



# A reliable data-based smoothing parameter selection method for circular kernel estimation

Jose Ameijeiras-Alonso<sup>1</sup>

Received: 22 February 2023 / Accepted: 10 January 2024 / Published online: 7 February 2024  
© The Author(s) 2024

## Abstract

A new data-based smoothing parameter for circular kernel density (and its derivatives) estimation is proposed. Following the plug-in ideas, unknown quantities on an optimal smoothing parameter are replaced by suitable estimates. This paper provides a circular version of the well-known Sheather and Jones bandwidths (J R Stat Soc Ser B Stat Methodol 53(3):683–690, 1991. <https://doi.org/10.1111/j.2517-6161.1991.tb01857.x>), with direct and solve-the-equation plug-in rules. Theoretical support for our developments, related to the asymptotic mean squared error of the estimator of the density, its derivatives, and its functionals, for circular data, are provided. The proposed selectors are compared with previous data-based smoothing parameters for circular kernel density estimation. This paper also contributes to the study of the optimal kernel for circular data. An illustration of the proposed plug-in rules is also shown using real data on the time of car accidents.

**Keywords** Circular data · Directional statistics · Kernel density estimation · Plug-in rule · Sheather and Jones bandwidth

## 1 Introduction

Circular data are observations that can be represented on the unit circumference and where periodicity must be taken into account. Classic examples appear when the goal is to model orientations or a periodic phenomenon with a known period. Several applications of circular data can be found, e.g., in Ley and Verdebout (2018) or SenGupta and Arnold (2022). The complicated features that circular data can exhibit on real data applications lead to several new flexible models in the statistical literature. A recent review of flexible parametric circular distributions can be found in Ameijeiras-Alonso and Crujeiras (2022).

When trying to obtain a flexible fit of the density function, an alternative to parametric models is the kernel density estimation. Kernel density estimation for circular data dates back to Beran (1979) and Hall et al. (1987), while the estimator of the density derivatives was studied by Di Marzio et al. (2011). Additionally, the extension of density (and its derivatives) to the (hyper-)toroidal and (hyper-)spherical cases was

considered by Di Marzio et al. (2011) and Klemelä (2000), respectively. It is well-known that, for all these estimators, the choice of the smoothing parameter is crucial when using these kernel methods.

In the usual linear inferential framework, where random variables are supported on the Euclidean space, one can find many approaches for selecting the “best” data-driven bandwidth parameter (see, e.g., Jones et al. 1996, for a discussion on this topic). Due to its good performance, one of the most-employed bandwidth selectors is the plug-in bandwidth proposed by Sheather and Jones (1991). The relevance of this plug-in selector is evident from the impressive number of citations that Sheather and Jones (1991) have received, and although other authors have introduced new bandwidth selectors, none of the proposals outperforms, in general, their plug-in selector.

There exists some literature on smoothing parameter selection for circular data, some of these ideas are based on cross-validation (Hall et al. 1987), rule-of-thumb (Taylor 2008), plug-in (Oliveira et al. 2012; García-Portugués 2013; Tenreiro 2022), or bootstrap (Di Marzio et al. 2011) techniques. But none of the approaches introduced so far presents an outstanding performance with respect to its competitors, as is the case with the proposal by Sheather and Jones (1991) in the linear case. Hence, the goal of this paper is to provide the needed theory to derive an algorithm that replicates the

✉ Jose Ameijeiras-Alonso  
jose.ameijeiras@usc.es

<sup>1</sup> CITMAga, Department of Statistics, Mathematical Analysis and Optimization, Universidade de Santiago de Compostela, Rúa Lope Gómez de Marzoa s/n, 15782 Santiago de Compostela, A Coruña, Spain

idea of the two-stage direct and solve-the-equation plug-in bandwidth selectors for circular kernel density estimation. In addition, the developed theory can be also employed when estimating the density derivatives.

Regarding the kernel choice, most of the results for circular density estimation fix the kernel to be the von Mises density function. In this paper, we study the asymptotic results for a more general class of kernels. This allow us to obtain the optimal kernels in the circular estimation context.

This paper is organized as follows. Section 2 is devoted to the definition of the circular kernel density derivative estimate. Also, the key function needed to derive the optimal smoothing parameter is introduced in this section. For a general circular kernel, the asymptotic mean integrated squared error and the optimal smoothing parameter of the derivative estimator are derived in Sect. 3. Section 4 is devoted to the estimation of density functionals needed to derive the plug-in rules. In Sect. 5, we discuss the kernel choice. Section 6 provides all the essential details to compute the two-stage direct and solve-the-equation plug-in smoothing selectors. We analysed the performance of the proposed plug-in rules through a simulation study. The summarized conclusions appear in Sect. 7. Section 8 uses a real data example to show the applicability of the proposed selector. Some final remarks are provided in Sect. 9. Section 10 describes the software that implements the proposed smoothing parameters. The complete simulation study showing that the proposed rules provide a competitive smoothing parameter, some additional comments regarding how to use the software, and the proofs of the theoretical results are provided as Supplementary Material.

## 2 Circular kernel density derivative estimation

Given a circular random sample in angles  $\Theta_1, \dots, \Theta_n \in [-\pi, \pi)$ , with an associated density function  $f$ , the circular kernel density estimator (see, for example, Oliveira et al. 2012) can be defined as follows:

$$\hat{f}_v(\theta) = \frac{1}{n} \sum_{i=1}^n K_v(\theta - \Theta_i), \quad (1)$$

where  $K_v$  is the circular kernel with a smoothing parameter  $v \in [0, 1]$ , representing the mean resultant length. It is worth noting that, in contrast to previous circular literature (Di Marzio et al. 2009, 2011), we use the notation  $K_v$  instead of  $K_\kappa$  since our smoothing parameter focuses on the mean resultant length rather than the concept of “concentration”. This choice allows us to establish a theory that is valid for both types of kernels, whether they depend on concentration,

such as the von Mises, or on the mean resultant length, such as the wrapped normal (see Sect. 5 for a formal introduction to these kernels).

The estimator (1) can be easily extended, when the objective is estimating the  $r$ -th derivative of  $f$ , denoted by  $f^{(r)}$ . In that case, the estimator of  $f^{(r)}$  can be defined as (Di Marzio et al. 2011),

$$\hat{f}_v^{(r)}(\theta) = \frac{1}{n} \sum_{i=1}^n K_v^{(r)}(\theta - \Theta_i). \quad (2)$$

The first question is which function can be employed as a kernel? In this case, we will assume that  $K_v$  satisfies assumptions reminiscent of the canonical conditions for the linear kernel (Wand and Jones 1995, Ch. 2). In particular, a circular kernel  $K_v$  (reflectively) symmetric about zero and square-integrable in  $[-\pi, \pi)$  would admit the following convergent Fourier series representation (see Mardia and Jupp 1999, Section 4.2),

$$K_v(\theta) = \frac{1}{2\pi} \left( 1 + 2 \sum_{j=1}^{\infty} \alpha_{K,j}(v) \cos(j\theta) \right), \quad (3)$$

where only the values of  $\alpha_{K,j}(v) \in [0, 1]$ , for  $j \in \{1, 2, \dots\}$ , depend on the employed kernel and the mean resultant length  $v$ . In Sect. 3, we will impose that Fourier series representation for the kernel  $K_v$ .

Secondly, a crucial element in kernel density estimation is the smoothing parameter. Generally, this parameter is taken as a non-random sequence, depending on the sample size, and then as a fixed value for a sample realization. In this work, we will introduce the function  $h$  depending on  $v$ ,

$$h \equiv h_K(v) = \frac{\pi^2}{3} + 4 \sum_{j=1}^{\infty} \frac{(-1)^j \alpha_{K,j}(v)}{j^2}. \quad (4)$$

We denote  $h$  as the *circular bandwidth* and the sequence of numbers  $h_n \equiv h_K(v_n)$  must satisfy  $\lim_{n \rightarrow \infty} h_n = 0$ . It is important to note that this circular bandwidth is constructed from trigonometric moments, distinguishing it from the concept of “bandwidth” used in scale-family kernels. The distinction arises because circular distributions, such as the von Mises, do not belong to the scale family. In this paper, we aim to derive an explicit expression for the optimal circular bandwidth without imposing assumptions that resemble scalar theory. Specifically, we do not require conditions like  $K_g(\theta) = K(\theta/g)/g$  (Tenreiro 2022) or  $K_{g_n}(\theta) = \lambda g_n^{-2} K((1 - \cos(\theta))/g_n^2)$ , for some constant  $\lambda$ , when  $g_n \rightarrow 0$  (García-Portugués 2013).

### 3 Asymptotic results

In this section, we establish the results needed for deriving the asymptotic mean integrated squared error (AMISE) and the AMISE-optimal smoothing parameter. Throughout this section, for a given derivative order  $r$ , we will employ the following assumptions to derive the asymptotic results.

