles, θ , where $0 \leq \theta < 2\pi$ radians, which represent the positive angle of rotation from some conveniently chosen origin.

It is clear that standard statistics are not always appropriate for circular data. For example, the pair of twenty four hour clock times 23:30 and 00:30 are closer to each other than are the pair 01:00 and 11:00 even though the numerical difference between the times is smaller in the second case. Also, the linear mean of the first pair of times is 12:00 which is clearly not a good measure of the average of the first pair of data. Instead, the “mean direction” is 24:00. A full review of circular moments and alternative statistics for circular data is contained in e.g. Mardia and Jupp (1999).

A number of parametric models for circular data have been developed using a variety of techniques. One approach is to “wrap” standard distributions on the line onto the circle, that is to define the circular density at some angle, θ , to be the sum of the linear densities at $\theta, \theta \pm 2\pi, \theta \pm 4\pi, \dots$, see e.g. Mardia and Jupp (1999), Jammalamadaka and Kozubowski(2003). Other models have been developed specifically for circular data. In particular, the von Mises distribution is an appropriate model for symmetric, unimodal, circular data and has many similar properties to the normal distribution on the line, see e.g. Evans et al (2000).

Most models developed thus far for circular data are unimodal and symmetric. However, in many cases, both multimodal and asymmetric data may be encountered. In such cases, non-parametric or semi-parametric models might be considered. Fully non-parametric, kernel based approaches are considered by Fisher (1989). However, semi-parametric models will usually have more predictive power. Fernández-Durán (2004) has developed a semi-parametric model based on trigonometric sums and Mooney et al (2003) have considered mixtures of von Mises distributions. Both these approaches have certain disadvantages. In the first case, as noted by Ferreira et al (2005), the trigonometric sums approach can often fit wiggly densities which are unrealistic in practice and in the second case, as the von Mises distribution is symmetric, large numbers of terms may be needed to well approximate asymmetric data samples. In this paper, we introduce an alternative semi-

parametric, mixture model based on the use of shifted, scaled, beta (SSB) distributions.

The article is organized as follows. In Section 2 we define the SSB distribution and show how its circular moments may be calculated. We then introduce mixtures of SSB distributions and relate this mixture model to Bernstein polynomials which have been developed as an approach to estimating densities with support $[0, 1]$ by e.g. Vitale (1975), Petrone (1999a,b), Petrone and Wassermann (2002), Babu et al (2002) and Kakizawa (2004). In Section 3 we consider maximum likelihood estimation for the SSB mixture model and we fit our model to two, well known, real data sets, comparing the results obtained with mixtures of von Mises distributions and with Bernstein polynomials. In Section 4 we describe a Bayesian algorithm for estimating the predictive density and fit the model to a third group of real data. Finally, some concluding remarks and possible extensions are given in Section 5.

2 The SSB distribution and mixtures of SSB distributions

2.1 The SSB distribution

Let θ be a random variable with support $[0, 2\pi)$. Then we shall say that θ has a SSB distribution with parameters $0 \leq \nu < 2\pi$ and $\alpha, \beta \in \mathbb{N}$, denoted by $\theta \sim \mathcal{SSB}(\nu, \alpha, \beta)$ if

$$f(\theta) = \begin{cases} \frac{1}{(2\pi)^{\alpha+\beta-1}B(\alpha,\beta)}(2\pi + \theta - \nu)^{\alpha-1}(\nu - \theta)^{\beta-1} & \text{if } 0 \leq \theta < \nu \\ \frac{1}{(2\pi)^{\alpha+\beta-1}B(\alpha,\beta)}(\theta - \nu)^{\alpha-1}(2\pi + \nu - \theta)^{\beta-1} & \text{if } \nu \leq \theta < 2\pi \end{cases} \quad (1)$$

where $B(\alpha, \beta) = (\alpha - 1)!(\beta - 1)!/(\alpha + \beta - 1)!$ is the beta function.

Clearly, we can represent the distribution of θ as the distribution of a shifted, scaled, beta distributed random variable, so that

$$\theta = \text{mod}(2\pi X + \nu, 2\pi)$$

where $\text{mod}(a, b)$ means a modulo b and X is a beta distributed random variable, $X \sim \mathcal{B}(\alpha, \beta)$, such that

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} \quad \text{for } 0 \leq x < 1. \quad (2)$$

The restriction in the definition of the SSB distribution that α and β are natural numbers is not strictly necessary. However, when fitting mixtures of SSB distributions to data, it is useful to include this restriction in order to avoid the possibility of fitting asymptotes which are unrealistic and would occur if α or β are below 1. Secondly, given that α and β are natural numbers, exact values for the circular or trigonometric moments of the SSB distribution can be derived as is shown in the following subsection.

2.2 Trigonometric moments of the SSB distribution

As noted in the introduction, the standard mean and other moments are not appropriate for circular random variables. Instead, for any circular random variable, θ , the p 'th trigonometric moment is defined to be

$$\begin{aligned} \mu'_p &= E[e^{ip\theta}] = E[\cos p\theta] + iE[\sin p\theta] \quad \text{where } i = \sqrt{-1} \\ &\stackrel{\text{def}}{=} \rho_p (\cos \mu'_p + i \sin \mu'_p) \quad \text{where} \\ \rho_p &= \sqrt{E[\cos p\theta]^2 + E[\sin p\theta]^2} \quad \text{and} \end{aligned} \tag{3}$$

$$\mu'_p = \begin{cases} \tan^{-1} \frac{E[\sin p\theta]}{E[\cos p\theta]} & \text{if } E[\sin p\theta] > 0 \text{ and } E[\cos p\theta] > 0 \\ \tan^{-1} \frac{E[\sin p\theta]}{E[\cos p\theta]} + \pi & \text{if } E[\cos p\theta] < 0 \\ \tan^{-1} \frac{E[\sin p\theta]}{E[\cos p\theta]} + 2\pi & \text{if } E[\sin p\theta] < 0 \text{ and } E[\cos p\theta] > 0 \end{cases} \tag{4}$$

for $p = 1, 2, \dots$. In particular, when $p = 1$, we write μ for μ'_1 and ρ for ρ_1 . Then, μ is the mean direction and ρ is the mean resultant length, see e.g. Mardia and Jupp (1999) for more details.

