On Smooth Density Estimation for Circular Data[1]

Yogendra P. Chaubey
Department of Mathematics and Statistics, Concordia University,
Montréal, QC H3G 1M8
e-mail: yogen.chaubey@concordia.ca

Abstract

Fisher (1989: *J. Structural Geology* **11**, 775-778) outlined an adaptation of the linear kernel estimator for density estimation that is commonly used in applications. However, better alternatives are now available based on circular kernels; see e.g. Di Marzio, Panzera, and Taylor, 2009: *Statistics & Probability Letters* **79**, 2066-2075. This paper provides a short review on modern smoothing methods for density and distribution functions dealing with the circular data. We highlight the usefulness of circular kernels for smooth density estimation in this context and contrast it with smooth density estimation based on orthogonal series. It is seen that the wrapped Cauchy kernel as a choice of circular kernel appears as a natural candidate as it has a close connection to orthogonal series density estimation on a unit circle. In the literature the use of von Mises circular kernel is investigated (see Taylor, 2008: *Computational Statistics & Data Analysis* **52**, 3493-3500), that requires numerical computation of Bessel function. On the other hand, the wrapped Cauchy kernel is much simpler to use. This adds further weight to the considerable role of the wrapped Cauchy distribution in circular statistics.

Keywords: circular kernels; kernel density estimator; orthogonal polynomials; orthogonal series density

1. Introduction

Given an i.i.d. $d-$ variate random sample $\{X_1, ..., X_n\}$ from a continuous distribution function F with density f, the Parzen-Rosenblatt kernel density estimator is given by

$$\tilde{f}(x; h) \equiv n^{-1} h^{-d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{1.1}$$

where h is known as the window-width or band-width and K is called the kernel function. The band-width h typically tends to 0 as the sample size n tends to infinity and K is typically a symmetric density function centered around zero with unit variance. This estimator was proposed by Rosenblatt's (1956), that was further studied by Parzen (1962) and popularized in many subsequent papers. An alternative motivation for the kernel density estimator is provided recently in Chaubey *et al.* (2012) by smoothing the *empirical distribution function* that justifies the use of asymmetric kernels while considering density estimation for non-negative random variables.

[1]Expanded version of the Invited Paper for the invited session IPS-114, "Nonparametric Methods : Theory and Applications" at ISI-2017, Marrakech July 16-21, 2017

In what follows we consider estimation of the density for circular data, i.e. an absolutely continuous (with respect to the Lebesgue measure) circular density $f(\theta), \theta \in [-\pi, \pi]$, *i.e* $f(\theta)$ is $2\pi-$periodic,

$$f(\theta) \geq 0 \text{ for } \theta \in \mathbb{R} \text{ and } \int_{-\pi}^{\pi} f(\theta) d\theta = 1. \tag{1.2}$$

Given a random sample $\{\theta_1, ...\theta_n\}$ for the above density, the kernel density estimator may be written as

$$\tilde{f}(\theta; h) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{\theta - \theta_i}{h}\right). \tag{1.3}$$

Fisher (1989) proposed non-parametric density estimation for circular data by adapting the linear kernel density estimator (1.1) with a quartic kernel [see also Fisher (1993), §2.2 (iv) where an improvement is suggested], defined on $[-1, 1]$, that is given by

$$K(\theta) = \begin{cases} .9375(1-\theta^2)^2 & \text{for } -1 \leq \theta \leq 1; \\ 0 & \text{otherwise.} \end{cases} \tag{1.4}$$

[see Eqs. (4.40) and (4.41) of Fisher (1993).] Since the resulting estimator is not necessarily periodic, Fisher (1993) suggested to perform the smoothing by replicating the data to 3 to 4 cycles and considering the part in the interval $[-\pi, \pi]$. This problem is easily circumvented by using circular kernels, that has been investigated by Di Marzio et al. (2011). Taylor (2008) considered the von Misses circular normal distribution with concentration parameter κ for K, that gives the estimator for f as

$$\hat{f}_{vM}(\theta; \kappa) = \frac{1}{n} \sum_{i=1}^{n} K_{vM}(\theta; \theta_j, \kappa), \tag{1.5}$$

where

$$K_{vM}(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos(\theta - \mu)\}, \quad -\pi \leq \theta \leq \pi, \tag{1.6}$$

and discussed determination of the optimal data based choice for κ. Note that the von Mises distribution gets concentrated around μ for large κ.

In Section 2, I present a simple approximation theory motivation for considering the circular kernel density estimator given in (1.5). It may be noted that the wrapped Cauchy distribution with location parameter μ and concentration parameter ρ is given by

$$K_{WC}(\theta; \mu, \rho) = \frac{1}{2\pi} \frac{1-\rho^2}{1+\rho^2 - 2\rho\cos(\theta - \mu)} \;, -\pi \leq \theta < \pi, \tag{1.7}$$

that becomes degenerate at $\theta = \mu$ as $\rho \to 1$. The estimator of $f(\theta)$ based on the above kernel is given by

$$\hat{f}_{WC}(\theta; \rho) = \frac{1}{n} \sum_{i=1}^{n} K_{WC}(\theta; \theta_j, \rho). \tag{1.8}$$

Section 3 describes the approach of approximation using orthogonal functions in deriving density estimators and Fourier series density estimator is highlighted as a special case. This section also establishes the equivalence between the circular kernel estimator using wrapped Cauchy kernel and orthogonal series estimation of a specific complex function defined over a unit circle. In Section 4, some alternative approaches based on transformations are provided. The last section provides some examples and conclusions.