- (A1) The circular density  $f$  is such that its derivative  $f^{(r+2)}$  is continuous.
- (A2) For all  $v$ , the kernel  $K_v$  admits the Fourier series representation (3).
- (A3) For  $t \in \{1, 2\}$ , define the function  $R_{K;r,t}$  as:

$$R_{K;r,t}(v) = \begin{cases} (2\pi)^{-1} \left(1 + 2 \sum_{j=1}^{\infty} \alpha_{K,j}^t(v)\right) & \text{if } r = 0, \\ \mathfrak{s}\pi^{-1} \sum_{j=1}^{\infty} j^{tr} \alpha_{K,j}^t(v) & \text{otherwise,} \end{cases}$$

$$\mathfrak{s} = \begin{cases} -1 & \text{if } t = 1 \text{ and } r \text{ modulo } 4 = 1, \\ -1 & \text{if } t = 1 \text{ and } r \text{ modulo } 4 = 2, \\ 1 & \text{otherwise.} \end{cases}$$

The smoothing parameter satisfies that  $\lim_{n \rightarrow \infty} h_K(v_n) = 0$  and  $\lim_{n \rightarrow \infty} n^{-1} R_{K;r,2}(v_n) = 0$ .

As mentioned in Sect. 1, Assumptions (A1) and (A2) are reminiscent of those employed in the standard linear case. For a large class of kernels, including the von Mises or the wrapped normal (see Sect. 5), we will see that Assumption (A3) translates into the standard conditions on the bandwidth, namely,  $\lim_{n \rightarrow \infty} h_n = 0$  and  $\lim_{n \rightarrow \infty} nh_n^{(2r+1)/2} = \infty$ . The result in Theorem 1 (see Section S3 of the Supplementary Material for a formal proof) states the AMISE order of the kernel derivative estimator (2). If we also assume the following two extra conditions, we can derive an explicit expression of the AMISE.

- (E1)  $\int_{-\pi}^{\pi} \theta^4 K_{v_n}(\theta) d\theta = o(h_n)$ , as  $n \rightarrow \infty$ .
- (E2)  $\lim_{n \rightarrow \infty} R_{K;r,2}(v_n) = \infty$  and, as  $n \rightarrow \infty$ ,  $\int_{-\pi}^{\pi} \theta^2 (K_{v_n}^{(r)}(\theta))^2 d\theta = o[R_{K;r,2}(v_n)]$ .

**Theorem 1** Under the Assumptions (A1)–(A3), we have

$$AMISE \left[ \hat{f}_v^{(r)} \right] = O \left( h_K^2(v_n) \right) + O \left( n^{-1} R_{K;r,2}(v_n) \right).$$

If we also assume Conditions (E1) and (E2), then the AMISE has the following explicit expression.

$$AMISE \left[ \hat{f}_v^{(r)} \right] = \frac{1}{4} h_K^2(v_n) \int_{-\pi}^{\pi} \left( f^{(r+2)}(\theta) \right)^2 d\theta + \frac{1}{n} R_{K;r,2}(v_n). \tag{5}$$

Theorem 1 states the expression of the AMISE which depend on the sample size  $n$ , the derivative of the density function  $f^{(r+2)}$ , the kernel  $K_v$ , and the mean resultant length  $v$ . The complete expression of the asymptotic mean squared error (AMSE) is provided in Section S3 of the Supplementary Material.

**Remark 1** Di Marzio et al. (2009, 2011) also analysed the AMISE of the circular kernel estimation obtaining a similar result as in (5), replacing  $h_K(v_n)$  by  $(1 - \alpha_{K,2}(v_n))/2$ . Note that, although not stated in their paper, additional conditions must be imposed to derive their asymptotic results. We further elaborate on this matter in Section S3.1 of the Supplementary Material. In particular, their results hold true assuming Condition (E1). Under this condition, we can approximate  $h_K(v_n)$  by  $(1 - \alpha_{K,2}(v_n))/2$ , as  $n$  increases. Thus, asymptotically, both results coincide, but the expression provided in this paper will allow us to derive an expression of the optimal smoothing parameter.

The issue, under the general AMISE expression given in (5), is that it is not straightforward to know how to derive an explicit expression of the optimal smoothing parameter, unless a specific kernel, such as the von Mises is chosen (see, e.g., Di Marzio et al. 2011). This problem can be solved when the following extra condition is assumed.

- (E3) As  $n \rightarrow \infty$ ,  $R_{K;r,2}(v_n) = Q_{K;r,2} h_n^{-(2r+1)/2} + o(h_n^{-(2r+1)/2})$ , where  $Q_{K;r,2}$  is a constant only depending on the kernel family and the derivative order  $r$ .

Note that under the Condition (E3), the Assumption (A3) simplifies to  $\lim_{n \rightarrow \infty} h_n = 0$  and  $\lim_{n \rightarrow \infty} nh_n^{(2r+1)/2} = \infty$ . Using this last assumption, we can observe the classic variance-bias trade-off, where the bias is reduced if  $h_n$  is “small” and the variance decreases if  $h_n$  is “large”. The optimal smoothing parameter with respect to the AMISE criteria can be obtained using the following corollary of Theorem 1, which is a direct consequence of  $h_K(v) > 0$  (see Section S3 of the Supplementary Material).

**Corollary 1** Consider the Assumptions (A1)–(A3) and (E1)–(E3). Then, for the kernel derivative estimator of order  $r$  (see (2)), we have that, asymptotically, the optimal (AMISE) value of  $h_n$  can be obtained from,

$$h_{K;r;AMISE} = \left( \frac{(2r + 1) Q_{K;r,2}}{n \int_{-\pi}^{\pi} (f^{(r+2)}(\theta))^2 d\theta} \right)^{2/(2r+5)}. \tag{6}$$

Under the previous assumptions, the minimal AMISE of  $\hat{f}_v^{(r)}$  is equal to

$$\inf_{0 \leq v < 1} AMISE \left[ \hat{f}_v^{(r)} \right] = \left( \frac{(2r + 1) Q_{K;r,2}}{n} \right)^{4/(2r+5)}$$

$$\times \frac{2r + 5}{8r + 4} \left( \int_{-\pi}^{\pi} (f^{(r+2)}(\theta))^2 d\theta \right)^{(2r+1)/(2r+5)} \tag{7}$$

**Remark 2** From Corollary 1, it can be observed that the optimal order for circular bandwidth density estimation  $h_{K;0;AMISE}$ , which is  $O(n^{-2/5})$ , does not coincide with the optimal bandwidth order,  $O(n^{-1/5})$ , as presented in Tenreiro (2022). The reason for this discrepancy, as explained in Sect. 2, is their lack of alignment. However, it is worth noting that, for example, in the case of the von Mises kernel (see Sect. 5.1), when translating this result to the optimal concentration  $\kappa$  for density estimation, both theoretical outcomes yield the same asymptotic optimal concentration of order  $O(n^{2/5})$ .

**Remark 3** Note that the AMISE expression in (7) aligns with the one obtained in the linear case (Wand and Jones 1995, Chapter 2), except for terms related to the kernel. If  $K_\ell$  is a scalar kernel with a variance of one, the same AMISE expression in the linear case can be obtained by replacing  $Q_{K;r,2}$  in (7) with  $\int_{-\infty}^{\infty} (K_\ell^{(r)}(x))^2 dx$ .

The optimal mean resultant length  $\nu$  is the solution to the equation  $h_K(\nu) = h_{K;r;AMISE}$ , see Equation (4). Alternatively, to avoid the infinite sum in (4), one can also obtain the optimal  $\nu$  by solving the equation  $\alpha_{K,2}(\nu) = 1 - 2h_{K;r;AMISE}$  (see Remark 1).

As usual, when using data-based plug-in smoothing parameters, the main problem with employing the optimal  $h_{K;r;AMISE}$  in Corollary 1, is that its value depends on the unknown value of  $\int_{-\pi}^{\pi} (f^{(r+2)}(\theta))^2 d\theta$ . A rule-of-thumb smoothing selector can be obtained by replacing  $f$  with a simple and standard density function, such as the von Mises density. One can also follow the Ćwik and Koronacki (1997) approach and replace  $f$  with a mixture model, such as the mixture of von Mises. Both techniques were already proposed and studied in the circular literature when employing the von Mises density as the kernel  $K_\nu$  (see Taylor 2008; Oliveira et al. 2012; García-Portugués 2013). Following Sheather and Jones (1991), an alternative is to estimate density functionals related to  $\int_{-\pi}^{\pi} (f^{(s)}(\theta))^2 d\theta$ . In the next section, we study how to obtain a kernel estimator of this last quantity, in the circular context.

