Density estimation of circular data with Bernstein polynomials

J.A. Carnicero *, M.P. Wiper† and M.C. Ausín‡

Abstract

This paper introduces a new, non-parametric approach to the modeling of circular data, based on the use of Bernstein polynomial densities. The model generalizes the standard Bernstein polynomial model to account for the specific characteristics of circular data. In particular, it is shown that the trigonometric moments of the proposed circular Bernstein polynomial distribution can all be derived in closed form. Secondly, we introduce an approach to circular Bernstein polynomial density estimation given a sample of data and examine the properties of this estimator. Finally our method is illustrated with a simulation study and a real data example.

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 $0 \le \theta < 2\pi$ radians, which represent the positive angle of rotation from some arbitrarily chosen origin, $\theta = 0$.

A number of parametric models for circular data have been developed using a variety of techniques, see e.g. Mardia and Jupp (1999) for a full review. However, most parametric

---
*Hospital Universitario de Getafe, Madrid, Spain.

†Departamento de Estadística, Universidad Carlos III de Madrid
Email : michael.wiper@uc3m.es Corresponding author.

‡Departamento de Estadística, Universidad Carlos III de Madrid

models developed for circular data are unimodal and symmetric but, in many cases, both multimodal and asymmetric data may be encountered. In such cases, semi-parametric or non-parametric approaches might be preferred. Semi-parametric approaches based on trigonometric sums and mixtures of von Mises or circular normal distributions have been introduced in Fernández-Durán (2004) and Mooney et al. (2003) respectively. Non-parametric, kernel based approaches have also been studied in e.g. Hall et al. (1987), Bai et al. (1988), Fisher (1989), Klemelä (2000) and Taylor (2008).

In this paper, we introduce an alternative non-parametric approach based on the use of Bernstein polynomials. It is well known that the Bernstein polynomial is a useful tool for interpolating functions defined on a closed interval and can therefore be used to approximate a density function on such an interval. Furthermore, Bernstein polynomials are an example of a polynomial approximation with a simple interpretation in terms of probability and possess good shape preservation properties, see e.g. Carnicer and Peña (1993). Bernstein polynomials have been proposed as density estimators for variables with finite support in a number of articles, see e.g. Vitale (1975), Petrone (1999a,b), Petrone and Wassermann (2002), Babu et al. (2002) and Kakizawa (2004).

Given that computations for kernel density estimations using standard circular densities such as the von Mises are often highly involved, this means that the use of Bernstein polynomials, which are computationally straightforward is attractive in this context. However, a problem with generalizing standard Bernstein polynomial density approaches to circular data is that these can lead to fitted densities with discontinuities, which seems unreasonable for continuous, circular data. The main objective of this paper is to show that it is possible to adapt linear Bernstein polynomial estimators to ensure that circular density estimators are produced. One advantage of the proposed Bernstein polynomial density approach is then that circular moments can be calculated in a closed form.

The article is organized as follows. Firstly, in Section 2, we define the Bernstein polynomial density approximation and show how this can be extended to circular variables. Secondly, in Section 3, we demonstrate how to calculate the circular moments of a Bernstein polynomial density. Then, in Section 4, we introduce a circular Bernstein polynomial density estimation procedure based on the linear estimator of Vitale (1975) and examine the convergence properties of the proposed estimator in comparison with the Vitale estimator. Our results are illustrated with a simulation study and a real data set in Section 5. Finally, we draw conclusions and consider various extensions in Section 6.

2. Bernstein polynomial approximations for circular variables

Let F be a continuous function on $[0, 1]$. Then, the Bernstein polynomial of order k for F is defined to be

$$(2.1) \qquad B_k(x) = \sum_{j=0}^{k} F\left(\frac{j}{k}\right) \left(\begin{array}{c} k \\ j \end{array} \right) x^j (1-x)^{k-j} \quad \text{for } 0 \le x \le 1 \text{ and } k \in \mathbb{N}.$$

In particular, it is well known that $B_k(x)$ converges uniformly to $F_X(x)$ as k goes to infinity, see e.g. Lorentz (1997).

In particular, if X is a strictly continuous random variable with support $[0, 1]$ and cumulative distribution function $F = F_X(\cdot)$, then (2.1) is also a distribution function. Then, the derivative of this function is the Bernstein density function which is defined as:

$$(2.2) \qquad b_k(x) = \sum_{j=1}^{k} \left(F_X\left(\frac{j}{k}\right) - F_X\left(\frac{j-1}{k}\right) \right) \beta(x \mid j, k-j+1),$$

where $\beta(\cdot \mid a, b)$ is a beta density function:

$$(2.3) \qquad \beta(x \mid a, b) = \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1}$$

and $B(a,b) = (a-1)!(b-1)!/(a+b-1)!$, for $a, b \in \mathbb{N}$, is the beta function.

Derivatives of Bernstein polynomials of a given function on the unit interval are also well known to converge uniformly to the function derivatives, see e.g. Phillips (2003), Chapter 7, and, in particular, from Butzer (1955), it is easy to show that, as long as f_X is twice differentiable over the unit interval, then

$$(2.4) \qquad b_k(x) - f_X(x) = \frac{1}{2k} \left[(1-2x)f_X'(x) + x(1-x)f_X''(x) \right] + o\left(\frac{1}{k} \right).$$

Floater (1995) also gives an error bound for the Bernstein polynomial approximation as

$$(2.5) \qquad |b_k(x) - f(x)| \leq \frac{1}{2k} \left[|1-2x| \|f_X'\| + x(1-x)\|f_X''\| \right]$$

where $\| \cdot \|$ is the max norm on $[0,1]$.

Clearly it is straightforward to extend the use of Bernstein polynomials to densities defined on any closed interval, e.g. $[0, 2\pi]$. However, in order to define a distribution on the circle, it is first necessary to formally define the density function of a circular random variable.

The density function, $f_\Theta(\theta)$ of a continuous, circular random variable, Θ, is a non-negative, continuous function such that

$$f_\Theta(\theta + 2\pi r) = f_\Theta(\theta), \quad \text{for } \theta \in \mathbb{R} \text{ and any } r \in \mathbb{Z}$$

and $\int_0^{2\pi} f_\Theta(\theta)\, d\theta = 1$.