2. Motivation for the Circular Kernel Density Estimator

We consider the following theorem from approximation theory (see Mhaskar and Pai (2000)) to motivate the circular kernel density estimator. Before giving the theorem we will need the following definition:

Definition 2.1 *Let $\{K_n\} \subset C^*$ where C^* denotes the set of periodic analytic functions with a period 2π. We say that $\{K_n\}$ is an approximate identity if*

A. $K_n(\theta) \geq 0 \; \forall \; \theta \in [-\pi, \pi]$;

B. $\int_{-\pi}^{\pi} K_n(\theta) = 1$;

C. $\lim_{n\to\infty} \max_{|\theta|\geq\delta} K_n(\theta) = 0$ for every $\delta > 0$.

The definition above is motivated from the following theorem which is similar to the one used in the theory of linear kernel estimation (see Prakasa Rao (1983)). Also, note that we have replaced K_n of Mhaskar and Pai (2000)) by $2\pi K_n$ without changing the result of the theorem.

Theorem 2.1 *Let $f \in C^*$, $\{K_n\}$ be approximate identity and for $n = 1, 2, ...$ set*

$$f^*(\theta) = \int_{-\pi}^{\pi} f(\eta) K_n(\eta - \theta) d\eta. \tag{2.9}$$

Then we have

$$\lim_{n\to\infty} \sup_{\theta\in[-\pi,\pi]} |f^*(\theta) - f(\theta)| = 0. \tag{2.10}$$

Note that taking the sequence of concentration coefficients $\rho \equiv \rho_n$ such that $\rho_n \to 1$, the density function of the Wrapped Cauchy will satisfy the conditions in the definition in place of K_n. In general K_n, appearing in the above theorem may be replaced by a sequence of periodic densities on $[-\pi, \pi]$, that converge to a degenerate distribution at $\theta = 0$.
For a given random sample of $\theta_1, ..., \theta_n$ from the circular density f, the Monte-Carlo estimate of f^* is given by

$$\tilde{f}(\theta) = \frac{1}{n} \sum_{i=1}^{n} K_n(\theta_i - \theta). \tag{2.11}$$

The kernel given by the wrapped Cauchy density satisfies the assumptions in the above theorem that provides the estimator proposed in (1.8). This gives the motivation for considering circular kernels for nonparametric density estimation for circular data as proposed in discussed in a more detailed by Marzio MD, et al. (2009). However, their development considers circular kernels of order $r = 2$ that further requires

$$\int_{-\pi}^{\pi} \sin^j(\theta) K_n(\theta) d\theta = 0 \text{ for } 0 < j < 2.$$

The circular kernel density estimator based on the wrapped Cauchy weights is given by

$$\tilde{f}_{WC}(\theta) = \frac{1}{n} \sum_{i=1}^{n} f_{WC}(\theta_i - \theta). \tag{2.12}$$

that may be considered more convenient in contrast to the von Mises kernel due to the fact that it does not require computation of an integral $I_0(\kappa)$. Another justification of the circular kernel density estimator may presented by smoothing the empirical distribution function, an approach investigated in Babu, Canty and Chaubey (2002) and Babu and Chaubey (2006)[see also the recent paper by Chaubey et al. (2012)].

In this approach we approximate the distribution function instead of the density function using the approximation

$$\begin{aligned} F^*(\theta) &= \int_{-\pi}^{\pi} F(\eta) K_n(\eta-\theta) d\eta \\ &= 1 - \int_{-\pi}^{\pi} \Psi_{\theta,n}(\eta) dF(\eta) \end{aligned} \tag{2.13}$$

where $\Psi_{\theta,n}(.)$ is the sequence of distribution functions corresponding to the circular densities $K_n(. - \theta)$
Since F is unknown, using the edf as a plug-in estimate to estimate F^*, results into an smooth estimator of F given by

$$\begin{aligned} \tilde{F}(\theta) &= 1 - \int_{0}^{2\pi} \Psi_{\theta,n}(\eta) dF_n(\eta) \\ &= 1 - \frac{1}{n}\sum_{i=1}^{n} \Psi_{\theta,n}(\theta_i). \end{aligned} \tag{2.14}$$

Considering circular distributions with mean μ and concentration parameter $\rho_n \to 0$ as $n \to \infty$, let the density function corresponding to $\Psi_{\theta,n}$ correspond to a location family given by $\psi(\theta-\mu;\rho)$ that has mean μ and concentration parameter ρ, then a smooth density estimator $\tilde{f}(\theta)$ (as the derivative of $\tilde{F}(\theta)$ is given by

$$\tilde{f}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \psi(\theta_i - \theta; \rho). \tag{2.15}$$

which is of the same form as the circular kernel density estimator given in (2.11).
In the next section we provide details of the approach where the circular density is represented as a linear form using a set of basis functions.