### 4 Estimation of density functionals

As mentioned in the previous section, an issue with using, in practice, the optimal smoothing parameter (6) is its dependence on the unknown quantity  $\int_{-\pi}^{\pi} (f^{(r+2)}(\theta))^2 d\theta$ . In this section, we will see how to estimate this last quantity using kernel techniques. For doing so, we first define the functional

of the form

$$\psi_s = \int_{-\pi}^{\pi} f^{(s)}(\theta) f(\theta) d\theta.$$

Note that under sufficient smoothness assumptions on  $f$  (e.g., the needed conditions to apply integration by parts), we obtain that,

$$\int_{-\pi}^{\pi} (f^{(s)}(\theta))^2 d\theta = (-1)^s \psi_{2s}. \tag{8}$$

Since  $\psi_s = \mathbb{E}(f^{(s)}(\Theta))$ , the following estimator can be employed to estimate the unknown quantity  $\int_{-\pi}^{\pi} (f^{(r+2)}(\theta))^2 d\theta$  on the optimal smoothing parameter (6),

$$\hat{\psi}_{s;\rho} = \frac{1}{n} \sum_{i=1}^n \hat{f}_\rho^{(s)}(\Theta_i) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n L_\rho^{(s)}(\Theta_i - \Theta_j), \tag{9}$$

where  $L_\rho$  and  $\rho$  are a kernel and a mean resultant length parameter. In the following theory, we will assume that  $L_\rho$  and  $\rho$  may differ from  $K_\nu$  and  $\nu$ . However, it is important to note that, in practice, the same kernel is typically employed for both the density (derivative) and the density functionals estimators. Later on, we will observe that, conversely,  $\nu$  and  $\rho$  usually take distinct values, and we will delve into how to derive a data-driven value for  $\rho$ .

Using the estimator (9), a direct plug-in estimator of the smoothing parameter can be obtained from (6), replacing the quantity depending on the true  $f$ , by its estimator, in the following way,

$$h_{K;r;PI} = \left( \frac{(2r + 1) Q_{K;r,2}}{n(-1)^{r+2} \hat{\psi}_{2r+4;\rho}} \right)^{2/(2r+5)} \tag{10}$$

The problem with the direct plug-in estimator (10), is that it still depends on a choice of the pilot mean resultant length  $\rho$ . Below, we establish the asymptotic theory to derive the optimal smoothing parameter of  $\hat{\psi}_{2r+4;\rho}$ . For obtaining that result, the following conditions on the pilot parameter are required.

- (A4) The smoothing parameter satisfies that  $\lim_{n \rightarrow \infty} h_L(\rho_n) = 0$  and  $\lim_{n \rightarrow \infty} n^{-2} R_{L;s,2}(\rho_n) = 0$ .
- (A5)  $\lim_{n \rightarrow \infty} R_{L;s,1}(\rho_n) = \infty$  and  $\lim_{n \rightarrow \infty} n^{-1} R_{L;s,1}(\rho_n) = 0$ .

Note that when replacing the kernel  $K_\nu$  by  $L_\rho$ , the mean resultant length  $\nu_n$  by  $\rho_n$ , and the order of the derivative  $r$  by  $s$ , Assumption (A3) implies Condition (A4). Below, we provide the AMSE for the estimator  $\hat{\psi}_{s;\rho}$  (see Section S4 of the Supplementary Material for a formal proof).

**Theorem 2** *If  $f$  satisfies Assumption (A1) (replacing  $r$  by  $s$ ), the kernel  $L_\rho$  and  $\rho_n$  verifies (A2), (A4) and (A5); then, for an even value of  $s$ ,*

$$AMSE[\hat{\psi}_{s;\rho}] = O(n^{-2}R_{L;s,1}^2(\rho_n)) + O(h_L^2(\rho_n)) + O(n^{-2}R_{L;s,2}(\rho_n)) + O(n^{-1}). \tag{11}$$

*If, in addition, Conditions (E1) and (E2) (using  $L_\rho$ , instead of  $K_v$ ;  $\rho_n$ , instead  $v_n$ ; and  $r = s$ ) are verified, we have*

$$AMSE[\hat{\psi}_{s;\rho}] = \left(n^{-1}R_{L;s,1}(\rho_n) + \frac{1}{2}h_L(\rho_n)\psi_{s+2}\right)^2 + 2n^{-2}\psi_0R_{L;s,2}(\rho_n) + 4n^{-1}\left(\int_{-\pi}^{\pi} (f^{(s)}(\theta))^2 f(\theta)d\theta - \psi_s^2\right). \tag{12}$$

As for Corollary 1, in order to derive a simpler expression of the AMSE and the optimal value of  $h_n \equiv h_L(\rho_n)$ , we will assume the following extra conditions:

- (E4)  $R_{L;s,1}(\rho_n) = Q_{L;s,1}h_n^{-(s+1)/2} + o(h_n^{-(s+1)/2})$ , as  $n \rightarrow \infty$ , where  $Q_{L;s,1}$  is a constant only depending on the kernel and the even number  $s$ . The sign of  $Q_{L;s,1}$  is equal to the sign of  $(-1)^{s/2}$ .
- (E5)  $\int_{-\pi}^{\pi} \theta^4 L_{\rho_n}(\theta)d\theta = o(h_n^{5/4})$ , as  $n \rightarrow \infty$ .

From the AMSE expression in (12), we can see that the optimal  $h_n$  value, in terms of AMSE, will depend on the relation between  $R_{L;s,1}(\rho_n)$ ,  $R_{L;s,2}(\rho_n)$ , and  $h_L(\rho_n)$ . Conditions (E3) and (E4) help to establish this relation, from which the following corollary is derived (see Section S4 of the Supplementary Material).

**Corollary 2** *Consider the assumptions of Theorem 2, Conditions ((E3); replacing  $K_v$  by  $L_\rho$ ; using  $\rho_n$ , instead  $v_n$ ; and  $r = s$ ), and (E4). Then, for the kernel estimator of  $\psi_s$  (see (9)), we have that, asymptotically, the optimal value, in terms of the AMSE expression, of  $h_n$  can be obtained from,*

$$h_{L;s;AMSE} = \left(-\frac{2Q_{L;s,1}}{n\psi_{s+2}}\right)^{2/(s+3)}. \tag{13}$$

*Under the previous assumptions, if Condition (E5) is also assumed, the minimal AMSE for (9) is of order  $O(n^{-\min(5,s+3)/(s+3)})$ . If  $s$  is an even number greater than 2, the minimal AMSE would be equal to*

$$\inf_{0 \leq \rho < 1} AMSE[\hat{\psi}_{s;\rho}] = 2\psi_0 Q_{K;r,2} \times \left(-\frac{\psi_{s+2}}{2Q_{L;s,1}}\right)^{(2s+1)/(s+3)} n^{-5/(s+3)}. \tag{14}$$

*If  $s = 0$ , the minimal AMSE is equal to*

$$\inf_{0 \leq \rho < 1} AMSE[\hat{\psi}_{s;\rho}] = 4n^{-1} \times \left(\int_{-\pi}^{\pi} (f^{(s)}(\theta))^2 f(\theta)d\theta - \psi_s^2\right). \tag{15}$$

*When  $s = 2$ , the minimal AMSE is equal to the sum of the right-hand sides of (14) and (15).*

Note that Condition (E5) is only assumed to obtain the same AMSE optimal order as in the linear case. If that condition is not fulfilled, then, when  $s > 0$ , the leading term in the AMSE could be of a larger order (see Section S4 of the Supplementary Material). Another important consideration is that the sign of  $\psi_{s+2}$  is the same as that of  $(-1)^{s/2+1}$  and, using Condition (E4), it also coincides with the sign of  $-Q_{L;s,1}$ . Therefore, we always have that  $h_{L;s;AMSE} \geq 0$  in (13).

Using the quantity  $h_{L;2r+4;AMSE}$  in (13), one can obtain the pilot mean resultant length  $\rho$  needed to derive the direct plug-in estimator (10). The issue, as in the linear case, is that  $h_{L;2r+4;AMSE}$  still depends on the unknown value of the functional  $\psi_{2r+6}$ . We comment on how to overcome this difficulty in Sect. 6.

### 5 The kernel choice

Theoretical results in Sects. 3 and 4 provide mathematical support for deriving the optimal smoothing parameter. However, we still did not discuss how to obtain the plug-in mean resultant length parameter  $v_{r;AMSE}$  from  $h_{K;r;AMSE}$ . As mentioned in Sect. 3, we must solve the equation  $h_K(v) = h_{K;r;AMSE}$ . We also need to obtain the values of  $Q_{K;r,1}$  and  $Q_{K;r,2}$  in (6) and (13).

In this section, we study what happens when the “most common” circular models are employed as kernels. For doing so, we restrict our attention to the four standard choices of circular densities (see Mardia and Jupp 1999, Section 3.5): cardioid, von Mises, wrapped normal, and wrapped Cauchy. Although all of them satisfy Assumption (A2), the cardioid kernel,  $K_{c;v}(\theta) = (1 + 2v \cos(\theta))/(2\pi)$ , with  $|v| < 1/2$ , does not meet Assumption (A3) and will be hence discarded from the analysis. This can be seen by observing that, for any  $v$ ,  $h_{Kc}(v) = \pi^2/3 - 4v \geq \pi^2/3 - 2$ . The other three circular densities are studied in Sects. 5.1 and 5.2.