In order to define a cumulative distribution function of a circular random variable, it is necessary to establish an origin, say $\nu \in [0, 2\pi)$. Given this origin, the cumulative distribution function is:

$$F_\Theta^\nu(\theta) = \int_\nu^{\nu+\theta} f_\Theta(u)\, du, \quad \text{for } 0 \leq \theta < 2\pi.$$

With respect to the origin ν, we shall also define the (shifted) density function

$$f_\Theta^\nu(\theta) \stackrel{\text{def}}{=} f_\Theta(\nu + \theta) \quad \text{where} \quad f_\Theta^0(\theta) = f_\Theta(\theta).$$

If we wish to consider the Bernstein polynomial density approximation of order k, with respect to the origin ν, that is

$$(2.6) \qquad b_k^\nu(\theta) = \frac{1}{2\pi} \sum_{j=1}^{k} \left(F_\Theta^\nu\left(\frac{2\pi j}{k} \right) - F_\Theta^\nu\left(\frac{2\pi(j-1)}{k} \right) \right) \beta\left(\frac{\theta}{2\pi} \,\middle|\, j, k-j+1 \right)$$

for this to be a strictly continuous, circular density then it is necessary that,

$$(2.7) \qquad F_\Theta^\nu\left(\frac{2\pi}{k} \right) = 1 - F_\Theta^\nu\left(\frac{2\pi(k-1)}{k} \right),$$

where $F_\Theta^\nu(\theta)$ represents the cumulative distribution function with respect to the origin, ν. The following theorem guarantees the existence of at least one origin satisfying (2.7).

2.1. Theorem. *Let f_Θ be a density function for a continuous, circular random variable, Θ. Then, for any $k \in \mathbb{N}$, there exists at least one point, $\nu = \nu_k \in [0, 2\pi)$ such that,*

$$\int_\nu^{\nu + \frac{2\pi}{k}} f_\Theta(\theta)\, d\theta = \int_{\nu - \frac{2\pi}{k}}^\nu f_\Theta(\theta)\, d\theta.$$

Proof. Define $G(\nu) = \int_{\nu}^{\nu+\frac{2\pi}{k}} f_\Theta(\theta)\, d\theta - \int_{\nu-\frac{2\pi}{k}}^{\nu} f_\Theta(\theta)\, d\theta$. If there exist two points, $0 \leq \nu_1 \neq \nu_2 < 2\pi$ such that $G(\nu_1) \leq 0$ and $G(\nu_2) \geq 0$, then by Bolzano's intermediate value theorem, there exists at least one point, $0 \leq \nu_0 < 2\pi$ such that $G(\nu_0) = 0$. Otherwise, suppose that $G(\nu)$ is always positive. Then, we have

$$\int_{\nu-\frac{2\pi}{k}}^{\nu} f_\Theta(\theta)\, d\theta < \int_{\nu}^{\nu+\frac{2\pi}{k}} f_\Theta(\theta)\, d\theta < \int_{\nu+2\pi-\frac{2\pi}{k}}^{\nu+2\pi} f_\Theta(\theta)\, d\theta$$

which is impossible, as, due to the periodicity of f, we have that

$$\int_{\nu-\frac{2\pi}{k}}^{\nu} f_\Theta(\theta)\, d\theta = \int_{\nu+2\pi-\frac{2\pi}{k}}^{\nu+2\pi} f_\Theta(\theta)\, d\theta.$$

Similarly, G cannot always be negative and so the theorem is proved. $\qquad\square$

For many distributions, the set of origins, which satisfy the conditions of Theorem 2.1 may contain more than one, or even an uncountable number of elements. Define $\boldsymbol{\nu}_k$ to be this set of origins. Then, for example, a density which is symmetric about π has $\boldsymbol{\nu}_k \supseteq \{0, \pi\}$ and for a uniform density, $\boldsymbol{\nu}_k = \{\nu : \nu \in [0, 2\pi)\}$.

2.1. Asymptotic properties. It is important to examine how the convergence properties of the Bernstein polynomial approximation as $k \to \infty$ are influenced by the dependence of the origin on k. From now on, we shall assume throughout that $f_\Theta(\theta)$ has finite second derivative for all $\theta \in \mathbb{R}$. Then, applying (2.5) to the circular case and considering a fixed point, θ, we have that for any origin, ν_k,

$$
\begin{aligned}
|b_k^{\nu_k}(\mathrm{mod}(\theta - \nu_k, 2\pi)) - f_\Theta(\theta)| &\leq \frac{1}{4\pi k}\left[|1 - 2x_\theta^{\nu_k}|\max_\vartheta |f_\Theta'(\vartheta)| + \right. \\
&\qquad \left. x_\theta^{\nu_k}(1 - x_\theta^{\nu_k})\max_\vartheta |f_\Theta''(\vartheta)|\right] \\
&\leq \frac{1}{4\pi k}\left[\max_\vartheta |f_\Theta'(\vartheta)| + \frac{1}{4}\max_\vartheta |f_\Theta''(\vartheta)|\right]
\end{aligned}
$$

(2.8)

for all θ, where $\mathrm{mod}(a, b)$ represents the value of a modulo b and, for brevity of notation, we write

$$(2.9) \qquad x_\theta^{\nu_k} \stackrel{\mathrm{def}}{=} \mathrm{mod}\left(\frac{\theta - \nu_k}{2\pi}, 1\right).$$

Thus, it is clear that however the origins, ν_k, are chosen, then the Bernstein polynomial approximation still converges uniformly to f_Θ with the same order of convergence as the linear Bernstein polynomial although the exact error will depend on the procedure for choosing the origins.

It is interesting to consider the asymptotic behaviour of the set of origins, $\boldsymbol{\nu}_k$, satisfying Theorem 2.1. In particular, it is straightforward to see that as $k \to \infty$, then the origins cluster around the turning points of the density f_Θ.

2.2. Theorem. *Let $\nu_k \in \boldsymbol{\nu}_k$. Then, there exists at least one point, $\nu \in \left[\nu_k - \frac{2\pi}{k}, \nu_k + \frac{2\pi}{k}\right]$ such that $f_\Theta'(\nu) = 0$.*

Proof. Suppose that $f_\Theta'(\nu) > 0$ for all ν in $\left[\nu_k - \frac{2\pi}{k}, \nu_k + \frac{2\pi}{k}\right]$. Then this implies that

$$G(\nu_k) = \int_{\nu_k}^{\nu_k + \frac{2\pi}{k}} f_\Theta(\theta)\, d\theta - \int_{\nu_k - \frac{2\pi}{k}}^{\nu_k} f_\Theta(\theta)\, d\theta > 0$$

which is impossible from Theorem 2.1 and similarly if $f_\Theta'(\nu)$ is negative over the whole interval which is a contradiction. $\qquad\square$