3. Density Estimators using Approximation by Orthogonal Functions

Here we will consider approximating continuous bounded functions $f(x)$ in a compact interval $\mathbb{I} = [a, b] \subset \mathbb{R}$. For a given nonnegative function $w(x)$ defined on $\mathbb{I}$, the L_2 weighted norm of $f(x)$ is defined as

$$\|f\|_2^w = \int_a^b |f(x)|^2 \; w(x)dx. \tag{3.16}$$

The space of such functions will be denoted by $\mathbb{L}_2^w$. The general method of approximation of functions $f \in \mathbb{L}_2^w$ involves the set of basis functions $\{\varphi_k(x)\}_0^\infty$ and a non-negative weight function $w(x)$ such that

$$< \varphi_k, \varphi_{k'} >_w = \int_a^b \varphi_k(x)\varphi_{k'}(x)w(x)dx = \begin{cases} 0 & \text{for } k \neq k' \\ 1 & \text{for } k = k' \end{cases} \tag{3.17}$$

Then for $f \in \mathbb{L}_2^w$ the partial sum

$$f_N(x) = \sum_{k=0}^{N} g_k \varphi_k(x), \tag{3.18}$$

where

$$g_k = \int_a^b f(x)\varphi_k(x)w(x)dx, \tag{3.19}$$

is considered to be the 'best' approximation in terms of the fact that the coefficients g_k are such that $a_k = g_k$ minimise

$$\|f - f_N\|_2^w = \int_a^b |f(x)|^2 \; w(x)dx. \tag{3.20}$$

The original idea is attributed to Čencov (1962) that considered the cosine basis

$$\{\varphi_0(x) = 1, \varphi_j(x) = \sqrt{2}\cos(\pi j x), j = 1, 2, ...\}$$

and $w(x) = 1$. In recent literature, many other type of basis functions including trigonometric, polynomial, spline, wavelet and others have been considered. The reader may refer to Devroy and Györfi (1985), Efromvich (1999), Hart (1997), Walter (1994) for a discussion of different bases and their properties. Efromvich (2010) presents an extensive overview of density estimation by orthogonal series concentrated on the interval $[0, 1]$. As mentioned in Efromvich (2010) the choice of the basis function primarily depends on the support of the function. Thus for the densities on $(-\infty, \infty)$, or on $[0, \infty)$, Hermite and Laguerre series are recommended; see Devroye and Györfy (2001), Walter (1994), Hall (1980) and Walter (1977). For compact intervals, *trigonometric (or Fourier)* series are recommended; discussion about these can be found in Čencov (1980), Devroy and Györfy (1985), Efromvich (1999), Hart (1997), Silverman (1986), Hall (1981), Tarter and Lock (1993). Classical orthogonal polynomials such as Chebyshev, Jacobi, Legendre and Gegenbauer are also popular; see Trefthen (2013)), Rudzkis and Radavicius (2005) and Buckland (1992). Wavelet bases are becoming increasingly popular, due to their ability in visualizing local frequency fluctuations and discontinuities, even though their explicit form is not available.

Once the basis functions are chosen, the density $f(x)$ for a random sample $\{x_1, ..., x_n\}$ may be estimated by

$$\hat{f}_N(x) = \sum_{k=0}^{N} \hat{g}_k \varphi_k(x), \tag{3.21}$$

where

$$\hat{g}_k = \frac{1}{n}\sum_{i=1}^{n} \varphi_k(x_i). \tag{3.22}$$

Efromvich (2010) discusses in detail various strategies of selecting N, *albeit* in a more general setting by considering the density estimators of the form

$$\hat{f}(x) = \hat{f}(x, \{\hat{w}_k\}) = \sum_{k=0}^{\infty} \hat{w}_k \hat{g}_k \varphi_k(x) \tag{3.23}$$

that includes the truncated estimator $\hat{f}_J$ as well as *hard-thresholding* and *block-thresholding* estimators, commonly studied in the wavelet literature. However, this modification will not be pursued in further discussion.

The estimators using orthogonal series of cosine functions and Fourier series are easy to implement. Using the truncated cosine series, the density estimator is given by

$$\hat{f}_{OC}(\theta) = \frac{1}{2\pi} + \sum_{k=1}^{N} \hat{g}_k \cos(k\theta). \tag{3.24}$$

where

$$\hat{g}_k = \frac{1}{n\pi} \sum_{i=1}^{n} \cos(k\theta_i).$$

This is appropriate for circular density functions that are symmetric around zero, however, the Fourier series is more general. The truncated Fourier series of $f(\theta)$ is given by

$$f(\theta) \approx \frac{1}{2} a_0 + \sum_{k=1}^{N} \{a_k \cos(k\theta) + b_k \sin(k\theta)\}, \tag{3.25}$$

where

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(\theta) \cos(k\theta) d\theta, \;\; k = 0, 1, ..., N \tag{3.26}$$

$$b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(\theta) \sin(k\theta) d\theta, \;\; k = 1, ..., N. \tag{3.27}$$

Considering these coefficients as expectations of appropriate functions, they can be estimated as

$$\hat{a}_k = \frac{1}{n\pi} \sum_{i=1}^{n} \cos(k\theta_i); k = 0, 1, 2.... \tag{3.28}$$

$$\hat{b}_k = \frac{1}{n\pi} \sum_{i=1}^{n} \sin(k\theta_i); k = 1, 2.... \tag{3.29}$$

Thus, the Fourier series density estimator is given by

$$\hat{f}_{FS}(\theta) = \frac{1}{2\pi} + \sum_{1}^{N} \{\hat{a}_k \cos(k\theta) + \hat{b}_k \sin(k\theta)\}. \tag{3.30}$$

N is considered a smoothing parameter and may be determined using the cross-validation method described in Efromvich (2010). A common problem with truncation in these estimators is that it may not produce a true density. In order to alleviate this problem Efromvich (1999) considers L_2 projection of $\hat{f}$ onto a class of non-negative densities given by

$$\breve{f}(x) = \max(0, \hat{f}(x) - c), \tag{3.31}$$

where c is chosen to make $\breve{f}$ a proper density.