Once the behaviour of the standard kernels is studied, a second question to discuss is which is the optimal kernel in the circular context. As in Muller (1984), we study the kernel optimality in terms of the AMISE expression in (7). Fixing  $f$  and the sample size  $n$ , in Sect. 5.3, we obtain the circular kernel that minimizes the AMISE.

### 5.1 Von Mises and wrapped normal kernels

First, denote by  $\mathcal{I}_j$  to the modified Bessel function of the first kind and order  $j$ , the density expression of the von Mises (VM) kernel is

$$K_{\text{VM};\nu}(\theta) = \frac{\exp(\kappa \cos(\theta))}{(2\pi \mathcal{I}_0(\kappa))} = \frac{1}{2\pi} \left( 1 + 2 \sum_{j=1}^{\infty} \frac{\mathcal{I}_j(\kappa)}{\mathcal{I}_0(\kappa)} \cos(j\theta) \right),$$

where  $\nu = \mathcal{I}_1(\kappa)/\mathcal{I}_0(\kappa)$ , being  $\kappa \geq 0$ . The density expression of the wrapped normal (WN), with a mean resultant length  $\nu \in [0, 1]$ , is

$$K_{\text{WN};\nu}(\theta) = \frac{1}{2\pi} \left( 1 + 2 \sum_{j=1}^{\infty} \nu^{j^2} \cos(j\theta) \right).$$

For both kernels, if  $\lim_{n \rightarrow \infty} \nu_n = 1$  (equivalently  $\lim_{n \rightarrow \infty} \kappa_n = \infty$ , for the von Mises kernel), then  $\lim_{n \rightarrow \infty} h_K(\nu_n) = 0$ . Even more, it is easy to show that, in that setting, Conditions (E1) and (E2) are satisfied. Thus, as  $n \rightarrow \infty$ , we obtain the following asymptotic approximations:

$$h_{K_{\text{VM}}(\nu_n)} = \frac{1}{2} \left( 1 - \frac{\mathcal{I}_2(\kappa_n)}{\mathcal{I}_0(\kappa_n)} \right) = \frac{1}{\kappa_n},$$

$$h_{K_{\text{WN}}(\nu_n)} = \frac{1}{2} \left( 1 - \nu_n^4 \right).$$

From the previous equalities, we can see that  $\kappa_n$  or  $\nu_n$  are easily derived, after computing the optimal/plug-in value of  $h_n$  (see, e.g., (6) or (10)).

$$\kappa_n \stackrel{\text{VM}}{=} h_n^{-1} \quad \text{and} \quad \nu_n \stackrel{\text{WN}}{=} (1 - h_n)^{1/4}. \tag{16}$$

Again, considering  $\lim_{n \rightarrow \infty} \nu_n = 1$ , it can be seen that Conditions (E3)–(E5) are satisfied for the von Mises and wrapped normal kernels. For both kernels and a non-negative integer number  $r$ , the values of the constants in (E3) and (E4) are the following ones:

$$Q_{K;r,1} = (-1)^{r/2} \frac{r!}{2^{r/2}(r/2)!\sqrt{2\pi}} \quad (r \text{ being even}),$$

$$Q_{K;r,2} = \frac{(2r)!}{2^{2r+1}r!\pi^{1/2}}. \tag{17}$$

We can see that the AMSE and AMISE results derived in Sects. 3 and 4 can be easily computed in practice from these last quantities. These optimal asymptotic results will coincide for both the von Mises and wrapped normal kernels.

### 5.2 Wrapped Cauchy kernel

The expression of the wrapped Cauchy kernel, with a mean resultant length  $\nu \in [0, 1]$ , is given by

$$K_{\text{WC};\nu}(\theta) = \frac{1}{2\pi} \left( 1 + 2 \sum_{j=1}^{\infty} \nu^j \cos(j\theta) \right).$$

From the density expression, we derive that  $h_{K_{\text{WC}}}(\nu) = \pi^2/3 + 4\text{Li}_2(-\nu)$ , where  $\text{Li}_s$  is the polylogarithm of order  $s$ . The last implies  $\lim_{n \rightarrow \infty} h_{K_{\text{WC}}}(\nu_n) = 0$  if  $\lim_{n \rightarrow \infty} \nu_n = 1$ . Also, the expressions of  $R_{K_{\text{WC}};r,2}(\nu)$  can be obtained in terms of a polylogarithm,

$$R_{K_{\text{WC}};r,2}(\nu) = \begin{cases} (2\pi)^{-1} (1 + 2\text{Li}_0(\nu^2)) & \text{if } r = 0, \\ \pi^{-1} \text{Li}_{-2r}(\nu^2) & \text{otherwise,} \end{cases} = O(h_n^{-(2r+1)}), \text{ as } n \rightarrow \infty. \tag{18}$$

If we consider a value of  $h_n$  such that  $\lim_{n \rightarrow \infty} h_n = 0$  and  $\lim_{n \rightarrow \infty} nh_n^{2r+1} = \infty$ , Condition (A3) is satisfied. Therefore, from Theorem 1, if  $f$  also satisfies Assumption (A1), we obtain the following AMISE for the wrapped Cauchy kernel,

$$\text{AMISE} \left[ \hat{f}_\nu^{(r)} \right] = O(h_n^2) + O(n^{-1}h_n^{-(2r+1)}). \tag{19}$$

The AMISE expression (19) would be minimized if  $h_n$  is of order  $n^{-1/(2r+3)}$ . Using the previous result, we obtain that the AMISE order of the wrapped Cauchy kernel is worse than that obtained with the von Mises or the wrapped normal kernel (see Corollary 1). In particular, its minimal AMISE is equal to

$$\inf_{0 \leq \nu < 1} \text{AMISE} \left[ \hat{f}_\nu^{(r)} \right] = O(n^{-2/(2r+3)}).$$

When searching for an explicit expression of the optimal smoothing parameter, one could be tempted to combine (5) and (18). But note that this cannot be done as Condition (E1) is not verified (see Section S3.1). Thus, the explicit expression of the asymptotic bias cannot be obtained following the steps in Section S3 of the Supplementary Material. This implies that the AMISE does not have a closed expression as given in (5). On the contrary, Tsuruta and Sagae (2017) were able to obtain an optimal smoothing parameter for the wrapped Cauchy kernel from the results by Di Marzio et al. (2009). Nevertheless, it should be noted that Di Marzio et al. (2009) results must not be employed for this kernel as Condition (E1) is not satisfied (see Section S3.1 for further details).

### 5.3 Wrapped bounded-support kernels

To find the optimal kernels, consider a wrapped kernel  $K_v(\theta) = \sum_{\ell=-\infty}^{\infty} K_{X;\lambda}(\theta + 2\ell\pi)$ , whose associated linear density  $K_{X;\lambda}$  has bounded support, i.e.,  $K_{X;\lambda}(x) = 0$  if  $|x| > \lambda$ . Then,  $K_v(\theta) = K_{X;\lambda}(\theta)$ , for all  $\theta \in [-\pi, \pi)$ , when  $\lambda < \pi$ . In that case,  $h = \int_{-\lambda}^{\lambda} x^2 K_{X;\lambda}(x) dx$  and  $R_{K;r,2}(v) = \int_{-\lambda}^{\lambda} (K_{X;\lambda}^{(r)}(x))^2 dx$ .

Consider the asymptotic case for linear bounded-support kernels, i.e.,  $\lim_{n \rightarrow \infty} \lambda_n = 0$ . Fixing the unknown linear density function  $f_X$ , Muller (1984) gives explicit expressions of the bounded-support kernels minimizing the AMISE when estimating the density function and its derivatives, in the linear case.

Assume now that the wrapped bounded-support kernel satisfies the assumptions of Corollary 1. Except for the values not depending on the kernel, the minimal AMISE of  $\hat{f}_v^{(r)}$  in (7) is equal to the minimal AMISE obtained in the linear case. As a consequence, Theorem 2.4 of Muller (1984) can be employed to see that the optimal kernels in terms of AMISE coincide with the wrapped version of those employed on linear kernel estimation.

The latter means that the optimal kernel for circular density estimation is the wrapped Epanechnikov, whose associated linear density is  $K_{X;\lambda}(x) = 3(1 - (x/\lambda)^2)/(4\lambda)$ . When  $\lambda < \pi$ , the mean resultant length is  $v = (3 \sin(\lambda) - 3\lambda \cos(\lambda))/\lambda^3$ ,  $h = \lambda^2/5$ , and  $R_{K;0,2}(v) = 3/(5\lambda)$ . Thus, for circular density estimation with the wrapped Epanechnikov, the optimal (AMISE) value of  $h_n$  is obtained by taking  $Q_{K;0,2} = 3/(5\sqrt{5})$  in (6). Regardless of the employed kernel, the minimum achievable AMISE for estimating a circular density function is given by:

$$\inf_{0 \leq v < 1} \text{AMISE} \left[ \hat{f}_v \right] = \frac{5}{4} \left( \frac{3}{5\sqrt{5n}} \right)^{4/5} \times \left( \int_{-\pi}^{\pi} \left( f^{(2)}(\theta) \right)^2 d\theta \right)^{1/5}.$$

This is the AMISE obtained with the wrapped Epanechnikov kernel. Note that this optimality is established among kernels that satisfy the conditions outlined in Corollary 1. The optimality of the Epanechnikov kernel for circular density estimation is also established by Tenreiro (2022) when restricting to kernels that satisfy alternative necessary assumptions. For the derivatives of the density, we refer to Muller (1984). There, we can see, e.g., that the wrapped Biweight would be the optimal kernel for the first derivative of  $f$ .