Similar sample equivalents can also be defined. Thus, given a sample of circular data, $\theta_1, \dots, \theta_n$, then the circular mean, $\bar{\theta}$, is defined as

$$\bar{\theta} = \begin{cases} \tan^{-1} \frac{\sum_{i=1}^n \sin \theta_i}{\sum_{i=1}^n \cos \theta_i} & \text{if } \sum_{i=1}^n \sin \theta_i > 0 \text{ and } \sum_{i=1}^n \cos \theta_i > 0 \\ \tan^{-1} \frac{\sum_{i=1}^n \sin \theta_i}{\sum_{i=1}^n \cos \theta_i} + \pi & \text{if } \sum_{i=1}^n \cos \theta_i < 0 \\ \tan^{-1} \frac{\sum_{i=1}^n \sin \theta_i}{\sum_{i=1}^n \cos \theta_i} + 2\pi & \text{if } \sum_{i=1}^n \sin \theta_i < 0 \text{ and } \sum_{i=1}^n \cos \theta_i > 0 \end{cases} \tag{5}$$

and the sample mean resultant length, r , is

$$r = \sqrt{\frac{1}{n} \left(\left(\sum_{i=1}^n \sin \theta_i \right)^2 + \left(\sum_{i=1}^n \cos \theta_i \right)^2 \right)}. \tag{6}$$

The circular moments of an SSB random variable with origin zero can be calculated using Theorem 1.

Theorem 1. Let $\theta \sim \mathcal{SSB}(0, \alpha, \beta)$. Then, for $p \in \mathbb{N}$,

$$\begin{aligned} E[\cos(p\theta)] &= \frac{1}{B(\alpha, \beta)} \sum_{j=0}^{\beta-1} (-1)^j \binom{\beta-1}{j} I_p(\alpha + j - 1) \\ &= \frac{1}{B(\alpha, \beta)} \sum_{j=0}^{\alpha-1} (-1)^j \binom{\alpha-1}{j} I_p(\beta + j - 1) \\ E[\sin(p\theta)] &= \frac{1}{B(\alpha, \beta)} \sum_{j=0}^{\beta-1} (-1)^j \binom{\beta-1}{j} J_p(\alpha + j - 1) \\ &= \frac{1}{B(\alpha, \beta)} \sum_{j=0}^{\alpha-1} (-1)^{j+1} \binom{\alpha-1}{j} J_p(\beta + j - 1) \end{aligned}$$

where $I_p(0) = J_p(0) = I_p(1) = 0$, $J_p(1) = -\frac{1}{2\pi p}$ and for $C > 1$,

$$\begin{aligned} I_p(C) &= \sum_{c=1}^{\lfloor \frac{C}{2} \rfloor} (-1)^{c-1} \frac{C!}{(C-2c+1)!} \frac{1}{(2\pi p)^{2c}} \\ J_p(C) &= \sum_{c=1}^{\lfloor \frac{C+1}{2} \rfloor} (-1)^c \frac{C!}{(C-2c+2)!} \frac{1}{(2\pi p)^{2c-1}}. \end{aligned}$$

‡

The circular moments of a general SSB distribution can be derived as a simple corollary of Theorem 1.

Corollary 1. Let $\theta \sim \mathcal{SSB}(\nu, \alpha, \beta)$. Then, for $p \in \mathbb{N}$,

$$\begin{aligned} E[\cos(p\theta)] &= \frac{1}{B(\alpha, \beta)} \left\{ \sum_{j=0}^{\beta-1} (-1)^j \binom{\beta-1}{j} (\cos(p\nu)I_p(\alpha + j - 1) - \sin(p\nu)J_p(\alpha + j - 1)) \right\} \\ &= \frac{1}{B(\alpha, \beta)} \left\{ \sum_{j=0}^{\alpha-1} (-1)^j \binom{\alpha-1}{j} (\cos(p\nu)I_p(\beta + j - 1) + \sin(p\nu)J_p(\beta + j - 1)) \right\} \\ E[\sin(p\theta)] &= \frac{1}{B(\alpha, \beta)} \left\{ \sum_{j=0}^{\beta-1} (-1)^j \binom{\beta-1}{j} (\sin(p\nu)I_p(\alpha + j - 1) + \cos(p\nu)J_p(\alpha + j - 1)) \right\} \\ &= \frac{1}{B(\alpha, \beta)} \left\{ \sum_{j=0}^{\alpha-1} (-1)^j \binom{\alpha-1}{j} (\sin(p\nu)I_p(\beta + j - 1) - \cos(p\nu)J_p(\beta + j - 1)) \right\} \end{aligned}$$

‡

Proofs of both results are given in the Appendix.