Recently, Chaubey (2016) demonstrated an interesting connection between the circular kernel density estimator using the Cauchy kernel and orthogonal series on a circle for the function

$$W(z) = \int \left(\frac{\mathrm{e}^{i\tau} + z}{\mathrm{e}^{i\tau} - z} \right) f(\tau) d\tau. \tag{3.32}$$

This involves the real and complex Poisson kernels that are defined as

$$P_r(\theta,\varphi) = \frac{1-r^2}{1+r^2-2r\cos(\theta-\varphi)} \tag{3.33}$$

for $\theta, \varphi \in [-\pi, \pi)$ and $r \in [0, 1)$ and by

$$C(z,\omega) = \frac{\omega + z}{\omega - z} \tag{3.34}$$

for $\omega \in \partial\mathbb{D}$ and $z \in \mathbb{D}; \mathbb{D} = \{z \mid |z| < 1\}$, is the open unit disk and $\partial\mathbb{D} = \{z \mid |z| = 1\}$ is the boundary of the unit disk. The connection between these kernels is given by the fact that

$$P_r(\theta,\varphi) = \text{Re } C(r\mathrm{e}^{i\theta}, \mathrm{e}^{i\varphi}) = (2\pi) f_{WC}(\theta; \varphi, \rho). \tag{3.35}$$

where $z = r\mathrm{e}^{i\theta}$ for $r \in [0, 1), \theta \in [-\pi, \pi]$ and $i = \sqrt{-1}$. Using the result (see (ii) in §5 of Simon (2005)) that for Lebesgue a.e. θ,

$$f(\theta) = \frac{1}{2\pi} \lim_{r \uparrow 1} \text{Re } W(r\mathrm{e}^{i\theta}), \tag{3.36}$$

a smooth density estimator is proposed to be

$$\hat{f}_r(\theta) = \frac{1}{2\pi} \text{Re } W_n(r\mathrm{e}^{i\theta}) \tag{3.37}$$

by appropriately choosing r, where

$$W_n(z) = \frac{1}{n} \sum_{j=1}^{n} \left(\frac{\mathrm{e}^{i\theta_j} + z}{\mathrm{e}^{i\theta_j} - z} \right). \tag{3.38}$$

Thus the sample estimate of $\hat{f}_r(\theta)$ can be written as

$$\hat{f}_r(\theta) \quad = \quad \frac{1}{n} \sum_{j=1}^{n} f_{WC}(\theta; \theta_j, r). \tag{3.39}$$

On the other hand the Fourier expansion of $W(z)$ with respect to the basis $\{1, z, z^2, ...\}$ is given by

$$W(z) = 1 + 2 \sum_{j=1}^{\infty} c_j z^j \tag{3.40}$$

where

$$c_j = \int \mathrm{e}^{-ij\theta} f(\theta) d\theta,$$

is the j^{th} trigonometric moment. The series is truncated at some term N^* so that the the error is negligible. However, we show below that estimating the trigonometric moments $c_j, j = 1, 2, ...$ as

$$\hat{c}_j = \frac{1}{n} \sum_{k=1}^{n} \mathrm{e}^{-ij\theta_k},$$

the estimator of $W(z)$ given by $\hat{W}(z) = 1 + 2\sum_{j=1}^{\infty} \hat{c}_j z^j$ is the same as $W_n(z)$. We have

$$\begin{aligned}
\hat{W}(z) &= 1 + \frac{2}{n}\sum_{j=1}^{n}\{\sum_{k=1}^{\infty} \mathrm{e}^{-ik\theta_j} z^k\} \\
&= 1 + \frac{2}{n}\sum_{j=1}^{n}\{\sum_{k=1}^{\infty} (\bar{\omega}_j z)^k\}; \omega_j = \mathrm{e}^{i\theta_j}; \bar{\omega}_j = \mathrm{e}^{-i\theta_j} \\
&= 1 + \frac{2}{n}\sum_{j=1}^{n}\left(\frac{\bar{\omega}_j z}{1-\bar{\omega}_j z}\right) \\
&= \frac{2}{n}\sum_{j=1}^{n}\left(\frac{1}{2} + \frac{\bar{\omega}_j z}{1-\bar{\omega}_j z}\right) \\
&= \frac{1}{n}\sum_{j=1}^{n}\left(\frac{1+\bar{\omega}_j z}{1-\bar{\omega}_j z}\right) \\
&= \frac{1}{n}\sum_{j=1}^{n} C(z, \omega_j),
\end{aligned}$$

which is the same as $W_n(z)$ given in (3.38). This ensures that the orthogonal series estimator of the density coincides with the circular kernel estimator, using the wrapped Cauchy kernel.