### 6 Plug-in smoothing parameters

In this section, we will study how to implement, in practice, the plug-in smoothing parameter  $h_{K;r;\text{PI}}$  in (10). As mentioned in Sect. 4, the issue of directly employing (10) is that the AMISE-optimal smoothing parameter for  $\hat{\psi}_{s;\rho}$  will always depend on an unknown value of  $\psi_{s+2}$ . A way to overcome this difficulty is to provide an  $l$ -stage direct plug-in smoothing selector. This procedure consists in estimating  $\psi_s$  with  $\hat{\psi}_{s+2;\rho}$ , in an iterative process, until some point in which we replace  $\psi_{2r+2l+4}$  by its value obtained with a simple density (see Sect. 6.1 for more details). An alternative selector can be computed by noting that the smoothing parameter for  $\hat{f}^{(r)}$  can be obtained as a function of the smoothing parameter for  $\hat{\psi}_{2r+4}$ . This allows us to construct a *solve-the-equation* rule. In Sect. 6.2, we discuss in more detail how this selector is derived.

#### 6.1 Two-stage direct plug-in smoothing selector

The procedure to obtain the  $l$ -stage direct plug-in smoothing selector consists in, at stage 0, using a simple rule of thumb to compute the smoothing parameter of  $\hat{\psi}_{2r+4+2l;\rho}$ . Once this initial step is achieved, from that estimator of  $\psi_{2r+4+2l}$ , the following functional estimators are derived in an iterative process (see Algorithm 1 for details).

Two decisions remain from this brief explanation: the number of stages  $l$  and which reference density should be employed at stage 0. Regarding the number of stages  $l$ , we suggest employing  $l = 2$  for two reasons. First,  $l = 2$  is a common choice in the linear case (see, e.g., Wand and Jones 1995, Section 3.6). A second reason was obtained when replicating the simulation study in Sect. 7, with  $l = 3$ . In that case, similar results were achieved with respect to those obtained with  $l = 2$ , with the added inconvenience of the extra computational time. The only scenario in which we would recommend increasing the value of  $l$  is when complex multimodal patterns are expected.

As mentioned before, at stage 0, a reference density is needed to compute the smoothing parameter of  $\hat{\psi}_{2r+2l+4;\rho}$ . Here, in the same spirit of the original rule of thumb proposed by Taylor (2008), a natural selection would be to replace  $\psi_{2r+2l+4}$  in (13) by the quantity obtained when assuming that the true density follows a von Mises. In the circular case, the main pitfall of employing that simple strategy is that a uniform estimation of the density can be obtained even if the true distribution is not uniform. This occurs, for example, when considering distributions with antipodal symmetry (see also Oliveira et al. 2012, for further discussion on this topic). The consequence would be to have a value of  $\hat{\psi}_{2r+2l+4;\rho}$  close to zero, which leads to a “large” value of the smoothing parameter at the next stage  $h_{L;2r+2l+2;\text{PI}}$ .

To avoid that last issue, while still having a simple model, one can use as the reference density the following mixture of  $M$  von Mises, all having the same concentration parameter  $\kappa \geq 0$ :

$$f_{\text{MvM}}(\theta; M; \boldsymbol{\mu}, \kappa, \boldsymbol{w}) = \frac{1}{(2\pi \mathcal{I}_0(\kappa))} \times \sum_{m=1}^M w_m \exp(\kappa \cos(\theta - \mu_m)), \tag{20}$$

where the parameters  $\mu_m \in [-\pi, \pi)$ ,  $w_m \in [0, 1]$ , for all  $m \in \{1, \dots, M\}$ , and with  $\sum_{m=1}^M w_m = 1$ . The value of  $\psi_{2r+2l+4}$  is calculated from the density (20), replacing its parameters by their maximum likelihood estimates obtained from the sample. An algorithm providing the maximum likelihood estimates for the density (20) was implemented in the R (R Core Team 2022) library `movMF` by Hornik and Grün (2014). In practice, following Oliveira et al. (2012), the value of  $M$  can be chosen, using the Akaike Information Criterion (AIC), by comparing the results obtained with the mixtures (20) of  $M = 1, \dots, M_{\text{max}}$  components.

In Algorithm 1, we summarize the steps that are needed to obtain our proposed two-stage direct plug-in smoothing parameter. There, for simplicity, we use the same kernel for both  $K$  and  $L$ , and this kernel satisfies Assumptions (A1)–(A5) and the extra Conditions (E1)–(E4).

**Algorithm 1** Two-stage direct plug-in smoothing selector.

Step 1: Use a rule of thumb to obtain the estimator of  $\psi_{2r+8}$ , namely  $\hat{\psi}_{2r+8;RT}$ . This can be achieved as follows.

Step 1.a: For every  $M = 1, \dots, M_{\text{max}}$ ; obtain the maximum likelihood estimators of the parameters in the mixture of  $M$  von Mises, all having the same concentration parameter, i.e., of  $\boldsymbol{\mu}, \kappa$ , and  $\boldsymbol{w}$  in (20).

Step 1.b: Select the number of components in the mixture,  $M_{\text{AIC}}$ , employing the AIC.

Step 1.c: Compute the value  $\hat{\psi}_{2r+8;RT}$ , using the density  $f_{\text{MvM}}(\cdot; M_{\text{AIC}}; \hat{\boldsymbol{\mu}}, \hat{\kappa}, \hat{\boldsymbol{w}})$  (see Equality (8)).

Step 2: Estimate  $\psi_{2r+6}$  using the estimator  $\hat{\psi}_{2r+6;\rho_1}$ , where  $\rho_1$  is the plug-in mean resultant length relying on  $\hat{\psi}_{2r+8;RT}$ .

Step 2.a: Obtain  $h_{K;2r+6;\text{PI}}$  from (13), replacing  $\psi_{2r+8}$  by  $\hat{\psi}_{2r+8;RT}$ .

Step 2.b: The value  $\rho_1$  is the one that satisfies  $h_K(\rho_1) = h_{K;2r+6;\text{PI}}$ , see Equation (4).

Step 3: Estimate  $\psi_{2r+4}$  using the estimator  $\hat{\psi}_{2r+4;\rho_2}$ , where  $\rho_2$  is the plug-in mean resultant length relying on  $\hat{\psi}_{2r+6;\rho_1}$ .

Step 4: Compute the AMISE-optimal smoothing parameter for  $\hat{f}^{(r)}$ , relying on  $\hat{\psi}_{2r+4;\rho_2}$ .

Step 4.a: Obtain the two-stage direct plug-in smoothing parameter  $h_{K;r;\text{PI}}$  from (10), with the pilot mean resultant length  $\rho_2$ .

Step 4.b: The selected mean resultant length  $\nu_{\text{DPI}}$  is obtained from the value that satisfies  $h_K(\nu_{\text{DPI}}) = h_{K;r;\text{PI}}$ .

The uniform distribution belongs to all the reference distributions mentioned before. Thus, in practice, the denominator of (10) or (13) could be equal to zero. If that occurs, we suggest directly returning a value of the mean resultant length  $\nu_{\text{DPI}} = 0$ , which would correspond to the uniform estimation of the density.

Given a finite value of  $n$  in (10) or (13), the equation  $h_K(\nu) = h_{K;r;\text{PI}}$ , or its approximation  $\alpha_{K,2}(\nu) = 1 - 2h_{K;r;\text{AMISE}}$ , cannot be solved for “large” values of  $h_{K;r;\text{PI}}$ . The reason is that  $h_K(\nu)$  and  $\alpha_{K,2}(\nu)$  are non-negative and bounded for any value of  $\nu$ . Since a “large” value of  $h_K(\nu)$  corresponds to  $\nu = 0$ , if the equation  $h_K(\nu) = h_{K;r;\text{AMISE}}$  cannot be solved, we also suggest employing the uniform as the density estimator.

### 6.2 Solve-the-equation plug-in smoothing selector

An alternative to the previous smoothing selector is the solve-the-equation rule. This rule consists in searching for the smoothing parameter that satisfies

$$h_{K;r;\text{STE}} = \left( \frac{(2r+1)Q_{K;r,2}}{n(-1)^{r+2}\hat{\psi}_{2r+4;\rho_{K;r;\text{STE}}}} \right)^{2/(2r+5)}, \tag{21}$$

with  $h_K(\rho_{K;r;\text{STE}}) = \gamma(h_{K;r;\text{STE}})$ .