Exact results for the circular mean can be derived in the case where the SSB distribution is symmetric. In particular, when $\alpha = \beta = 1$, then it is clear that the SSB distribution reduces to the circular uniform distribution whose trigonometric moments are all zero. When $\alpha = \beta > 1$, then by symmetry, the mean circular direction is clearly $\mu = \text{mod}(\nu + \pi, 2\pi)$ and, the mean resultant length is

$$\rho = \frac{1}{B(\alpha, \alpha)} \left| \sum_{j=0}^{\alpha-1} (-1)^j \binom{\alpha-1}{j} I_p(\alpha + j - 1) \right|.$$

2.3 Mixtures of SSB distributions and Bernstein polynomials

The simple SSB model is not sufficiently flexible to model most types of circular data. Thus, from now on, we will consider mixtures of SSB distributions of form

$$f(\theta) = \sum_{j=1}^k w_j f(\theta | \nu_j, \alpha_j, \beta_j)$$

where $f(\cdot | \nu, \alpha, \beta)$ is the $\mathcal{SSB}(\nu, \alpha, \beta)$ density function as in (1) and $w_j > 0$ for $j = 1, \dots, k$ and $\sum_{j=1}^k w_j = 1$. We shall represent this distribution by writing $\theta \sim \mathcal{SSB}_k(\mathbf{w}, \boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ where $\mathbf{w} = (w_1, \dots, w_k)$, $\boldsymbol{\nu} = (\nu_1, \dots, \nu_k)$, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_k)$ and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)$.

The circular moments of the SSB mixture distribution can be derived directly from Corollary 1.

Corollary 2. Let $\theta \sim \mathcal{SSB}_k(\mathbf{w}, \boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\beta})$. Then, for $p \in \mathbb{N}$,

$$\boldsymbol{\mu}'_p = \sum_{j=1}^k w_j (E[\cos p\theta_j] + iE[\sin p\theta_j])$$

where $E[\cos p\theta_j]$ and $E[\sin p\theta_j]$ can be derived from Corollary 1.

‡

A particular case of SSB mixture density is the scaled, shifted, Bernstein polynomial density. For any variable, X , with support $[0, 1]$ and continuous distribution function $F_X(\cdot)$, then the Bernstein polynomial of order k is

defined to be

$$B_k(x) = \sum_{j=1}^k F_X\left(\frac{j}{k}\right) \binom{k}{j} x^j (1-x)^{k-j}$$

and it is well known that the Bernstein polynomial converges uniformly to F_X as $k \rightarrow \infty$. The associated Bernstein density function is

$$b_k(x) = \sum_{j=1}^k \left(F_X\left(\frac{j}{k}\right) - F_X\left(\frac{j-1}{k}\right) \right) f(x|j, k-j+1)$$

where $f(\cdot|\alpha, \beta)$ is a beta density function as in (2).

This has motivated the use of the Bernstein polynomial density

$$f_{BP}(x) = \sum_{j=1}^k w_j f(x|j, k-j+1)$$

as a semi-parametric model for samples of data with support $[0, 1]$. The good convergence properties of this model have been illustrated by e.g. Vitale (1975) and Kakizawa (2004) from a classical viewpoint and Petrone (1999a,b) and Petrone and Wassermann (2002) using a Bayesian approach. Clearly, rescaling onto $[0, 2\pi)$ and shifting the origin to ν , the Bernstein polynomial density of order k corresponds to $\mathcal{SSB}(\mathbf{w}, \nu \mathbf{1}, (1 : k), (k : 1))$ where $\mathbf{1}$ is a unit vector, $(1 : k) = (1, 2, \dots, k)$ and $(k : 1) = (k, k-1, \dots, 1)$.

3 Classical inference for the SSB mixture model

Assume now that we observe a sample of data, $\boldsymbol{\theta} = \{\theta_1, \dots, \theta_n\}$ to which we wish to fit a mixture of SSB distributions. Firstly, it is important to note that it is straightforward to calculate the maximum likelihood parameter estimates when the data are fitted by a single SSB distribution. For a given origin, ν , the maximum likelihood estimates of the beta parameters, α and β , may be calculated by first rescaling the data onto the data to the $[0, 1)$ interval and then calculating the unconstrained maximum likelihood estimates using, for example the `MATLAB` routine, `betafit`. Then the constrained maxima may be found by simply checking the likelihoods at the pairs of integer valued α and β around the unconstrained maxima. Finally, the global maxima may be estimated by using the same procedure over a grid of values of ν .

For fitting a k dimensional mixture distribution, then various approaches are possible. One possibility is to use the EM algorithm, see e.g. Dempster et

al (1977), McLachlan and Peel (2000) but we have found that this procedure is very sensitive to the choice of initial values and thus, we prefer to use an alternative procedure based on direct likelihood maximization.

Firstly, when $k = 2$, or $k = 3$, it is possible to directly maximize the likelihood function in terms of the component weights, \mathbf{w} , by searching over a grid of values of $\boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\beta}$ although for higher dimensional mixtures, this procedure is too computationally intensive. In these cases, we use an iterative algorithm which automatically finds a (local) maximum. Given that we wish to fit a mixture of $k > 3$ terms, we assume that we have calculated the maximum likelihood estimates for the $k - 1$ dimensional mixture and proceed by setting the initial values, $(\nu_i, \alpha_i, \beta_i)$ for $i = 1, \dots, k - 1$ equal to the maximum likelihood estimates for this model and then estimating all component weights and the parameters, $(\nu_k, \alpha_k, \beta_k)$ of the k 'th mixture component by direct likelihood maximization using a grid search procedure. Then, the parameters of components $2, \dots, k$ are fixed and the weights and $(\nu_1, \alpha_1, \beta_1)$ are re-estimated by direct likelihood maximization, the parameters of components, $1, 3, \dots, k$ are fixed and the weights and $(\nu_2, \alpha_2, \beta_2)$ are re-estimated and so on until convergence. We find in practice that this algorithm usually converges to the global maximum likelihood, due to the fact that the different beta components are well distinguished from each other due to the discreteness of the beta parameters. Thus, the main mixture terms are usually well identified in low dimensional mixtures and the effect of adding an extra component is essentially that of fitting a new component which well fits a small amount of the observed data.