4. Transformation Based Density Estimators

In this section we outline some simple transformation estimators that are based on the fact that if we transform the angular data on $(-\pi, \pi)$ to some interval $\mathbb{I}$, where the properties of approximations on $\mathbb{I}$ are well known. Let $x = t(\theta)$ denote a one-to-one 2π periodic transformation from $(-\pi, \pi)$ to I and let $p(x)$ denote the density of the transformed data, then the density of the original data is given by

$$f(\theta) = p(t(\theta))|\frac{dt(\theta)}{d\theta}|. \tag{4.41}$$

4.1. Transformation for use with the kernel estimator on the real line

Denoting the angular random variable by Θ, the kernel density estimator on the real line may be applied using the transformation

$$X = \tan(\Theta/2), \tag{4.42}$$

that transforms the interval $[pi, \pi]$ to $(-\infty, \infty)$ and the kernel density estimator of X is given by

$$\hat{p}(x; h) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - \tan(\theta_i/2)}{h}\right). \tag{4.43}$$

and the transformation based kernel density estimator of $f(\theta)$ is given by

$$\hat{f}(\theta; h) = \frac{1}{1+\cos(\theta)}\ \hat{p}\left(\frac{\sin\theta}{1+\cos\theta}; h\right). \tag{4.44}$$

An attractive feature of the above procedure in contrast to Fisher's adaptation of the linear method is that the latter method gives a periodic estimator, however the former does not.

4.2. Transformation for use with Bernstein polynomial density estimator

Babu and Chaubey (2006) consider estimating the distributions defined on a hypercube, extending the univariate Bernstein polynomials (Babu, Canty and Chaubey (2002), Vitale (1973)). Denoting the empirical distribution function of a random sample of $n-$ observations from a random variable $X \in [0,1]$, by G_n, the Bernstein polynomial density estimator is given by

$$\hat{p}_B(x;m) = m \sum_{j=1}^{m} \left[F_n\left(\frac{j}{m}\right) - F_n\left(\frac{j-1}{m}\right) \right] \beta(x; j, m-j+1), x \in [0,1], \tag{4.45}$$

where $\beta(x;a,b)$ is given by

$$\beta(x;a,b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}, \tag{4.46}$$

and $B(a,b) = (a+b-1)!/[(a-1)!(b-1)!]$. We consider the transformation

$$t(\theta) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1}(c \tan(\theta/2)), \tag{4.47}$$

that maps the interval $[-\pi, \pi]$ to $[0,1]$ in a one-to-one monotonic transformation for all $c > 0$. Note that this provides a periodic transformation in contrast to the linear transformation $t(\theta) = \theta/(2\pi)$ for transforming the interval $[0, 2\pi]$ to $[0,1]$, as considered Carnicero et al. (2010). This transformation offers an extra parameter c that may be optimally chosen for a given random sample. The transformed estimator of $f(\theta)$ is given by

$$\hat{f}_B(\theta;m) = \frac{1}{2\pi} \hat{p}_B(t(\theta);m) \frac{c(1+tan^2(\theta/2)}{1+c^2 tan^2(\theta/2)}. \tag{4.48}$$

4.3. Transformation for use with orthogonal polynomials

Orthogonal polynomials of the Chebyshev's class on $[-1,1]$ can be converted to orthogonal polynomials on a circle $\mathbb{C} = \{z | \|z\| = 1\}$ through the transformation

$$x = \frac{1}{2}(z + z^{-1}).$$

This has been quite popular in numerical approximation of functions (see for example Trefethen (2013), Chapter 3). the kth *Chebyshev polynomial* can be defined by the real part of the function z^k on the unit circle:

$$x = \frac{1}{2}(z + z^{-1}) = \cos\theta, \ \theta = \cos^{-1} x, \tag{4.49}$$

$$T_k(x) = \frac{1}{2}(z^k + z^{-k}) = \cos(k\theta). \tag{4.50}$$

The following theorems justify the use of orthogonal polynomial estimators (see Rudin (1976)).

Theorem 4.2 *If h is Lipschitz continuous on $[-1,1]$, it has a unique representation as Chebyshev series,*

$$h(x) = \frac{1}{2}\, a_0 + \sum_{k=1}^{\infty} a_k T_k(x), \tag{4.51}$$

which is absolutely and uniformly convergent. The coefficients are given by the formula

$$a_k = \frac{2}{\pi}\int_{-1}^{1}\frac{h(x)T_k(x)}{\sqrt{(1-x^2)}}dx, \tag{4.52}$$

and for $k = 0$, by the same formula with the factor $2/\pi$ changed to $1/\pi$.

If $h(x)$ represents a density on $[-1, 1]$, a_k can be estimated by

$$\hat{a}_0 = \frac{1}{n\pi}\sum_{i=1}^{n}\frac{1}{\sqrt{(1-x_i^2)}}, \tag{4.53}$$

$$\hat{a}_k = \frac{2}{n\pi}\sum_{i=1}^{n}\frac{T_k(x_i)}{\sqrt{(1-x_i^2)}}; k = 1, 2, ... \tag{4.54}$$

$$\tag{4.55}$$

For using Chebyshev's polynomials in order to provide circular density estimator we transform the circular data as $x_i = 2\tan^{-1}(\tan(\theta_i/2))/\pi$ that essentially provides a 2π periodic transformation to the interval $[-1, 1]$, and the Chebyshev's polynomial circular density estimator is given by

$$\hat{f}_{CP}(\theta) = \frac{1}{2\pi}\hat{a}_0 + \frac{1}{\pi}\sum_{k=1}^{N}\hat{a}_k T_k(\theta/\pi) \tag{4.56}$$