Now, the smoothing parameter for  $\hat{\psi}_{2r+4}$  is a function of the smoothing parameter for  $\hat{f}^{(r)}$ . We suggest taking the function  $\gamma$  by looking at the relation between the optimal smoothing parameter of these two estimators,

$$\eta_{L;2r+4;\text{AMSE}} = \left( (-1)^{r+1} \frac{2Q_{L;2r+4,1}}{(2r+1)Q_{K;r,2}} \right)^{2/(2r+7)} \times (\psi_{2r+4}/\psi_{2r+6})^{2/(2r+7)} h_{K;r;\text{AMISE}}^{(2r+5)/(2r+7)}.$$

Using a plug-in rule, this last relation suggests taking the following function:

$$\gamma(h) = \left( (-1)^{r+1} \frac{2Q_{L;2r+4,1}}{(2r+1)Q_{K;r,2}} \right)^{2/(2r+7)} \times \left( \hat{\psi}_{2r+4;\rho_1} / \hat{\psi}_{2r+6;\rho_2} \right)^{2/(2r+7)} h^{(2r+5)/(2r+7)}. \tag{22}$$

The two smoothing parameters of the density functional estimators,  $\hat{\psi}_{2r+4}$  and  $\hat{\psi}_{2r+6}$ , can be obtained from (13).

Those smoothing parameters will depend on some other density functionals,  $\psi_{2r+6}$  and  $\psi_{2r+8}$ . Following Sheather and Jones (1991), we suggest estimating these two new functionals with a rule of thumb. In Algorithm 2, we summarize the steps that are needed to obtain our proposed solve-the-equation plug-in smoothing parameter. Again, for simplicity, we use the same kernel for both  $K$  and  $L$ , a kernel that satisfies Assumptions (A1)–(A5) and the extra Conditions (E1)–(E4).

**Algorithm 2** Solve-the-equation plug-in smoothing selector.

- Step 1: Use a rule of thumb to obtain the estimator of  $\psi_{2r+6}$  and  $\psi_{2r+8}$ , namely  $\hat{\psi}_{2r+6;RT}$  and  $\hat{\psi}_{2r+8;RT}$ . This can be done as in Step 1 of Algorithm 1.
- Step 2: Estimate  $\psi_{2r+4}$  and  $\psi_{2r+6}$  using the estimators  $\hat{\psi}_{2r+4;\rho_1}$  and  $\hat{\psi}_{2r+6;\rho_2}$ , where  $\rho_1$  is the plug-in mean resultant length relying on  $\hat{\psi}_{2r+6;RT}$ , and  $\rho_2$  the one relying on  $\hat{\psi}_{2r+8;RT}$ .
- Step 3: Employ the estimators  $\hat{\psi}_{2r+4;\rho_1}$  and  $\hat{\psi}_{2r+6;\rho_2}$  to obtain the function  $\gamma$  in (22).
- Step 4: Using the  $\gamma$  function of Step 3, select the value of  $h_{K;r;STE}$  by solving Equation (21). The solve-the-equation mean resultant length  $v_{STE}$  is obtained from the value that satisfies  $h_K(v_{STE}) = h_{K;r;STE}$ .

Note that two extra functional estimators must be computed in Step 1 of Algorithm 2. For this reason, the obtained smoothing parameter could be considered as a *two-stages solve-the-equation plug-in selector*. As for the direct rule, the number of stages could be increased to estimate the density functionals. In particular, a rule of thumb can be employed to estimate  $\psi_{2r+10}$  or  $\psi_{2r+12}$  and then, in an iterative process, the values of  $\rho_1$  and  $\rho_2$  are obtained (as in Steps 2 and 3 of Algorithm 1). The simulation study in Sect. 7 was carried out under the same conditions, using three (relying on  $\hat{\psi}_{2r+8;RT}$  and  $\hat{\psi}_{2r+10;RT}$ ) and four (relying on  $\hat{\psi}_{2r+10;RT}$  and  $\hat{\psi}_{2r+12;RT}$ ) stages. The results in practice were similar to those obtained with Algorithm 2, with the drawback of the extra computational time.

### 7 Simulation study

In Section S1 of the Supplementary Material, we performed a simulation study to analyse the performance of the direct two-stage plug-in smoothing parameter (see Algorithm 1) and the solve-the-equation plug-in smoothing selector (see Algorithm 2). We focused only on the density estimation case ( $r = 0$  in (2)), with  $K_v$  being the von Mises kernel. The reason is that their effectiveness can be compared with other smoothing parameters proposed in the literature. But note that our data-driven smoothing parameters could be employed to estimate any derivative and with other kernels.

In particular, in this framework, slightly better results are expected by employing the wrapped Epanechnikov kernel (see Sect. 5.3).

A complete report on the results obtained when comparing the proposed smoothing selectors with other data-based smoothing selectors available in the circular literature (see Sect. 1) is provided in Section S1. In the current section, Table 1 presents the results obtained, for  $n = 100$ , using the proposed smoothing selectors and other data-driven rules available in the literature. Table 1 contains the average integrated square error (ISE) computed from 1000 samples of the reference Models M5–M20 that can be found in Oliveira et al. (2012). Models M1–M4 were not included here as they can be well-approximated by a von Mises density. In those cases, as expected, the rule of thumb provides the best results. The full discussion of the obtained results is provided in Section S1. Below, as a summary of our findings, we provide our recommendations. In this summary, in parentheses, we provide the abbreviated names of the analysed smoothing rules. The same notation is used in the tables presenting the simulation results.

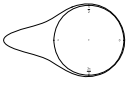
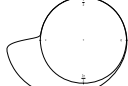
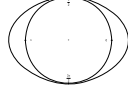
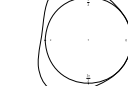
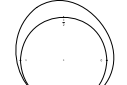

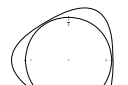
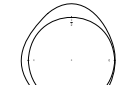
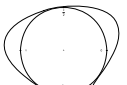
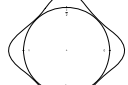
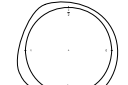

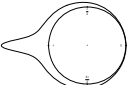
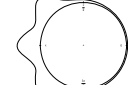

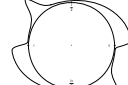
- For small sample sizes (say  $n = 50$ ), employ the proposed direct plug-in rule with  $M_{max} = 1 (v_{DPI}; 1)$ . The only exception to this recommendation is if there could be evidence of the density function being multimodal and  $k$ -fold rotational symmetric. In that case, employ the solve-the-equation plug-in smoothing selector with  $M_{max} = 1 (v_{STE}; 1)$ .
- For moderate sample sizes (say  $n = 100$  or  $n = 250$ ), employ the solve-the-equation plug-in smoothing selector with  $M_{max} = 1 (v_{STE}; 1)$ . The direct plug-in rule with  $M_{max} = 1 (v_{DPI}; 1)$  is also recommendable if there is evidence of unimodality.
- For large sample sizes (say  $n = 1000$ ), employ the Fourier series-based plug-in approach of Tenreiro (2022) ( $v_{PT}$ ).

### 8 Real data application

In this section, we revisit the car accident data that can be found in Ameijeiras-Alonso and Crujeiras (2022). This real dataset consists of the time of the day (at a resolution of one minute) at which the car crash happened in El Paso County (Texas, USA) in 2018. A total of 85 observations were recorded on the webpage of the National Highway Traffic Safety Administration of the United States.

In Ameijeiras-Alonso and Crujeiras (2022), several features of the shape of the distribution are analysed with different inferential tools. The conclusion is that the density is unimodal and (reflectively) asymmetric. More specifically among several parametric models studied by Ameijeiras-

**Table 1** Average ISE ( $\times 100$ ) and standard deviations ( $\times 100$ , in parentheses) computed from 1000 samples of sample size  $n = 100$  for Models M5–M20 of Oliveira et al. (2012)

Model	M5	M6	M7	M8
				
$\nu_{DPI; 1}$	3.411 (1.896)	2.639 (1.054)	1.437 (0.617)	1.393 (0.745)
$\nu_{DPI; 5}$	<b>3.138 (1.783)</b>	2.874 (1.398)	1.268 (0.684)	1.466 (0.838)
$\nu_{STE; 1}$	3.188 (2.162)	<b>2.614 (1.122)</b>	<b>1.172 (0.606)</b>	<b>1.384 (0.752)</b>
$\nu_{RTT}$	8.587 (2.629)	3.173 (0.962)	10.554 (0.368)	3.727 (0.678)
$\nu_{RTGP}$	7.664 (2.467)	3.055 (0.932)	8.88 (1.237)	2.708 (0.679)
$\nu_{PIO}$	8.554 (2.638)	3.102 (0.916)	10.547 (0.394)	3.683 (0.678)
$\nu_{PIGP}$	3.201 (1.863)	2.722 (1.181)	1.183 (0.644)	1.387 (0.731)
$\nu_{PIT}$	3.431 (1.982)	3.036 (1.721)	1.3 (0.831)	1.438 (1.004)
$\nu_{LCV}$	6.827 (2.769)	2.859 (1.241)	1.259 (0.69)	1.49 (0.821)
Model	M9	M10	M11	M12
				