2.2. An alternative approximation. The results of the previous section suggest that an alternative Bernstein polynomial approximation based on using a fixed origin, ν, such that $f'_\Theta(\nu) = 0$ can also be considered. As it may be that $\nu \notin \boldsymbol{\nu}_k$, for some k, then the standard Bernstein polynomial approximation needs to be corrected to provide a circular approximation. Thus, writing

$$b_k^\nu(\theta) = \sum_{j=1}^k \omega_{kj}^\nu \beta\left(\frac{\theta}{2\pi}\bigg| j, k-j+1\right)$$

where $\omega_{kj}^\nu = F^\nu\left(\frac{2\pi j}{k}\right) - F^\nu\left(\frac{2\pi(j-1)}{k}\right)$ for $j = 1, \ldots, k$, then the corrected approximation is given by

$$c_k^\nu(\theta) = \sum_{j=1}^k \varpi_{kj}^\nu \beta\left(\frac{\theta}{2\pi}\bigg| j, k-j+1\right).$$

where $\varpi_{k1}^\nu = \varpi_{kk}^\nu = (\omega_{k1}^\nu + \omega_{kk}^\nu)/2$ and $\varpi_{kj}^\nu = \omega_{kj}^\nu$ for $j = 2, \ldots, k-1$. It is easy to show that this corrected approximation preserves the asymptotic properties of the usual Bernstein polynomial.

2.3. Theorem. *Assume that $\nu \in [0, 2\pi)$ satisfies $f'_\Theta(\nu) = 0$. Then*

$$|c_k^\nu(\theta) - b_k^\nu(\theta)| = o\left(\frac{1}{k}\right).$$

Proof. We have

$$\begin{aligned}
c_k^\nu(\theta) - b_k^\nu(\theta) &= \sum_{j=1}^k \left(\varpi_{kj}^\nu - \omega_{kj}^\nu\right) \beta\left(\frac{\theta}{2\pi}\bigg| j, k-j+1\right) \\
&= (\varpi_{k1}^\nu - \omega_{k1}^\nu) \beta\left(\frac{\theta}{2\pi}\bigg| 1, k\right) + (\varpi_{kk}^\nu - \omega_{kk}^\nu) \beta\left(\frac{\theta}{2\pi}\bigg| k, 1\right) \\
&= \frac{\omega_{kk}^\nu - \omega_{k1}^\nu}{2}\left(\beta\left(\frac{\theta}{2\pi}\bigg| 1, k\right) - \beta\left(\frac{\theta}{2\pi}\bigg| k, 1\right)\right)
\end{aligned}$$

so that

$$|c_k^\nu(\theta) - b_k^\nu(\theta)| \leq k\left|\frac{\omega_{kk}^\nu - \omega_{k1}^\nu}{2}\right| = \frac{1}{2}\left|\frac{\omega_{kk}^\nu - \omega_{k1}^\nu}{1/k}\right|.$$

Define $G_1(\vartheta) = \int_{\nu-\vartheta}^\nu f_\Theta(\theta)\, d\theta = \omega_{kk}^\nu$ and $G_2(\vartheta) = \int_\nu^{\nu+\vartheta} f_\Theta(\theta)\, d\theta = \omega_{k1}^\nu$ for $\vartheta > 0$. Observe that from Taylor's theorem, we have that

$$G_i(\vartheta) = G_i(0) + \vartheta G_i'(0) + \vartheta^2 G_i''(0) + o\left(\vartheta^2\right)$$

for $i = 1, 2$. Noting that $G_i(0) = 0$, $G_i'(0) = f_\Theta^\nu(0)$ and, because $f'_\Theta(\nu) = 0$, we have $G_i''(0) = 0$, for $i = 1, 2$, then we find

$$\frac{\omega_{kk}^\nu - \omega_{k1}^\nu}{1/k} = \frac{o\left(\frac{1}{k^2}\right)}{1/k} = o\left(\frac{1}{k}\right).$$

$\square$

2.3. Trigonometric moments. For a circular random variable, Θ, the p'th trigono-metric moment is defined to be

$$\boldsymbol{\mu}'_p \quad = \quad E\left[e^{ip\Theta}\right] = E[\cos p\Theta] + iE[\sin p\Theta] \quad \text{where } i = \sqrt{-1}$$

$$\overset{\text{def}}{=} \quad \rho_p \left(\cos \mu'_p + i \sin \mu'_p\right) \quad \text{where}$$

$$(2.10) \quad \rho_p \quad = \quad \sqrt{E[\cos p\Theta]^2 + E[\sin p\Theta]^2} \quad \text{and}$$

$$(2.11) \quad \mu'_p \quad = \quad \begin{cases} \tan^{-1} \frac{E[\sin p\Theta]}{E[\cos p\Theta]} & \text{if} \quad E[\sin p\Theta] > 0 \text{ and } E[\cos p\Theta] > 0 \\ \tan^{-1} \frac{E[\sin p\Theta]}{E[\cos p\Theta]} + \pi & \text{if} \quad E[\cos p\Theta] < 0 \\ \tan^{-1} \frac{E[\sin p\Theta]}{E[\cos p\Theta]} + 2\pi & \text{if} \quad E[\sin p\Theta] < 0 \text{ and } E[\cos p\Theta] > 0 \end{cases}$$

for $p = 1, 2, \ldots$. 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.

For some variables, exact calculation of the circular moments may be difficult and then, an alternative may be to use the circular Bernstein polynomial approximation of sufficiently high order to approximate these moments. The trigonometric moments of a circular variable with a Bernstein polynomial density can be easily derived from the following theorem.

2.4. Theorem. *Assume that Θ is a circular random variable with density $f_\Theta(\theta) = \sum_{j=1}^{k} \varpi_{kj}^0 \beta\left(\frac{\theta}{2\pi} \mid j, k - j + 1\right)$. The p'th trigonometric moments of Θ are given by*

$$\boldsymbol{\mu}'_p = \sum_{j=1}^{k} \varpi_{kj}^0 \left(E[\cos 2\pi p \mathcal{B}_j] + iE[\sin 2\pi p \mathcal{B}_j]\right)$$

where $\mathcal{B}_j$ is a beta random variable with density function $\beta(\cdot \mid j, k - j + 1)$ as defined in (2.3) such that

$$(2.12) \quad E[\cos(2\pi p \mathcal{B}_j)] \quad = \quad \frac{1}{B(j, k-j+1)} \sum_{r=0}^{k-j} (-1)^r \begin{pmatrix} k-j \\ r \end{pmatrix} I_p(j + r - 1)$$