Finally, in order to select the number of terms to include in the SSB mixture model, we use the Bayesian information criterion (BIC), see Schwarz (1978).

3.1 Examples

In this section, we consider two data sets that have been well analyzed in the circular data literature. Figure 1 shows a circular dot plot of the directions, in radians taken by 76 turtles after laying their eggs, taken from Stephens (1969). One can see that there is a clear bi-modality in these data due to the fact that most turtles return to the sea whereas a few turtles go in the approximately opposite direction towards the land.

Figure 2 gives the arrival times on the 24 hour clock of 254 patients at an intensive care unit taken from Cox and Lewis (1966). In this case, there

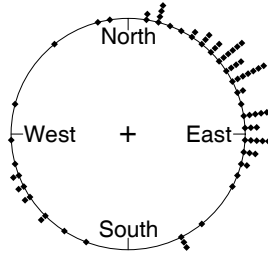


Figure 1: Directions taken by turtles after laying eggs

is no obvious grouping of the data apparent from the dot plot.

For each data set we fitted the (shifted) Bernstein polynomial density, using direct likelihood maximization to optimize the weights over a grid of possible origins and a mixture of von Mises distributions using the EM algorithm. Finally, the SSB mixture model was also fitted firstly with a common origin equal to zero, secondly with a common unknown origin, ν and finally with different origins for each mixture element.

Table 1 shows the log-likelihood and BIC for each of the models fitted to the turtles data. It can be seen that the optimal model according to the BIC is the full SSB mixture model although there is little difference between fitting this model, the common origin SSB mixture model and the von Mises mixture model.

Figure 3 shows a histogram of the turtles data with the predictive densities assuming the von Mises mixture model (dotted line), the common origin SSB mixture model (dashed line) and the full SSB mixture model (solid line) respectively. The fitted density for the Bernstein polynomial model was very similar to that of the common origin SSB mixture model, so this density is not shown.

For the turtles data, the estimated mean direction using the full SSB model is 1.12 radians which is very close to the sample mean direction of 1.11 radians.

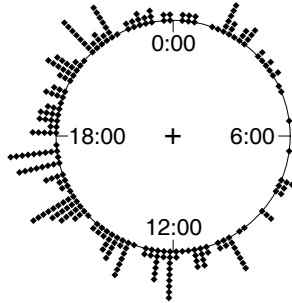


Figure 2: Arrival times of patients at an emergency ward

Table 1: Log likelihoods and BICs for the turtles data

Model	Origin	Mixture size	Log-likelihood	BIC
Bernstein poly.	1.96π	3	-116.65	246.29
von Mises	--	2	-108.62	238.89
SSB common origin	0.4π	2	-108.64	238.94
SSB full model	--	2	-106.13	238.24

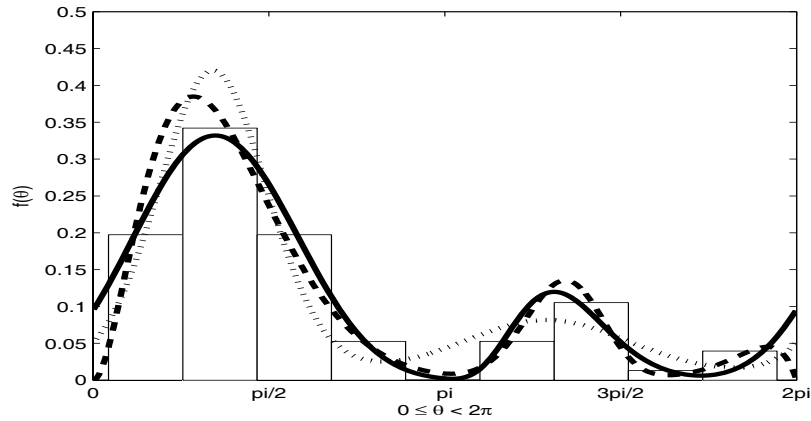


Figure 3: Histogram of the turtles data with fitted densities

Table 2 shows the log-likelihood and BIC for each of the models fitted to the emergencies data. As earlier, the optimal model according to the BIC criterion is the full SSB mixture model.

Table 2: Results for the emergencies data set

Model	Origin	Mixture size	Log-likelihood	BIC
Bernstein poly.	1.62π	5	-431.48	907.50
von Mises	--	2	-439.62	906.93
SSB common origin	1.56π	1	-434.95	887.01
SSB full model	--	2	-426.32	885.86

Finally, Figure 4 shows the fitted von Mises mixture (dotted line), common origin SSB (dashed line) and full SSB models (solid line). In this case, we can observe a jump in the density for the full model and for the common origin SSB model. Jumps can be observed whenever an SSB component, $\mathcal{SSB}(\nu, \alpha, \beta)$ has $\alpha = 1$ and $\beta > 1$ or $\alpha > 1$ and $\beta = 1$.

The estimated mean direction using the full SSB model is 4.49 radians (17.29 hours) whereas the sample mean is 4.612 radians (17.36 hours).