The Chebyshev weight function, however, is singular at the extremes of the interval of support. Arbitrary power singularities may be assigned to each extreme giving a general weight function

$$w(x) = (1-x)^{\alpha}(1+x)^{\beta} \tag{4.57}$$

where $\alpha, \beta > 0$ are parameters. The associated polynomials are known as *Jacobi polynomials,* usually denoted as $\{P_n^{(\alpha,\beta}\}$. The special case $\alpha = \beta$, gives orthogonal polynomials that are known as as *Gegenauer* or *ultraspherical polynomials* and are subject of much discussion in numerical analysis; see Koornwinder et al. (2010). The most special case of all $\alpha = \beta = 0$ gives a constant weight function and produces what are known as Legendre polynomials denoted by $P_n(x), n = 0, 1, 2, ...$ that define a orthogonal system for the interval $[-1, 1]$. They may be simply described as

$$P_0(x) = 1, P_1(x) = x \tag{4.58}$$

and the recurrence relation

$$(k+1)P_{k+1}(x) = (2k+1)xP_k(x) - kP_{k-1}(x). \tag{4.59}$$

Thus

$$P_2(x) = \frac{3}{2}x^2 - \frac{1}{2}, P_3(x) = \frac{5}{2}x^3 - \frac{3}{2}x, ...etc. \tag{4.60}$$

An explicit representation may be given by the following formula:

$$P_k(x) = 2^k\sum_{j=0}^{k}\binom{k}{j}\binom{\frac{k+j-1}{2}}{k}x^j \tag{4.61}$$

This avoids the possible numerical problem in computing the coefficients due to singularity at the extremes. Hence this will be a preferred alternative to the Chebyshev polynomials. The expansion of a function $h(x)$ in terms of the Legendre polynomials is given by

$$h(x) = \sum_{k=0}^{\infty} c_k P_k(x) \tag{4.62}$$

where

$$c_k = \frac{1+2k}{2} \int_{-1}^{1} h(x) P_k(x) dx. \tag{4.63}$$

Unbiased estimators of the coefficients c_k are given by

$$\hat{c}_k = \frac{1+2k}{2n} \sum_{i=1}^{n} P_k(x_i). \tag{4.64}$$

Thus the density estimator in the original scale is given by

$$\hat{f}_{LP}(\theta) = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{N} \hat{g}_k P_k(\theta/\pi) \tag{4.65}$$

where

$$\hat{g}_k = \frac{1+2k}{2n} \sum_{i=1}^{n} P_k(\theta_i/\pi); k = 1, 2, ... \tag{4.66}$$

5. Examples and Conclusions

5.1 Examples

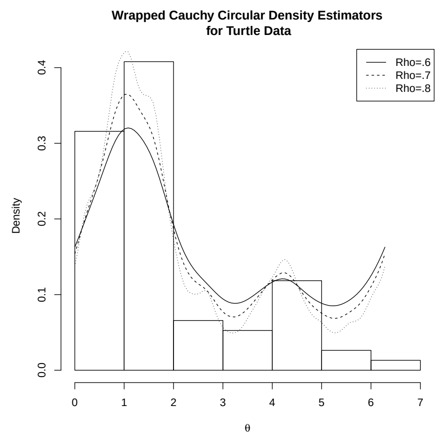
In this section we illustrate some of the density estimators by considering the wellknown Turtle data and Ants data. The Turtle data set gives the measurements of the directions taken by 76 turtles after treatment that is available from Appendix B.3 in Fisher (1993), whereas the Ants data set gives the measurements of the directions chosen by 100 ants in response to an evenly illuminated black target that is available from Appendix B.7 in Fisher (1993).

Figure 5.1 gives plots of the histogram with superimposed density estimators based on the wrapped Cauchy kernel and the von Misses kernel along with the transformed kernel estimators based on the classical kernel estimators based on Gaussian and logistic kernels for different values of the concentration and variance parameters, for Turtle data and Figure 5.2 presents the same for the Ants data. The kernel estimator for the transformed data on the real line is obtained from (1.3) where for the normal kernel

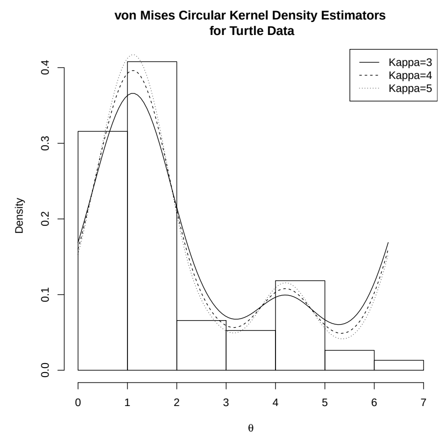
$$K(u) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}u^2}$$

and that for the logistic kernel

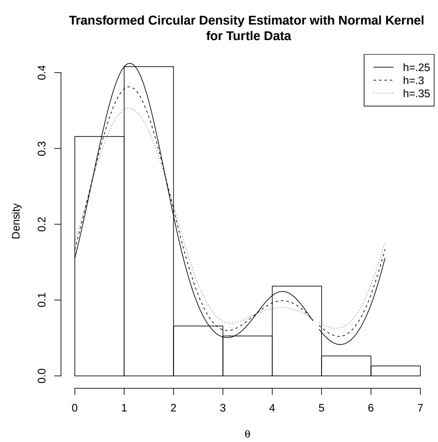
$$K(u) = \frac{\mathrm{e}^{-u}}{(1+\mathrm{e}^{-u})^2}.$$