$$(2.13) \quad \quad = \quad \frac{1}{B(j, k-j+1)} \sum_{r=0}^{j-1} (-1)^r \begin{pmatrix} j-1 \\ r \end{pmatrix} I_p(k - j + r)$$

$$(2.14) \quad E[\sin(2\pi p \mathcal{B}_j)] \quad = \quad \frac{1}{B(j, k-j+1)} \sum_{r=0}^{k-j} (-1)^r \begin{pmatrix} k-j \\ r \end{pmatrix} J_p(j + r - 1)$$

$$(2.15) \quad \quad = \quad \frac{1}{B(j, k-j+1)} \sum_{r=0}^{j-1} (-1)^{r+1} \begin{pmatrix} j-1 \\ r \end{pmatrix} J_p(k - j + r)$$

where $I_p(0) = J_p(0) = I_p(1) = 0$, $J_p(1) = -\frac{1}{2\pi p}$ and for $C = 2, 3, 4, \ldots,$

$$(2.16) \quad I_p(C) \quad = \sum_{c=1}^{\lfloor \frac{C}{2} \rfloor} (-1)^{c-1} \frac{C!}{(C-2c+1)!} \frac{1}{(2\pi p)^{2c}}$$

$$(2.17) \quad J_p(C) \quad = \sum_{c=1}^{\lfloor \frac{C+1}{2} \rfloor} (-1)^c \frac{C!}{(C-2c+2)!} \frac{1}{(2\pi p)^{2c-1}}.$$

Proof. The expressions for the circular moments in terms of the I_p and J_p functions are immediate by expanding the polynomial terms of the beta density and then, the values of these functions can be found in standard collections of integrals, e.g. Gradshteyn and Ryzhik (2007). □

The moments of a variable distributed as a Bernstein polynomial density with an origin distinct from zero can be derived immediately from the previous theorem by using the expressions for sine and cosine of a sum.

3. Estimation for the Circular Bernstein Polynomial

Given a sample of n data generated from a linear variable, X, with support $[0,1]$, then from (2.2), the natural Bernstein polynomial estimator of order k for the density of X is given by,

$$\hat{b}_k(x) = \sum_{j=1}^{k} \left(\hat{F}_X \left(\frac{j}{k} \right) - \hat{F}_X \left(\frac{j-1}{k} \right) \right) \beta(x \mid j, k-j+1),$$

for $x \in [0,1]$, where $\hat{F}_X(\cdot)$ is the empirical distribution function, see e.g. Vitale (1975).

Vitale (1975) demonstrates that as long as the true density function, f_X, is twice differentiable, then the bias of this estimator is,

$$(3.1) \qquad E\left[\hat{b}_k(x) - f_X(x) \right] = \frac{1}{2k} \left[(1-2x) f_X'(x) + x(1-x) f_X''(x) \right] + o\left(\frac{1}{k} \right)$$

where, in particular the estimator is free of boundary bias as the o term is uniform in $[0,1]$. Furthermore, the variance of the estimator is

$$(3.2) \qquad V\left[\hat{b}_k(x) \right] = \frac{\sqrt{k}}{n} \frac{f(x)}{2\sqrt{\pi x(1-x)}} + o\left(\frac{\sqrt{k}}{n} \right)$$

Vitale shows that the optimal choice of k is to set $k = O\left(n^{2/5} \right)$ and shows that in this case, the estimator converges at a rate $n^{-4/5}$, which is the same rate of convergence as the alternative, kernel based estimators.

Suppose now that we have a sample of n data, $\{\theta_1, \ldots, \theta_n\}$, observed from a continuous, unknown, circular density, $f_\Theta(\theta)$. Then, for any origin, ν, a standard, linear Vitale estimator having the same asymptotic properties could be defined by,

$$\hat{b}_k^\nu(\theta) = \frac{1}{2\pi} \sum_{j=1}^{k} \left(\hat{F}_\Theta^\nu \left(\frac{2\pi j}{k} \right) - \hat{F}_\Theta^\nu \left(\frac{2\pi(j-1)}{k} \right) \right) \beta\left(\frac{\theta}{2\pi} \,\middle|\, j, k-j+1 \right)$$

$$(3.3) \qquad = \frac{1}{2\pi} \sum_{j=1}^{k} \hat{\omega}_{kj}^\nu \beta\left(\frac{\theta}{2\pi} \,\middle|\, j, k-j+1 \right)$$

where $\hat{F}_\Theta^\nu(\cdot)$ is the empirical distribution function defined from an origin ν and $\hat{\omega}_{kj}^\nu = \hat{F}_\Theta^\nu \left(\frac{2\pi j}{k} \right) - \hat{F}_\Theta^\nu \left(\frac{2\pi(j-1)}{k} \right)$ for $j = 1, \ldots, k$.

However, similar to (2.7), this estimator will only be circular if $\hat{\omega}_{k1}^\nu = \hat{\omega}_{kk}^\nu$ and, in contrast to Theorem 2.1, it may be that there exists no origin $\nu \in [0, 2\pi)$ which leads to a circular density estimate. For example, in the case that $k = 2$, if an odd number of data are observed, then no origin, ν, satisfies $\hat{F}_\Theta^\nu(\pi) = 1/2$. Therefore, in order to produce a valid, circular, density estimator, a modification of the Vitale estimator is required. We propose the following procedure.

Firstly, for a given order, k, we need to select a suitable origin. To do this, define

$$(3.4) \qquad d(\nu) = \hat{\omega}_{k1}^\nu - \hat{\omega}_{kk}^\nu, \qquad \text{for } 0 \le \nu < 2\pi,$$

which measures the difference between the first and the last weight of the beta mixture density in (3.3).

Now let $\hat{\boldsymbol{\nu}}_k = \arg\min |d(\nu)|$ be the set of origins which minimize the absolute distance between the first and last weights. Then for $\hat{\nu}_k \in \hat{\boldsymbol{\nu}}_k$, we can define a circular Bernstein polynomial estimator as

$$\hat{c}_k^{\hat{\nu}}(\theta) = \frac{1}{2\pi} \sum_{j=1}^{k} \hat{\varpi}_{kj}^{\hat{\nu}_k} \beta\left(\frac{\theta}{2\pi} \,\middle|\, j, k-j+1 \right)$$

where $\hat{\varpi}_{k1}^{\hat{\nu}_k} = \hat{\varpi}_{kk}^{\hat{\nu}_k} = \frac{\hat{\omega}_{k1}^{\nu} + \hat{\omega}_{kk}^{\nu}}{2}$ and $\hat{\varpi}_{kj}^{\hat{\nu}_k} = \hat{\omega}_{kj}^{\hat{\nu}_k}$ for $j = 2, \ldots, k-1$. Thus, this estimator modifies the standard Bernstein polynomial estimator by averaging the first and last weights, analogous to the approach introduced in Section 2.2.