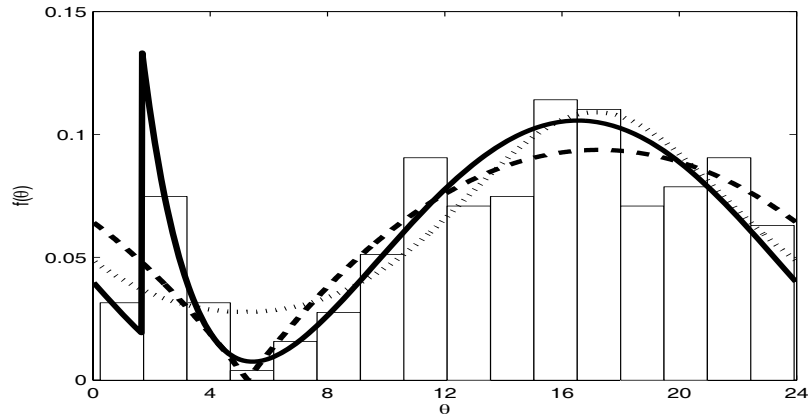


Figure 4: Histogram of the emergency times data with fitted SSB densities

4 Bayesian inference

In most real data examples we have analyzed, with the exception of the emergencies data, the fitted SSB mixture distributions from the classical algorithm include a uniform component. Thus, in this section, we shall consider an extended SSB mixture model including a uniform term, as well as k SSB elements. Bayesian inference for this model can be undertaken by extending ideas developed in Robert et al (2003) who consider a mixture of standard beta variables and a uniform term in the context of hypothesis testing.

Thus, the model we consider in this section is

$$f(\theta|\mathbf{w}, \boldsymbol{\nu}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = w_0 \frac{1}{2\pi} + \sum_{j=1}^k w_j f(\theta|\nu_j, \alpha_j, \beta_j)$$

where $\mathbf{w} = (w_0, \dots, w_k)$, $f(\theta|\nu, \alpha, \beta)$ is a standard SSB density and $w_j > 0$, for $j = 0, \dots, k$ and $\sum_{j=0}^k w_j = 1$.

In order to carry out Bayesian inference for this model, it is first necessary to define prior distributions for the model parameters. Here, we assume relatively uninformative priors:

$$\begin{aligned} \nu_j &\sim \mathcal{U}(0, 2\pi) \quad \text{a uniform distribution on } [0, 2\pi), \text{ for } j = 1, \dots, k \\ \alpha_j &\sim \mathcal{TG}(0.01), \quad \text{a geometric distribution truncated onto } \{1, 2, \dots, 50\} \text{ for } j = 1, \dots, k, \end{aligned}$$

$$\beta_j \sim \mathcal{TG}(0.01), \quad \text{a truncated geometric distribution as above}$$

$$\mathbf{w} \sim \mathcal{D}(\underbrace{1, \dots, 1}_{k+1}), \quad \text{a Dirichlet distribution.}$$

In order to ensure identifiability of the model, we assume that $\nu_1 < \nu_2 < \dots < \nu_k$.

Given a sample of data, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$, then Bayesian inference can be carried out by introducing the latent variables $\mathbf{Z} = (Z_1, \dots, Z_n)$ so that $P(Z_i = z | \mathbf{w}) = w_z$ for $z = 0, \dots, k$. Then Bayesian inference can be carried out using a standard, Metropolis within Gibbs sampler as in e.g. Robert et al (2003) as follows:

1. $t = 0$. Set initial values $\mathbf{w}^{(0)}, \boldsymbol{\nu}^{(0)}, \boldsymbol{\alpha}^{(0)}, \boldsymbol{\beta}^{(0)}$.
2. For $i = 1, \dots, n$ generate $z_i^{(t+1)}$ from $P(Z_i | \theta_i, \mathbf{w}^{(t)}, \boldsymbol{\nu}^{(t)}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{\beta}^{(t)})$.
3. Generate $\mathbf{w}^{(t+1)}$ from $f(\mathbf{w} | \mathbf{z}^{(t+1)})$.
4. For $j = 1, \dots, k$:
 - (a) Generate $\nu_j^{(t+1)}$ from $f(\nu_j | \boldsymbol{\theta}, \mathbf{z}^{(t+1)}, \alpha_j^{(t)}, \beta_j^{(t)})$.
 - (b) Generate $\alpha_j^{(t+1)}, \beta_j^{(t+1)}$ from $f(\alpha_j, \beta_j | \boldsymbol{\theta}, \mathbf{z}^{(t+1)}, \boldsymbol{\nu}^{(t+1)})$.
5. Reorder the SSB parameters if necessary so that $\nu_1^{(t+1)} < \dots < \nu_k^{(t+1)}$.
6. $t = t + 1$. Go to 2.

In this case, steps 2 and 3 are standard, as for any Gibbs sampling algorithm for mixtures. In step 4a, we use a Metropolis step, generating candidate values, $\tilde{\nu}_j$ from a uniform distribution on $\text{mod}(\nu_j^{(t)} \pm 0.05, 2\pi)$. In step 4b, we use a Metropolis Hastings proposal generating candidate values, $\tilde{\alpha}_j, \tilde{\beta}_j$ from independent, triangular distributions with modes $\alpha_j^{(t)}$ and $\beta_j^{(t)}$ respectively. The reordering in step 5 is included so that the identifiability restriction is complied with. A similar algorithm can be developed for fitting models where all SSB terms have a common origin.

In order to consider models containing an unknown number, k , of SSB mixture terms, then various approaches are possible. One method is to use reversible jump (Richardson and Green 1997) or birth-death (Stephens 2000)

MCMC approaches in order to mix over models of different dimensions. An alternative approach, which we follow here is to use a Bayesian model choice criterion to select a model of appropriate dimension.

A standard tool for Bayesian model choice is the deviance information criterion or DIC, see Spiegelhalter et al (2002). However, it is well known that the basic DIC is inappropriate in the context of mixture models, see e.g. Celeux et al (2006). Celeux et al (2006) recommend a number of variants of the DIC and, in particular, here, we use their DIC_3 criterion which is defined as follows.