$\nu_{DPI; 1}$	0.739 (0.458)	2.572 (0.87)	2.638 (0.905)	1.191 (0.519)
$\nu_{DPI; 5}$	0.818 (0.544)	2.726 (1.058)	1.497 (0.654)	1.188 (0.594)
$\nu_{STE; 1}$	0.82 (0.526)	<b>2.552 (0.927)</b>	<b>1.405 (0.6)</b>	<b>1.082 (0.533)</b>
$\nu_{RTT}$	0.781 (0.51)	2.895 (0.792)	6.485 (0.014)	4.265 (0.523)
$\nu_{RTGP}$	<b>0.699 (0.443)</b>	2.812 (0.798)	6.318 (0.133)	2.904 (0.577)
$\nu_{PIO}$	0.81 (0.523)	2.88 (0.827)	6.485 (0.015)	4.254 (0.522)
$\nu_{PIGP}$	0.749 (0.486)	2.814 (0.926)	1.538 (1.02)	1.15 (0.639)
$\nu_{PIT}$	0.971 (0.908)	2.988 (1.508)	1.461 (0.795)	1.219 (0.735)
$\nu_{LCV}$	0.809 (0.536)	3.08 (1.038)	1.492 (0.664)	1.171 (0.626)
Model	M13	M14	M15	M16
				
$\nu_{DPI; 1}$	3.458 (1.045)	6.393 (1.207)	0.768 (0.246)	7.865 (0.14)
$\nu_{DPI; 5}$	1.941 (0.824)	2.061 (0.817)	0.86 (0.403)	2.611 (1.247)
$\nu_{STE; 1}$	<b>1.824 (0.708)</b>	<b>1.904 (0.731)</b>	0.843 (0.374)	<b>2.324 (0.805)</b>
$\nu_{RTT}$	10.874 (0.125)	8.182 (0.043)	0.85 (0.153)	7.837 (0.045)
$\nu_{RTGP}$	10.188 (0.552)	8.212 (0.073)	<b>0.762 (0.193)</b>	7.912 (0.136)
$\nu_{PIO}$	10.869 (0.132)	8.183 (0.043)	0.839 (0.156)	7.839 (0.067)
$\nu_{PIGP}$	1.825 (0.734)	2.119 (1.332)	0.777 (0.227)	3.321 (2.314)
$\nu_{PIT}$	1.952 (0.969)	1.978 (0.916)	0.961 (0.557)	2.325 (0.917)
$\nu_{LCV}$	1.856 (0.732)	1.969 (0.78)	0.971 (0.552)	2.34 (0.849)
Model	M17	M18	M19	M20
				
$n = 100$				
$\nu_{DPI; 1}$	5.196 (1.59)	3.014 (0.64)	2.569 (0.581)	8.133 (0.815)
$\nu_{DPI; 5}$	4.695 (1.767)	2.982 (0.745)	<b>2.48 (0.749)</b>	<b>3.491 (0.977)</b>
$\nu_{STE; 1}$	4.36 (1.822)	<b>2.59 (0.958)</b>	2.483 (0.62)	4.067 (0.893)
$\nu_{RTT}$	7.93 (1.198)	3.563 (0.493)	3.875 (0.639)	10.961 (0.053)
$\nu_{RTGP}$	7.382 (1.148)	3.397 (0.484)	3.292 (0.545)	10.582 (0.258)
$\nu_{PIO}$	7.906 (1.209)	3.561 (0.482)	3.888 (0.657)	10.96 (0.053)
$\nu_{PIGP}$	6.258 (2.413)	3.413 (0.585)	2.56 (0.701)	4.028 (1.694)
$\nu_{PIT}$	<b>4.288 (1.82)</b>	2.998 (1.069)	2.718 (1.026)	3.644 (1.176)
$\nu_{LCV}$	5.695 (1.949)	2.976 (0.83)	2.519 (0.686)	3.514 (0.899)

In bold, smallest average ISE. Smoothing parameters: proposed direct plug-in rule with  $M_{\max} = 1$  ( $\nu_{DPI;1}$ ) and with  $M_{\max} = 5$  ( $\nu_{DPI;5}$ ), solve-the-equation plug-in selector with  $M_{\max} = 1$  ( $\nu_{STE;1}$ ), rule of thumb of Taylor (2008) ( $\nu_{RTT}$ ), rule of thumb of García-Portugués (2013) ( $\nu_{RTGP}$ ), plug-in rule of Oliveira et al. (2012) ( $\nu_{PIO}$ ), exact plug-in rule of García-Portugués (2013) ( $\nu_{PIGP}$ ), plug-in rule of Tenreiro (2022) ( $\nu_{PIT}$ ), and likelihood cross-validation of Hall et al. (1987) ( $\nu_{LCV}$ )

Alonso and Crujeiras (2022), the conclusion is that the “best” parametric fitting is achieved by the wrapped skew normal density of Pewsey (2000). A representation of the fitted wrapped skew normal density is provided in Fig. 1 (top, thick solid line).

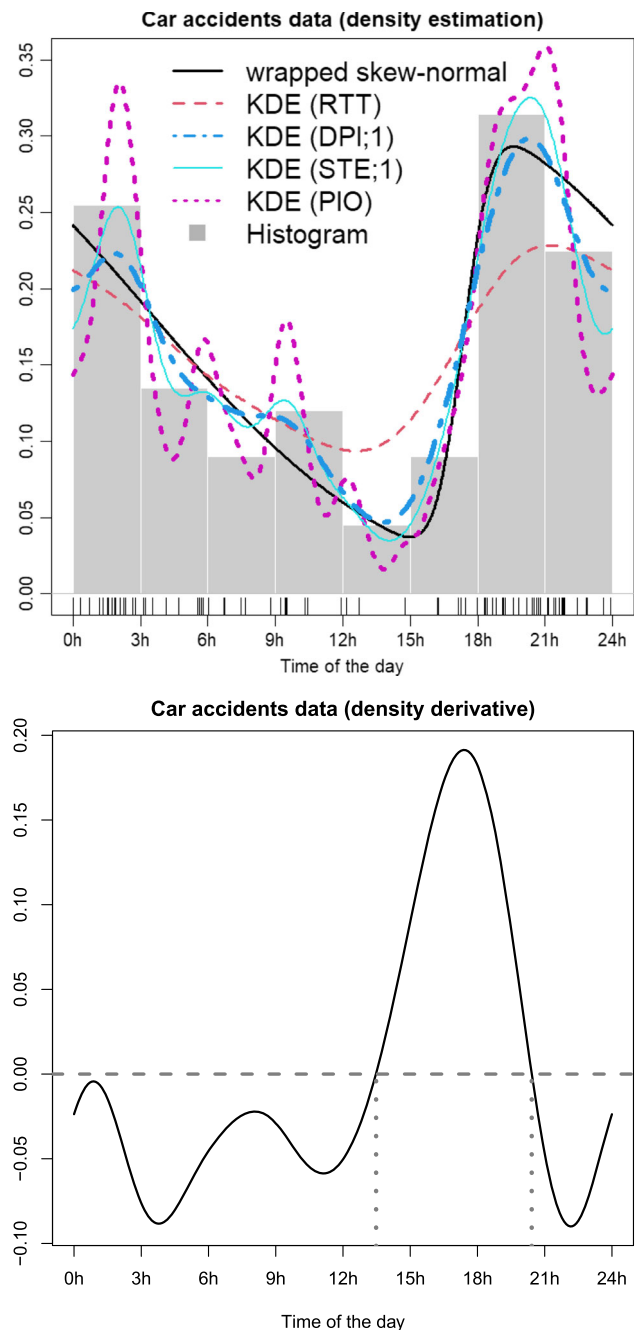
This real dataset constitutes a good example where we would recommend employing the proposed direct plug-in smoothing selector, with  $M_{\max} = 1$ , to estimate the density function as the sample size is “small” ( $n = 85$ ) and unimodality cannot be rejected for this sample. In Fig. 1 (top), we show the estimated density function employing this direct plug-in rule (DPI; 1), the proposed solve-the-equation plug-in selector (STE; 1), the rule of thumb of Taylor (2008) (RTT), and the plug-in rule of Oliveira et al. (2012) (PIO). We give the values of the smoothing parameters in Table 2. If we take the wrapped skew normal density as a reference, we can visually see that the closest kernel estimation is provided by the direct plug-in rule. This can be confirmed by computing the ISE (its expression is given in Equation (S1.1) in the Supplementary Material), replacing the “true”  $f$  with the fitted wrapped skew normal density. In Table 2, we computed the ISE obtained by different smoothing parameters proposed in the literature (see Section S1). There, we can observe that the “best” fitting (smallest ISE) is obtained with the proposed direct plug-in rule.