3.1. Asymptotic properties. Here we shall examine the asymptotic properties of the proposed estimator. In order to do this, we decompose the difference between the estimator and the underlying circular density as follows:

$$
\begin{aligned}
\hat{c}_k^{\hat{\nu}_k}(\mathrm{mod}(\theta - \hat{\nu}_k, 2\pi)) - f_\Theta(\theta) &= A(\theta, \hat{\nu}_k) + B(\theta, \hat{\nu}_k) + C(\theta, \hat{\nu}_k) \quad \text{where} \\
A(\theta, \hat{\nu}_k) &= \hat{c}_k^{\hat{\nu}_k}(\mathrm{mod}(\theta - \hat{\nu}_k, 2\pi)) - \hat{b}_k^{\hat{\nu}_k}(\mathrm{mod}(\theta - \hat{\nu}_k, 2\pi)) \\
B(\theta, \hat{\nu}_k) &= \hat{b}_k^{\hat{\nu}_k}(\mathrm{mod}(\theta - \hat{\nu}_k, 2\pi)) - b_k^{\hat{\nu}_k}(\mathrm{mod}(\theta - \hat{\nu}_k, 2\pi)) \\
C(\theta, \hat{\nu}_k) &= b_k^{\hat{\nu}_k}(\mathrm{mod}(\theta - \hat{\nu}_k, 2\pi)) - f_\Theta(\theta)
\end{aligned}
$$
(3.5)

Now we can estimate the mean and variance of each component of the above decomposition individually. Firstly, we shall show that the difference between the corrected and uncorrected estimators in $A(\cdot)$ is always small. To do this, we introduce the following generalization of a continuous function taken from Burgin (2010).

3.1. Definition. A function $f : \mathbb{R} \to \mathbb{R}$ is called r-continuous at a point $a \in \mathbb{R}$ if $f(x)$ is defined at a and for any $\varepsilon > 0$ there is a $\delta > 0$ such that for any x with $|a - x| < \delta$, we have that $|f(x) - f(a)| < r + \varepsilon$.

Thus, an r-continuous function is basically a non-continuous function with jumps smaller than a quantity r. The following result, taken from Burgin (2010), is Bolzano's intermediate value theorem for r-continuous functions,

3.1. Lemma. Let $f : [a, b] \to \mathbb{R}$ be an r-continuous function. If $f(a) < 0$ and $f(b) > 0$, then there is at least one point $c \in [a, b]$ such that $|f(c)| < r$.

Proof. See Burgin (2010). $\qquad\square$

Now we can demonstrate the existence of an origin $\hat{\nu}_k$ such that $|d(\hat{\nu}_k)| \leq 1/n$.

3.2. Lemma. Let $\{\theta_1, \ldots, \theta_n\}$ be a random sample from a strictly continuous, circular, random variable, Θ, with density function $f_\Theta(\cdot)$. Then, for $k = 2, 3, \ldots,$ there exists at least one point $\hat{\nu}_k \in [0, 2\pi)$ such that $|d(\hat{\nu}_k)| \leq 1/n$.

Proof. For any $0 \leq \nu < 2\pi$, write $d(\nu) = d_1(\nu) - d_2(\nu)$ such that $d_1(\nu) = \hat{F}_\Theta^\nu\left(\frac{2\pi}{k}\right)$ and $d_2(\nu) = 1 - \hat{F}_\Theta^\nu\left(\frac{2\pi(k-1)}{k}\right)$. For a sample from a strictly continuous density, then $d_1(\nu)$ and $d_2(\nu)$ are both step functions with steps of size $1/n$ and therefore, $d(\nu)$ is a step function taking steps of size $1/n$ or $2/n$ so that $d(\nu)$ is $2/n$-continuous.

Now, assume that there exist two points, $0 \leq \nu_1 \neq \nu_2 < 2\pi$ such that $d(\nu_1) < 0$ and $d(\nu_2) > 0$. Then from Lemma 3.1 there exists at least one point, $0 \leq \hat{\nu}_k < 2\pi$, such that $|d(\hat{\nu}_k)| < 2/n$ and recalling that $\hat{d}(\cdot)$ is a step function, we have that $|d(\hat{\nu}_k)| \leq 1/n$.

On the contrary, suppose that $d(\nu)$ is always positive. Observe that

$$
d_1(\nu) = \frac{1}{n} \sum_{i=1}^n I_{(\nu, \nu + 2\pi/k]}(\theta_i) \text{ and } d_2(\nu) = \frac{1}{n} \sum_{i=1}^n I_{(\nu - 2\pi/k, \nu]}(\theta_i)
$$

where $I(\theta)_{(a,b]}$ is an indicator function taking the value 1 if $\theta \in (a, b]$ and 0 otherwise. Then, we have that,

$$
\sum_{i=1}^n I_{\left(\nu - \frac{2\pi}{k}, \nu\right]}(\theta_i) < \sum_{i=1}^n I_{\left(\nu, \nu + \frac{2\pi}{k}\right]}(\theta_i) < \sum_{i=1}^n I_{\left(\nu + 2\pi - \frac{2\pi}{k}, \nu + 2\pi\right]}(\theta_i)
$$

which is impossible, as we have that,

$$
\sum_{i=1}^n I_{\left(\nu - \frac{2\pi}{k}, \nu\right]}(\theta_i) = \sum_{i=1}^n I_{\left(\nu + 2\pi - \frac{2\pi}{k}, \nu + 2\pi\right]}(\theta_i).
$$

Equally, $d(\nu)$ cannot always be negative and so the lemma is proved. $\qquad\square$

Finally, we can now show that, in a similar way to Theorem 2.3 the difference between the corrected and uncorrected estimators is small.