Assume that a circular variable, θ follows a parametric model \mathcal{M} with parameters ϕ such that $\theta|\mathcal{M}, \phi \sim f(\cdot|\mathcal{M}, \phi)$. Then given a sample of data, $\boldsymbol{\theta} = \theta_1, \dots, \theta_n$, the DIC_3 criterion for model selection is

$$DIC_3 = -4E_{\phi}[\log f(\boldsymbol{\theta}|\phi)|\boldsymbol{\theta}] + 2 \log \prod_{i=1}^n E_{\phi}[f(\theta_i|\phi, \boldsymbol{\theta})].$$

As with the BIC, small values of DIC_3 indicate a better model. As noted by Celeux et al (2006), this criterion has the advantage of being very straightforward to calculate given the output from a Gibbs sampler. The full properties of this criterion are also examined in more detail by Celeux et al (2006).

4.1 Example

In this section we consider a data set of crimes perpetrated in Chicago on May 11th 2007, obtained from www.chicagocrime.org. Figure 5 gives the 24 hour clock times of the 1240 crimes committed on this day.

For this data set we fitted the most general algorithm with the circular uniform distribution plus up to 5 SSB distribution components and also considered models with a uniform component and SSB terms with a common origin. The optimal model according to the DIC_3 criterion was the fixed origin model with 4 SSB components ($DIC_3 = 4575$) whereas the optimal full model contains five SSB components with $DIC_3 = 4582$. Figure 4.1 shows the fits of the optimal full model (solid line) and the optimal, common origin model (dashed line).

The fits of both predictive densities to the data appear are quite similar.

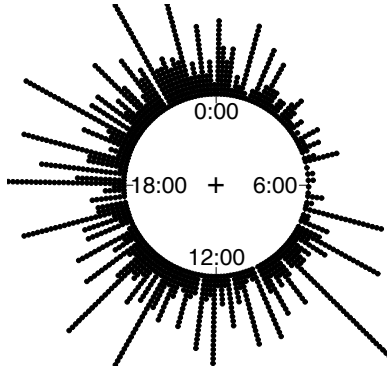


Figure 5: Perpetrated times of crimes on may 11th at Chicago

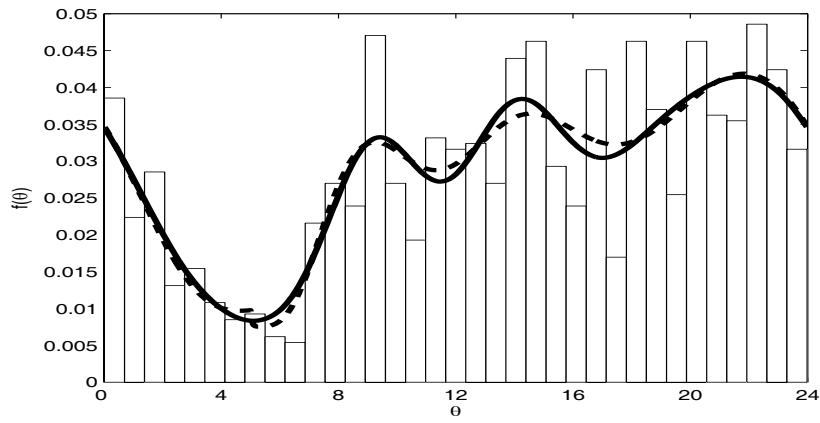


Figure 6: Histogram of the crime data with the predictive densities

5 Conclusions and extensions

In this article, we have developed a new, semi-parametric model for circular data which appears to compare favourably with the well known, von Mises mixture model. A number of extensions are possible.

Firstly, we have here used the BIC as a classical model choice criterion. It would also be possible to consider other criteria such as the Akaike (1974) information criterion (AIC). Our practical experience shows that the use of the AIC generally leads to overfitting for the Bernstein polynomial models and that the SSB mixture models still perform better than both the von Mises mixture and Bernstein polynomial models under this criterion. Alternative criteria such as the corrected AIC could also be considered. Also from a Bayesian viewpoint, alternative model selection criteria such as those proposed by Celeux et al (2006) could be considered. In practice, we have found that the models selected using these criteria are usually the same as for DIC_3 used here.

Secondly, we have considered Bayesian inference for a mixture of a uniform density and various SSB distributions of fixed size. It is also possible to consider the simple SSB mixture model without the assumption that a uniform component is always present. In practice, this seems to introduce little advantage as in the great majority of data sets we have examined, a uniform component is virtually always fitted. Furthermore, this introduces some extra technical restrictions in the setting up of the MCMC algorithm due to the necessity of maintaining the identifiability of the fitted model.

Thirdly, it is straightforward to implement Bayesian inference for SSB mixtures of random dimension via reversible jump (Green 1995, Richardson and Green 1997) or birth-death (Stephens 2000) MCMC techniques. One advantage of this approach is that this avoids jumps in the predictive density due to the fact that this is mixed over the parameter distribution.

Fourthly, it would be interesting to examine the possible increase in efficiency from using a mixture of SSB distributions with continuous beta parameters instead of restricting the parameters to be integers. Finally, from a Bayesian standpoint, instead of using a finite mixture of SSB distributions, we might consider using continuous, Dirichlet process mixtures of SSB distributions, along the lines of Kottas (2006).