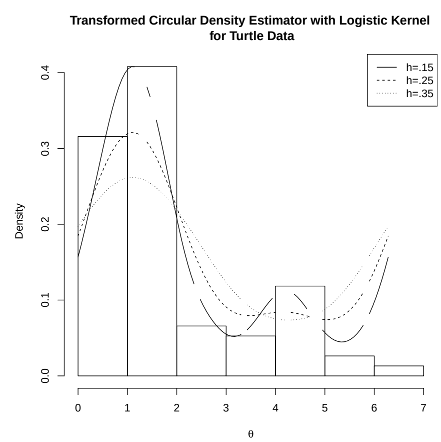
(a)



(b)



(c)



(d)

Figure 5.1: Histogram and Circular Kernel Density Estimators for Turtle data
(a) Wrapped Cauchy kernel, $\rho = .6, .7, .8.$; (b) von Mises kernel, $\kappa = 3, 4, 5.$;
(c) Transformed linear kernel with Gaussian kernel, $h = .25, .3, .35.$;
(d) Transformed linear kernel with logistic kernel, $h = .15, .25.35.$;

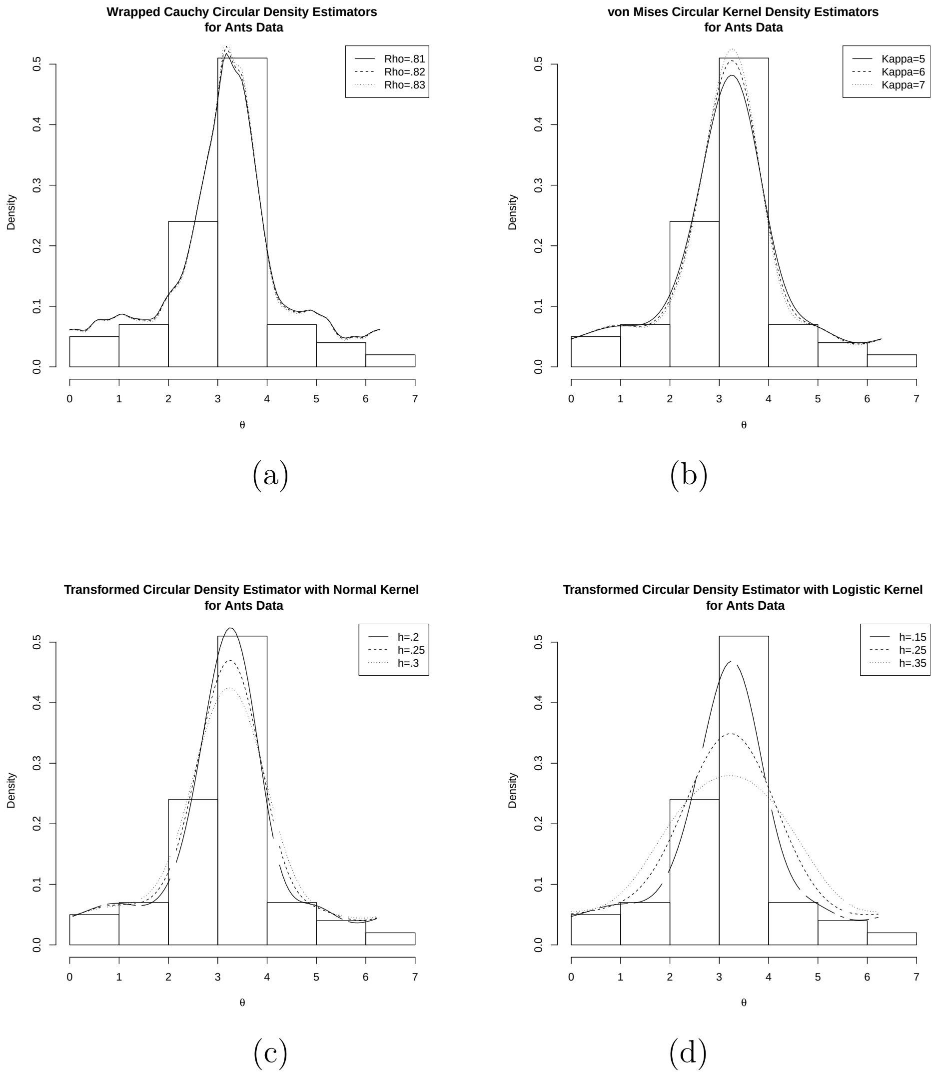


Figure 5.2: Histogram and Circular Kernel Density Estimators for Ants data
(a) Wrapped Cauchy kernel, $\rho = .81, .82, .83$; (b) von Mises kernel, $\kappa = 5, 6, 7.$;
(c) Transformed linear kernel with Gaussian kernel, $h = .2, .25, .3.$;
(d) Transformed linear kernel with logistic kernel, $h = .15, .25, .35$.

The von Mises kernel seems to provide smoother plots as compared to wrapped Cauchy, however, both the kernels provide similar estimators, qualitatively. The estimators obtained by IS transformation also produce similar results as to those given by the circular kernel estimators. Smoothing parameter may be selected using the proposal described in Taylor (2008). An enhanced strategy is to investigate a range of values around the value given by cross-validation.

5.2 Conclusions

This paper provides a simple approximation result for continuous function on defined on a unit circle to motivate circular kernel density estimator. An interesting connection between this estimator using the wrapped Cauchy kernel and an orthogonal series estimator is discovered that shows that truncation in the orthogonal series estimator may be avoided by using the wrapped Cauchy kernel. The approximation result used here also provides circular density estimator as the derivative of smooth distribution function estimator. Some further families of estimators are suggested using stereographic projection of circle on different intervals that might provide easily computable estimators from well known software. Further studies about the estimators and their properties is being carried out in ongoing research investigation of the author.