3.3. Lemma. *Let $\{\theta_1,\ldots,\theta_n\}$ be a random sample from a strictly continuous, circular, random variable, Θ, with density function $f_\Theta(\cdot)$ and let $\hat{\nu}_k \in \hat{\boldsymbol{\nu}}_k$. Then we have that*

$$\left| \hat{c}_k^{\hat{\nu}_k}(\theta) - \hat{b}^{\hat{\nu}_k}(\theta) \right| \le \frac{k}{4\pi n}.$$

Proof. We have:

$$\hat{c}_k^{\hat{\nu}_k}(\theta) - \hat{b}_k^{\hat{\nu}_k}(\theta) = \tfrac{1}{4\pi}\left\{ \hat{F}^{\hat{\nu}_k}\left(\tfrac{2\pi}{k}\right) + \hat{F}^{\hat{\nu}_k}\left(\tfrac{2\pi(k-1)}{k}\right) - 1 \right\}\left(\beta\left(\tfrac{\theta}{2\pi}\,\middle|\,k,1\right) - \beta\left(\tfrac{\theta}{2\pi}\,\middle|\,1,k\right)\right)$$

and the maximum difference between the two estimators occurs when either $\theta = 0$ or $\theta = 2\pi$, so that

$$\max \left| \hat{f}_k^{\hat{\nu}_k}(\theta) - \hat{b}_k^{\hat{\nu}_k}(\theta) \right| \le \frac{k}{4\pi}\left| \hat{F}^{\hat{\nu}_k}\left(\frac{2\pi}{k}\right) + \hat{F}^{\hat{\nu}_k}\left(\frac{2\pi(k-1)}{k}\right) - 1\right| \le \frac{k}{4\pi n},$$

from Theorem 3.2. $\qquad\square$

In particular, Lemma 3.3 implies:

3.4. Corollary.

$$
\begin{aligned}
|E\left[A\left(\theta,\hat{\nu}_k\right)\right]| &\le \frac{k}{4\pi n} \\
V\left[A\left(\theta,\hat{\nu}_k\right)\right] &\le \frac{k^2}{16\pi^2 n^2}
\end{aligned}
$$

Now consider the second component, $B(\cdot)$, of (3.5).

3.5. Lemma.

$$
\begin{aligned}
E\left[B\left(\theta,\hat{\nu}_k\right)\right] &= 0 \\
V\left[B\left(\theta,\hat{\nu}_k\right)\right] &\le \frac{1}{n}\left(f_\Theta(\theta) + \frac{1}{4\pi k}\left[\max_\vartheta |f'_\Theta(\vartheta)| + \frac{1}{4}\max_\vartheta |f''_\Theta(\vartheta)| \right]\right)^2 \\
&\quad + \frac{k}{2\pi n}\left(f_\Theta(\theta) + \frac{1}{4\pi k}\left[\max_\vartheta |f'_\Theta(\vartheta)| + \frac{1}{4}\max_\vartheta |f''_\Theta(\vartheta)| \right]\right)
\end{aligned}
$$

Proof. The proof for the mean is immediate by recalling that for any variables, X, Y, then $E[X] = E[E[X|Y]]$ and noting that here, we have

$$E\left[B\left(\theta,\hat{\nu}_k\right) \mid \hat{\nu}_k\right] = 0$$

as the expected value of the empirical cumulative distribution function is equal to the true distribution function.

For the variance term, recall that $V[X] = V[E[X|Y]] + E[V[X|Y]]$ and given that in our case, the first term is zero, we only need to consider

$$E\left[V\left[B\left(\theta,\hat{\nu}_k\right) \mid \hat{\nu}_k\right]\right] = E\left[V\left[\hat{b}_k^{\hat{\nu}_k}\left(\mathrm{mod}(\theta - \hat{\nu}_k, 2\pi)\right) \mid \hat{\nu}_k\right]\right].$$

Now following Vitale (1975), consider

$$V\left[\hat{b}_k^{\hat{\nu}_k}\left(\mathrm{mod}(\theta - \hat{\nu}_k, 2\pi)\right) \mid \hat{\nu}_k\right] = V\left[\frac{1}{2\pi}\sum_{j=1}^{k} \hat{\omega}_{kj}^{\hat{\nu}_k}\beta\left(x_\theta^{\hat{\nu}_k}\,\middle|\,j, k-j+1\right)\,\middle|\,\hat{\nu}_k\right]$$

where x_θ^ν is as in (2.9),

$$
= \frac{1}{4\pi^2 n}\left[\left(\sum_{j=1}^{k}\omega_{kj}^{\hat{\nu}_k}\beta\left(x_\theta^{\hat{\nu}_k}\Big| j, k-j+1\right)\right)^2 + \right.
$$

$$
\left. \sum_{j=1}^{k}\omega_{kj}^{\hat{\nu}_k}\beta\left(x_\theta^{\hat{\nu}_k}\Big| j, k-j+1\right)^2\right]
$$

$$
\leq \frac{1}{n}\left(f_\Theta(\theta) + \frac{1}{4\pi k}\left[\max_\vartheta|f_\Theta'(\vartheta)| + \frac{1}{4}\max_\vartheta|f_\Theta''(\vartheta)|\right]\right)^2
$$

$$
+\frac{1}{4\pi^2 n}\sum_{j=1}^{k}\omega_{kj}^{\hat{\nu}_k}\beta\left(x_\theta^{\hat{\nu}_k}\Big| j, k-j+1\right)^2
$$

applying (2.8) to the first term and, then recalling that $\beta(x|j, k-j+1) \leq k$ for $0 \leq x \leq 1$ and applying (2.8) to the second term, we have:

$$
V\left[\hat{b}_k^{\hat{\nu}_k}\left(\mathrm{mod}(\theta-\hat{\nu}_k, 2\pi)\right)\mid \hat{\nu}_k\right] \leq \frac{1}{n}\left(f_\Theta(\theta) + \frac{1}{4\pi k}\left[\max_\vartheta|f_\Theta'(\vartheta)| + \frac{1}{4}\max_\vartheta|f_\Theta''(\vartheta)|\right]\right)^2
$$

$$
+\frac{k}{2\pi n}\left(f_\Theta(\theta) + \frac{1}{4\pi k}\left[\max_\vartheta|f_\Theta'(\vartheta)| + \frac{1}{4}\max_\vartheta|f_\Theta''(\vartheta)|\right]\right)
$$

$\square$

Finally, consider $C(\cdot)$ in (3.5). The following result is a direct consequence of (2.8).

3.6. Corollary.

$$
|E\left[C\left(\theta, \hat{\nu}_k\right)\right]| \leq \frac{1}{4\pi k}\left[\max_\vartheta|f_\Theta'(\vartheta)| + \frac{1}{4}\max_\vartheta|f_\Theta''(\vartheta)|\right]
$$

$$
V\left[C\left(\theta, \hat{\nu}_k\right)\right] \leq \frac{1}{16\pi^2 k^2}\left[\max_\vartheta|f_\Theta'(\vartheta)| + \frac{1}{4}\max_\vartheta|f_\Theta''(\vartheta)|\right]^2.
$$

The previous results can be combined to give the following theorem.