References

- [1] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.
- [2] Babu, G.J., Canty, A.J. and Chaubey, Y.P. (2002). Application of Bernstein polynomials for smooth estimation of a distribution and density function. *Journal of Statistical Planning and Inference*, **105**, 377–392.
- [3] Celeux, G., Forbes, F., Robert, C. and Titterton, M. (2006) Deviance Information Criteria for missing data models. *Bayesian analysis*, **1**, 651–674.
- [4] Cox, D.R. and Lewis, P.A.W. (1966). *The Statistical Analysis of Series of Events*. Methuen: London.
- [5] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B*, **39**, 1–38.
- [6] Evans, M., Hastings, N. and Peacock, B. (2000). von Mises Distribution. In *Statistical Distributions*, 3rd ed., Wiley: New York, pp. 189–191.
- [7] Fernández-Durán, J.J. (2004). Circular distributions based on nonnegative trigonometric sums. *Biometrics*, **60**, 499–503.
- [8] Ferreira, J.T.A.S., Juarez, M.A. and Steel, M.F.J. (2005). Directional Log-spline Distributions. *Working Paper*, **05-17**, Centre for Research in Statistical Methodology, Warwick University.
- [9] Green, P.J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 711–732.
- [10] Jammalamadaka, S.R. and Kozubowski T.J. (2003). A new family of circular models: The wrapped Laplace distributions, *Advances and Applications in Statistics* 3(1), 77-103, 2003.
- [11] Kakizawa, Y. (2004). Bernstein polynomial probability density estimation. *Journal of Nonparametric Statistics*, **16**, 709–729.

- [12] Kottas, A. (2006). Dirichlet process mixtures of beta distributions, with applications to density and intensity estimation. *UCSC Department of Applied Math and Statistics Technical Reports*, **ams2006-12**.
- [13] Fisher, N.I. (1989). Smoothing a sample of circular data. *Journal of Structural Geology*, **11**, 775–778.
- [14] Mardia, K.V. and Jupp, P. E. (1999). *Directional Statistics*, Wiley: Chichester.
- [15] McLachlan, G. and Peel, (2000). *Finite Mixture Models*. Wiley: New York.
- [16] Mooney A., Helms P.J. and Jolliffe I.T. (2003). Fitting mixtures of von Mises distributions: a case study involving sudden infant death syndrome. *Computational Statistics and Data Analysis*, **41**, 505–513.
- [17] Petrone, S. (1999a). Random Bernstein polynomials. *Scandinavian Journal of Statistics*, **26**, 373–393.
- [18] Petrone, S. (1999b). Bayesian density estimation using random Bernstein polynomials. *Canadian Journal of Statistics*, **27**, 105–126.
- [19] Petrone, S. and Wassermann, L. (2002). Consistency of Bernstein polynomial posteriors. *Journal of the Royal Statistical Society Series B*, **64**, 79–100.
- [20] Richardson, S. and Green, P. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society Series B*, **59**, 731–792.
- [21] Robert, C. P. y Rousseau, J. (2003) A mixture approach to bayesian goodness of fit. *Technical report*, **2002-9**, Universit Paris Dauphine.
- [22] Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.
- [23] Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion), *Journal of the Royal Statistical Society Series B*, **64**, 583–616.

- [24] Stephens, M.A. (1969). Techniques for directional data. *Technical Report*, **150**, Department of Statistics, Stanford University.
- [25] Stephens, M. (2000). Bayesian analysis of mixture models with an unknown number of components – an alternative to reversible jump methods. *Annals of Statistics*, **28**, 40–74.
- [26] Vitale, R.A. (1975). A Bernstein polynomial approach to density estimation. In *Statistical inference and related topics* (ed. Madan Lal Puri), Vol 2, 87–100. Academic Press, New York.

Appendix: Proofs of theorems

Proof of Theorem 1

First note that if $\theta \sim \mathcal{SSB}(0, \alpha, \beta)$ then

$$\begin{aligned}
E[\cos(p\theta)] &= E[\cos(2\pi pX)] \quad \text{where } X \sim \mathcal{B}(\alpha, \beta) \\
&= \int_0^1 \cos(2\pi px) \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} dx \\
&= \frac{1}{B(\alpha, \beta)} \int_0^1 \cos(2\pi px) \sum_{j=0}^{\beta-1} \binom{\beta-1}{j} x^{\alpha-1+j} dx \quad \text{assuming } \beta \in \mathbb{N} \\
&= \frac{1}{B(\alpha, \beta)} \sum_{j=0}^{\beta-1} (-1)^j \binom{\beta-1}{j} I_p(\alpha+j-1) \quad \text{where} \\
I_p(C) &= \int_0^1 \cos(2\pi px) x^C dx \quad \text{and similarly,} \\
E[\sin(p\theta)] &= E[\sin(2\pi pX)] \\
&= \frac{1}{B(\alpha, \beta)} \sum_{j=0}^{\beta-1} (-1)^j \binom{\beta-1}{j} J_p(\alpha+j-1) \quad \text{where} \\
J_p(C) &= \int_0^1 \sin(2\pi px) x^C dx \quad \text{Also} \\
E[\cos(p\theta)] &= E[\cos(2\pi p(1-Y))] \quad \text{where } Y \sim \mathcal{B}(\beta, \alpha) \\
&= E[\cos(2\pi pY)] \\
&= \frac{1}{B(\alpha, \beta)} \sum_{j=0}^{\alpha-1} (-1)^j \binom{\alpha-1}{j} I_p(\beta+j-1) \\
E[\sin(p\theta)] &= E[\sin(2\pi p(1-Y))]
\end{aligned}$$

$$\begin{aligned}
&= -E[\sin(2\pi pY)] \\
&= \frac{1}{B(\alpha, \beta)} \sum_{j=0}^{\alpha+j-1} (-1)^{j+1} \binom{\alpha-1}{j} J_p(\beta + j - 1).
\end{aligned}$$

Now observe that

$$\begin{aligned}
I_p(0) &= \int_0^1 \cos(2\pi px) dx = 0 \\
J_p(0) &= \int_0^1 \sin(2\pi px) dx = 0 \\
I_p(1) &= \int_0^1 x \cos(2\pi px) dx = \frac{1}{2\pi p} [x \sin(2\pi px)]_0^1 - \frac{1}{2\pi p} \int_0^1 \sin(2\pi px) dx = 0 \\
J_p(1) &= \int_0^1 x \sin(2\pi px) dx = -\frac{1}{2\pi p} [x \cos(2\pi px)]_0^1 + \frac{1}{2\pi p} \int_0^1 \cos(2\pi px) dx = -\frac{1}{2\pi p}
\end{aligned}$$