References

Babu, G. J., Canty, A. and Chaubey, Y. P. (2002). Application of Bernstein polynomials for smooth estimation of a distribution and density function. *J. Statist. Plann. Inference* , **105**, 377-392.

Babu, G. J., and Chaubey, Y. P. (2006). Smooth estimation of a distribution and density function on hypercube using Bernstein polynomials for dependent random vectors. *Statistics and Probability Letters*, **76**, 959-969.

Buckland, S. T. (1992). Fitting density functions with polynomials. *J. R. Stat. Soc.*, **A41**, 6376.

Carnicero, J. A., Wiper, M. P. and Ausín, M. C. (2010). Circular Bernstein polynomial distributions. *Working Paper 10-25, Statistics and Econometrics Series 11*, Departamento de Estadstica Universidad Carlos III de Madrid. `e-archivo.uc3m.es/bitstream/handle/10016/8318/ws102511.pdf?sequence=1`

Čencov, N. N. (1962). Evaluation of an unknown distribution density from observations. *Soviet Math. Dokl.*, **3**, 1559-1562.

Čencov, N. N. (1980). *Statistical Decision Rules and Optimum Inference.* New York: Springer-Verlag.

Chaubey, Y. P. (2016). Smooth Kernel Estimation of a Circular Density Function: A Connection to Orthogonal Polynomials on the Unit Circle. *Preprint*, `https://arxiv.org/abs/1601.05053`

Chaubey, Y.P.; Li, J.; Sen, A. and Sen, P. K. (2012). A new smooth density estimator for non-negative random variables. *Journal of the Indian Statistical Association*, **50**, 83-104.

Devroye, L. and Györfi, L. (1985). *Nonparametric Density Estimation: The L_1 View.* New York: John Wiley & Sons

Di Marzio, M., Panzera, A., Taylor, C.C. (2009). Local Polynomial Regression for Circular Predictors. *Statistics & Probability Letters*, **79**(19), 2066–2075.

Di Marzio M, Panzera A, Taylor C. C. (2011). Kernel density estimation on the torus. *Journal of Statistical Planning & Inference*, **141**, 2156-2173.

Efromvich, S. (1999). *Nonparametric Curve Estimation: Methods, Theory, and Applications*, New York: Springer.

Efromvich, S. (2010). Orthogonal series density estimation. *Wiley Interdisciplinary Rreviews*, **2**, 467-476.

Eğecioğlu, Ömer and Srinivasan, Ashok (2000). Efficient nonparametric density estimation on the sphere with applications in fluid mechanics. Siam J. Sci. Comput. **22** 152-176.

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

Fisher, N. I. (1993). *Statistical Analysis of Circular Data.* Cambridge University Press, Cambridge.

Hall, P. (1980). Estimating a density on the positive half line by the method of orthogonal series. *Ann. Inst. Stat. Math.* **32**, 351362.

Hall, P. (1981). On trigonometric series estimates of densities.*Ann. Stat.* **9**, 683685.

Hart, J. D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests.* New York: Springer.

Koornwinder, T. H., Wong, R. S. C., Koekoek, R., and Swarttouw, R.F. (2010). Orthogonal polynomials. In *NIST Handbook of Mathematical Functions*, Eds: Olver, Frank W. J.; Lozier, Daniel M.; Boisvert, Ronald F.; Clark, Charles W., London: Cambridge University Press.

Mhaskar, H. N. and Pai, D. V. (2000). *Fundamentals of approximation Theory.* Narosa Publishing House, New Delhi, India.

Parzen, E. (1962). On Estimation of a Probability Density Function and Mode. *Ann. Math. Statist*, **33:3**, 1065-1076.

Prakasa Rao, B. L. S. (1983). *Non Parametric Functional Estimation.* Academic Press, Orlando, Florida.

Rudin, W. (1976). *Real and Complex Analysis*- Third Edition. McGraw Hill: New York.

Rudzkis, R. and Radavicius, M. (2005). Adaptive estimation of distribution density in the basis of algebraic polynomials. *Theory Probab Appl*, **49**, 93109.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Stat.*, **27:56**, 832837

Simon, B. (2005). *Orthogonal Polynomials on the Unit Circle*, Part 1: Classical Theory. American Mathematical Society, Providence, Rhode Island.

Silverman B. W. (1986). *Density Estimation for Statistics and Data Analysis.* London: Chapman & Hall.

Tarter, M. E. and Lock, M. D. (1993). *Model-Free Curve Estimation.* New York: Chapman & Hall.

Taylor, C. C. (2008). Automatic Bandwidth Selection for Circular Density Estimation. *Computational Statistics & Data Analysis*, **52**(7), 3493–3500.

Trefthen, L. N. (2013). *Approximation Theory and Approximation Practice.* SIAM: Philadelphia

Vitale, R. A. (1973). A Bernstein polynomial approach to density estimation. *Commun Stat*, **2**, 493-506.

Walter, G. G. (1977). Properties of Hermite series estimation of probability density. *Ann. Stat.*, **5**, 12581264.

Walter, G. G. (1994). *Wavelets and other Orthogonal Systems with Applications.* London: CRC Press.