3.7. Theorem. *Let* $k, n \to \infty$ *where* $k \leq \sqrt{n}$. *Then*

$$
\left|E\left[\hat{c}_k^{\hat{\nu}_k}\left(\mathrm{mod}(\theta-\hat{\nu}_k, 2\pi)\right) - f_\Theta(\theta)\right]\right| \leq \frac{1}{4\pi k}\left[\max_\vartheta|f_\Theta'(\vartheta)| + \frac{1}{4}\max_\vartheta|f_\Theta''(\vartheta)|\right] + o\left(\frac{1}{k}\right).
$$

Letting $k \to \infty$ *such that* $kn^{\frac{1}{3}} \to 1$ *then*

$$
E\left[\left(\hat{c}_k^{\hat{\nu}_k}\left(\mathrm{mod}(\theta-\hat{\nu}_k, 2\pi)\right) - f_\Theta(\theta)\right)^2\right] = \frac{f_\Theta(\theta)}{n^{2/3}} + o\left(n^{-\frac{2}{3}}\right).
$$

Proof. The proof for the mean is immediate by summing the formulae for the means in Lemma 3.5 and Corollaries 3.4 and 3.6. Recalling that the mean squared error of the proposed estimator is composed of the bias squared and the variance and observing that the dominant term in the variance comes from $V\left[\hat{b}_k^{\hat{\nu}_k}\left(\mathrm{mod}(\theta-\hat{\nu}_k, 2\pi)\right)\right]$ gives the result. $\square$

We should note that work in e.g. Hall et al (1987) that optimal, circular kernel based estimators achieve the same convergence rate of $n^{-4/5}$ as linear kernel based estimates and that therefore, in contrast to the linear case, the Bernstein polynomial based estimators have slightly poorer convergence properties than the circular kernel estimators.

Two further comments are in order here. Firstly note that the previous results are valid for any choice of origin, $\hat{\nu}_k$, in $\hat{\boldsymbol{\nu}}_k$ and that, typically, the cardinality of $\hat{\boldsymbol{\nu}}_k$ will

be greater than 1. In order to define a unique origin, we suggest selecting the origin, $\hat{\nu}_k \in \hat{\boldsymbol{\nu}}_k$, which maximizes the (pseudo) log-likelihood estimate,

$$\sum_{i=1}^{n} \log \sum_{j=1}^{k} \hat{f}_k^{\hat{\nu}_k}(\theta_i).$$

Of course, many alternative criteria could also be used, e.g. maximizing the p value for some goodness of fit test. Secondly, in order to choose k in practice, we propose using standard, least squares, cross validation (LSCV) as in e.g. Wand and Jones (1995).

4. Examples

In this section, we carry out both a simulation study to compare our approach with some of the alternatives and we also illustrate our approach with a real data example.

4.1. Simulated examples. Here, we compare the performance of the use of Bernstein polynomials with respect to some non-parametric alternatives, that is kernel density estimation (Bai et al, 1988) and non-negative trigonometric sums (Fernández-Durán, 2004). To do this, we have simulated 500 samples of size 500 of the following models:

(1) A cosine or cardioid distribution with circular mean π. See e.g. Mardia and Jupp (1999).
(2) A von Mises distribution, VM(π,1), with circular mean π and concentration parameter 1. See e.g. Mardia and Jupp (1999).
(3) A wrapped normal Laplace distribution, WNL(π,0.0001,0.25,1). See Reed and Pewsey (2009).
(4) A distribution based on non-negative trigonometric sums with density
$$f_\Theta(\theta) = \frac{1 + 0.6\cos(2\theta - \pi) + 0.4\sin(3\theta - \frac{\pi}{2})}{2\pi}.$$
(5) A mixture of two von Mises distributions: $0.7\text{VM}\left(\frac{\pi}{3}, 0.5\right) + 0.3\text{VM}\left(\frac{2\pi}{3}, 2\right)$.
(6) A mixture of two wrapped normal Laplace distributions:
$0.5\text{WNL}(\pi, 0.0001, 0.25, 1) + 0.5\text{WNL}(\frac{9}{5}\pi, 0.0001, 0.25, 1)$.

For each sample, we have used LSCV to choose the bandwidth for the kernel method and the degree of the polynomial for the Bernstein polynomial and the trigonometric sums. Then, the underlying distribution is fitted and the empirical mean integrated square error (MISE) is computed. Table 1 shows the values of one minus the mean empirical MISE (which approximates the expected MSE). The optimal model is given in bold.

Table 1. Estimated values of one minus the average empirical mean integrated square error

model	BP	Kernel	TS
1	**0.9980**	0.9972	0.9962
2	**0.9986**	**0.9986**	0.9960
3	0.9912	0.9951	**0.9959**
4	**0.9987**	0.9970	0.9976
5	**0.9987**	0.9986	0.9977
6	0.9921	**0.9957**	0.9951

As we can observe in Table 1 the simulations suggest that the circular Bernstein polynomial approach is competitive with existing nonparametric approaches.

4.2. Real data example. Here we consider data obtained from the Chicago Police Department website, `http://gis.chicagopolice.org/` which correspond to the reported, twenty four hour clock times of 1297 crimes perpetrated in Chicago on May 11th, 2007. A Bernstein polynomial density approximation of order $k = 20$ was fitted to these data. Also, a kernel density estimate based on a von Mises kernel was fitted for comparison. In both cases, LSCV was used for model selection. Figure 1 shows a rose plot and a histogram of the data of the data and the corresponding fitted densities based on an origin corresponding to midnight. It can be observed that both fits are quite similar
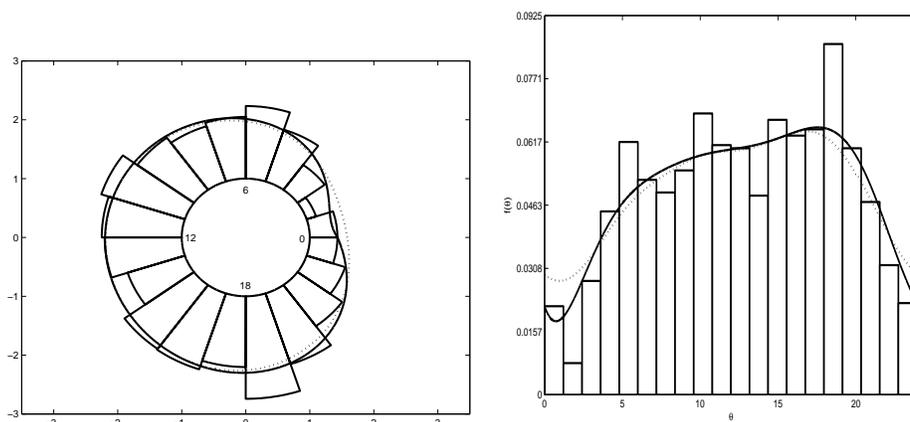

Figure 1. Roseplot (left hand side) and histogram of the Chicago crime data with fitted Bernstein polynomial (solid line) and kernel density (dashed line) approximations.

in nature although the Bernstein polynomial approach appears to provide a somewhat closer approach at the "boundary" points.