Now consider $I_p(C)$. For $C \geq 2$,

$$\begin{aligned}
I_p(C) &= \int_0^1 x^C \cos(2\pi px) dx \\
&= \frac{1}{2\pi p} [x^C \sin(2\pi px)]_0^1 - \frac{C}{2\pi p} \int_0^1 x^{C-1} \sin(2\pi px) dx \\
&= -\frac{C}{2\pi p} \int_0^1 x^{C-1} \sin(2\pi px) dx \\
&= \frac{C}{(2\pi p)^2} [x^{C-1} \cos(2\pi px)]_0^1 - \frac{C(C-1)}{(2\pi p)^2} \int_0^1 x^{C-2} \cos(2\pi px) dx \\
&= \frac{C}{(2\pi p)^2} - \frac{C(C-1)}{(2\pi p)^2} I_p(C-2) \tag{7}
\end{aligned}$$

and therefore $I_p(2) = \frac{2}{(2\pi p)^2}$ and $I_p(3) = \frac{3}{(2\pi p)^2}$ which satisfy the formula given in Theorem 1. Assume now that the formula is valid for $c = 2, \dots, C$. Then

$$\begin{aligned}
I_p(C+2) &= \frac{C+2}{(2\pi p)^2} - \frac{(C+2)(C+1)}{(2\pi p)^2} I_p(C) \quad \text{from (7)} \\
&= \frac{C+2}{(2\pi p)^2} - \sum_{c=1}^{\lfloor \frac{C}{2} \rfloor} (-1)^{c-1} \frac{C!}{(C-2c+1)!} \frac{1}{(2\pi p)^{2c}} \quad \text{from the induction assumption} \\
&= \frac{C+2}{(2\pi p)^2} + \sum_{c=1}^{\lfloor \frac{C}{2} \rfloor} (-1)^{c+1-1} \frac{(C+2)!}{(C+2-2(c+1)+1)!} \frac{1}{(2\pi p)^{2(c+1)}}
\end{aligned}$$

$$\begin{aligned}
&= \frac{C+2}{(2\pi p)^2} + \sum_{c=2}^{\lfloor \frac{C+2}{2} \rfloor} (-1)^{c-1} \frac{(C+2)!}{(C+2-2c+1)!} \frac{1}{(2\pi p)^{2c}} \\
&= \sum_{c=1}^{\lfloor \frac{C+2}{2} \rfloor} (-1)^{c-1} \frac{(C+2)!}{(C+2-2c+1)!} \frac{1}{(2\pi p)^{2c}}
\end{aligned}$$

which demonstrates the formula for $I_p(C)$.

Equally, we have the recurrence relation

$$J_p(C) = -\frac{1}{2\pi p} - \frac{C(C-1)}{(2\pi p)^2} J_p(C-2) \quad (8)$$

which implies that $J_p(2) = -\frac{1}{2\pi p}$ and $J_p(3) = -\frac{1}{2\pi p} + \frac{3!}{(2\pi p)^3}$ which satisfy the formula of Theorem 1. Assuming the formula is valid for $c = 2, \dots, C$ then

$$\begin{aligned}
J_p(C+2) &= -\frac{1}{2\pi p} - \frac{(C+2)(C+1)}{(2\pi p)^2} J_p(C) \quad \text{from (8)} \\
&= -\frac{1}{2\pi p} - \frac{(C+2)(C+1)}{(2\pi p)^2} \sum_{c=1}^{\lfloor \frac{C+1}{2} \rfloor} (-1)^c \frac{C!}{(C-2c+2)!} \frac{1}{(2\pi p)^{2c-1}} \\
&\quad \text{from the induction assumption} \\
&= -\frac{1}{2\pi p} - \sum_{c=1}^{\lfloor \frac{C+1}{2} \rfloor} \frac{(C+2)!}{(C-2c+2)!} \frac{1}{(2\pi p)^{2c+1}} \\
&= \sum_{c=1}^{\lfloor \frac{C+2+1}{2} \rfloor} (-1)^c \frac{(C+2)!}{(C+2-2c+2)!} \frac{1}{(2\pi p)^{2c-1}}
\end{aligned}$$

which demonstrates the formula for $J_p(C)$ and proves the theorem. \ddagger

Proof of Corollary 1

Let $\theta \sim \mathcal{SSB}(\nu, \alpha, \beta)$. Then:

$$\begin{aligned}
E[\cos(p\theta)] &= E[\cos(\text{mod}(p\nu + p\theta_0, 2\pi))] \quad \text{where } \theta_0 \sim \mathcal{SSB}(0, \alpha, \beta) \\
&= E[\cos(p\nu + p\theta_0)] \\
&= \cos p\nu E[\cos p\theta_0] - \sin p\nu E[\sin p\theta_0] \\
&= \frac{1}{B(\alpha, \beta)} \left\{ \sum_{j=0}^{\beta-1} (-1)^j \binom{\beta-1}{j} (\cos(p\nu) I_p(\alpha + j - 1) - \sin(p\nu) J_p(\alpha + j - 1)) \right\}.
\end{aligned}$$

The other expressions can be derived similarly by using the expansions of $\sin(A+B)$ and $\cos(A+B)$ and the fact that if $X \sim \mathcal{B}(\alpha, \beta)$, then $Y = 1 - X \sim \mathcal{B}(\beta, \alpha)$. \ddagger