Finally, the fitted mean direction for the Bernstein polynomial model, calculated using the results of Section 3, and the empirical mean direction are 17:45 hours and 17:42 hours and the fitted and empirical mean circular resultant lengths calculated for the Bernstein polynomial approximation are 0.200 and 0.189 respectively. Thus, as should be expected, there is good agreement between the fitted and empirical moments. Note that an interesting further line of research would be to examine the convergence properties of the moments of the Bernstein polynomial approximation.

5. Conclusions

In this work we have demonstrated how Bernstein polynomial density estimators can be adapted to the case of circular data. A number of extensions are possible.

Firstly, the results presented in Section 3.1 have not considered the convergence properties of the selected origin, $\hat{\nu}_k$. In particular, it would be interesting to explore the convergence of the origin and to see if this can be used in order to provide an improved convergence rate for the fitted Bernstein polynomial. Also, as we know that in the theoretical case, the valid origins cluster around the turning points of the density function, then it would also be reasonable to consider the use of non-parametric approaches to the estimation of turning points of a density, for example the mode, see e.g. Eddy (1980).

Secondly, various modifications designed to reduce the bias of the linear Vitale estimator have been proposed, see e.g. Kakizawa (2004). It would be useful to examine whether such approaches can be applied to the circular case. Furthermore, Bayesian, Bernstein polynomial density estimates are analyzed in e.g. Petrone (1999b) and the adaption of this approach to the circular data case could also be explored.

Thirdly, we could also consider extensions to spherical data or data defined on the torus. Such approaches could be implemented by exploring generalizations of multivariate Bernstein polynomial density estimators as in e.g. Babu and Chaubey (2006).

Finally, it is interesting to consider modeling correlated circular circular or circular linear data by combining the marginal density estimation techniques developed here with the use of non-parametric Bernstein polynomial copula estimates (Sancetta and Satchell, 2004) to develop a non-parametric approach to bivariate directional data modeling and to directional regression. See also Carnicero et al (2013) for work in this direction.

Acknowledgments

References

[1] Babu, G.J., Canty, A.J. and Chaubey, Y.P. (2002). Application of Bernstein polynomials for smooth estimation of a distribution and density function, *J. Stat. Plan. Infer.*, **105**, 377–392.

[2] Babu, G.J. and Chaubey, Y.P. (2006). Smooth estimation of a distribution and density function on a hypercube using Bernstein polynomials for dependent random vectors, *Stat. Probabil. Lett.*, **76** 959–969.

[3] Bai, Z.D., Radhakrishna Rao, C. and Zhao, L.C. (1988). Kernel estimators of density function of directional data, *J. Multivariate Anal.*, **27**, 24–39.

[4] Burgin, M. (2010). Continuity in Discrete Sets. *Discussion paper* Available at http://arxiv.org/abs/1002.0036v1

[5] Butzer, P.L. (1955). Summability of generalized Bernstein polynomials, *Duke Math. J.,* **22**, 617–622.

[6] Carnicer, J.M. and Peña, J.M. (1993). Shape preserving representations and optimality of the Bernstein basis. *Adv. Comput. Math.*, **1**, 173–196.

[7] Carnicero, J.A., Ausín, M.C. and Wiper, M.P. (2013). Non-parametric copulas for circular-linear and circular-circular data: an application to wind directions. *Stoch. Env. Res. Risk A.* **27**, 1991–2002.

[8] Eddy, W.F. (1980). Optimum kernel estimators of the mode. *Ann. Statist.*, **8**, 870-882.

[9] Fernández-Durán, J.J. (2004). Circular distributions based on nonnegative trigonometric sums, *Biometrics*, **60**, 499–503.

[10] Fisher, N.I. (1989). Smoothing a sample of circular data, *J. Structural Geology*, **11**, 775–778.

[11] Gradshteyn, I.S. and Ryzhik, I.M. (2007). *Table of Integrals, Series, and Products*, 7'th ed., Academic Press: Burlington.

[12] Hall, P., Watson, G.S. and Cabrera, J. (1987). Kernel density estimation for spherical data, *Biometrika*, **74**, 751–762.

[13] Kakizawa, Y. (2004). Bernstein polynomial probability density estimation, *J. Nonparametr. Stat.*, **16**, 709–729.

[14] Klemelä, J. (2000). Estimation of Densities and Derivatives of Densities with Directional Data, *J. Multivariate Anal.*, **73**, 18–40.

[15] Lorentz, G.G. (1997). *Bernstein Polynomials*, 2nd. ed., AMS Chelsea, New York.

[16] Mardia, K.V. and Jupp, P. E. (1999). *Directional Statistics*, Wiley, Chichester.

[17] Mooney A., Helms P.J. and Jolliffe I.T. (2003). Fitting mixtures of von Mises distributions: a case study involving sudden infant death syndrome, *Comput. Stat. Data An.*, **41**, 505–513.

[18] Petrone, S. (1999a). Random Bernstein polynomials, *Scand. J. Stat.*, **26**, 373–393.

[19] Petrone, S. (1999b). Bayesian density estimation using random Bernstein polynomials, *Can. J. Stat.*, **27**, 105–126.

[20] Petrone, S. and Wassermann, L. (2002). Consistency of Bernstein polynomial posteriors, *J. R. Stat. Soc. B*, **64**, 79–100.

[21] Phillips, G.M. (2003). *Interpolation and Approximation by Polynomials*, Springer, New York.

[22] Reed, W. J. and Pewsey, A. (2009). *Two nested families of skew-symmetric circular distributions*, *Test*, **18**, 516–528.

[23] Sancetta, A. and Satchell, S. (2004). The Bernstein copula and applications to modeling and approximations of multivariate distributions. *Econometric Theory*, **20**, 535–562

[24] Vitale, R.A. (1975). A Bernstein polynomial approach to density estimation. In *Statistical Inference and Related Topics*, Madan Lal Puri ed., Academic Press, New York, pp. 87–100.

[25] Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*. Chapman & Hall, Boca Raton.