Finally, another application of the results derived in this paper can be found when the interest is in the density derivative estimate. As before, we will focus on the results obtained with the direct plug-in rule, with  $M_{\max} = 1$ . In that case, the smoothing parameter for the kernel estimator can be derived with Algorithm 1 (with  $r = 1$ ). The obtained estimator, with the von Mises kernel, for the car accident data is plotted in Fig. 1 (bottom). There, one can see that the derivative estimation is positive between 13:29 and 20:25 and it is negative during the remaining times of the day. Thus, if one wants to answer the question of when the peak of car accidents is pro-

**Table 2** Value of different smoothing parameters and ISE ( $\times 100$ ) computed from the car accident data, taking the fitted wrapped skew normal density as the reference density  $f$  in (S1.1)

Rule	$\nu_{\text{DPI};1}$	$\nu_{\text{STE};1}$	$\nu_{\text{RTT}}$	$\nu_{\text{RTGP}}$
Value	6.07	11.17	1.65	2.91
ISE( $\nu$ ) ( $\times 100$ )	0.26	0.51	1.13	0.49
Rule	$\nu_{\text{PIO}}$	$\nu_{\text{PIGP}}$	$\nu_{\text{PIT}}$	$\nu_{\text{LCV}}$
Value	39.29	3.01	2.50	7.81
ISE( $\nu$ ) ( $\times 100$ )	1.77	0.47	0.62	0.32

Smoothing parameters: proposed direct plug-in rule with  $M_{\max} = 1$  ( $\nu_{\text{DPI};1}$ ), solve-the-equation plug-in selector with  $M_{\max} = 1$  ( $\nu_{\text{STE};1}$ ), rule of thumb of Taylor (2008) ( $\nu_{\text{RTT}}$ ), rule of thumb of García-Portugués (2013) ( $\nu_{\text{RTGP}}$ ), plug-in rule of Oliveira et al. (2012) ( $\nu_{\text{PIO}}$ ), exact plug-in rule of García-Portugués (2013) ( $\nu_{\text{PIGP}}$ ), plug-in rule of Tenreiro (2022) ( $\nu_{\text{PIT}}$ ), and likelihood cross-validation of Hall et al. (1987) ( $\nu_{\text{LCV}}$ )



**Fig. 1** Car accident data. Top, rug plot (with tick marks) and density estimation: fitted wrapped skew normal density (thick solid line), histogram (grey rectangles), and kernel density estimator with the von Mises kernel and different smoothing parameters. Smoothing parameters: rule of thumb of Taylor (2008) (RTT, dashed line), proposed direct plug-in rule with  $M_{\max} = 1$  (DPI; 1; dot-dashed line), proposed solve-the-equation plug-in rule with  $M_{\max} = 1$  (STE; 1; thin solid line), and plug-in rule of Oliveira et al. (2012) (PIO, dotted line). Bottom (solid line): kernel density derivative estimation, with a von Mises kernel and the proposed two-stage direct plug-in smoothing selector ( $M_{\max} = 1$ ). Bottom (dotted lines): estimated location of the modal and antimodal directions

duced, we see that, according to this estimator, one can find a modal direction at 20:25. The antimodal (valley) direction is achieved at 13:29. This is also remarkable as these values are close to the nonparametric modal (20:18) and antimodal (13:35) estimators by Ameijeiras-Alonso et al. (2019).

## 9 Concluding remarks

The main contributions of this paper are the new plug-in smoothing parameters for circular kernel density (and its derivatives) estimation. In the past, some papers provided data-driven smoothing parameters for estimating the circular density function, but still, a Sheather and Jones (1991) plug-in rule was missing in the circular literature. This paper fills that gap by providing all the theoretical results needed to derive an  $l$ -stage direct and solve-the-equation plug-in smoothing selectors for a general kernel satisfying some assumptions. The needed constants to obtain the optimal smoothing parameter are given for the “most-popular” circular kernels. Besides that, following the optimality criterion of Muller (1984), this paper also discusses kernel choice. The conclusion is that the optimal kernel for circular density estimation is the wrapped Epanechnikov.

An interesting point for future research would be to analyse the asymptotic relative rate of convergence of the smoothing selectors. However, it is worth noting, at least in the linear case (see Jones et al. 1996), that achieving better asymptotics does not always translate to improved performance in real scenarios, especially with small sample sizes. In this paper, we decided to include a simulation study to compare the different proposed selectors. This study confirms that the suggested plug-in rules yield competitive smoothing parameters when compared with other proposals available in the statistical literature, particularly for small or moderate sample sizes.

Although not done in this paper, the presented ideas could be extended to the multivariate toroidal setting, using the product kernel of Di Marzio et al. (2011). In that case, the optimal smoothing parameter could be derived if the same kernel and mean resultant length are employed in all the dimensions. Thus, in principle, one could obtain an explicit expression of the optimal smoothing parameter for the kernel estimators of the density derivative and the density functionals. Then, a similar scheme to that provided in Algorithms 1 or 2 could be employed to obtain a two-stage or a solve-the-equation plug-in rule for the toroidal kernel density (and its derivatives) estimator.

## 10 Availability

The proposed  $l$ -stage direct plug-in rule and the solve-the-equation selector have been added to the R library

`NPCirc` (Oliveira et al. 2014). Given a dataset  $x$ , the two-stage plug-in rule, described in Algorithm 1 (with  $M_{\max} = 5$ ), for density estimation, can be obtained with the function `bw.AA(x, method='dpi')`. The solve-the-equation smoothing selector, described in Algorithm 2 (with  $M_{\max} = 1$ ), for density estimation is computed with `bw.AA(x)` or `bw.AA(x, method='ste')`. Thus, `bw.AA` is the circular equivalent of the function `bw.SJ` of the `stats` R package.

The kernel density derivative estimate for circular data was also added to the `NPCirc` package, inside the `kern.den.circ` function. By default it employs, as the smoothing parameter, the proposed solve-the-equation plug-in selector (with  $M_{\max} = 1$ ). A representation of the circular kernel density estimator, using the von Mises kernel and that smoothing parameter, for a sample  $x$  could be obtained as follows.

```
R> library("NPCirc")
R> fderhat <- kern.den.circ(x)
R> plot(fderhat, plot.type="line")
```

Some other extra possibilities are available by changing the arguments in the functions `bw.AA` and `kern.den.circ`. This includes, among others, the kernel  $K_\nu$  choice (argument `kernel`) or the derivative order  $r$  (argument `deriv.order`). More details are provided in Section S2 of the Supplementary Material. There, also some additional examples of how to use the `bw.AA` function are included.

## Supplementary information

This article has the following accompanying supplementary files.

Supplementary Material for the article: It contains the full results of the simulation study in Sect. 7, further comments about the available software (see Sect. 10), and the proofs of the theoretical results in Sects. 3 and 4. (.pdf file)

`bw.AA` and `kern.den.circ` R functions: The R package `NPCirc` contains, in its function `bw.AA`, the code to perform the proposed  $l$ -stage direct plug-in rule and the solve-the-equation selector. The R code for kernel density derivative estimate for circular data is provided in `kern.den.circ`. Available on the website <https://CRAN.R-project.org/package=NPCirc>. (GNU zipped tar file)

Car-crashes-data: The dataset containing the 85 car crashes that happened in El Paso County (Texas, USA) in 2018 is available on the website <https://github.com/jose-ameijeiras/Car-crashes-data>. (.csv and RData file)

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11222-024-10384-x>.

**Acknowledgements** Supported by Grant PID2020-116587GB-I00 funded by MCIN/AEI/10.13039/501100011033 and the Competitive Reference Groups 2021-2024 (ED431C 2021/24) from the Xunta de Galicia. The author is grateful to Rosa M. Crujeiras and Alberto Rodríguez-Casal for helpful suggestions and comments. The author also expresses gratitude to three anonymous reviewers for providing valuable comments that significantly contributed to the enhancement of the paper.

**Author Contributions** The author confirms sole responsibility for the entire article.

**Funding** Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

## Declarations

**Conflict of interest** The author declares that he has no conflicts of interest related to this research.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

Ameijeiras-Alonso, J., Benali, A., Crujeiras, R.M., Rodríguez-Casal, A., Pereira, J.M.: Fire seasonality identification with multimodality tests. *Ann. Appl. Stat.* **13**(4), 2120–2139 (2019). <https://doi.org/10.1214/19-AOAS1273>

Ameijeiras-Alonso, J., Crujeiras, R.M.: Flexible Circular Modeling: A Case Study of Car Accidents, *Directional Statistics for Innovative Applications*, 93–116. Springer, Singapore (2022)

Beran, R.J.: Exponential models for directional data. *Ann. Stat.* **7**(6), 1162–1178 (1979). <https://doi.org/10.1214/aos/1176344838>

Ćwik, J., Koronacki, J.: A combined adaptive-mixtures/plug-in estimator of multivariate probability densities. *Comput. Stat. Data Anal.* **26**(2), 199–218 (1997). [https://doi.org/10.1016/S0167-9473\(97\)00032-7](https://doi.org/10.1016/S0167-9473(97)00032-7)

Di Marzio, M., Panzera, A., Taylor, C.C.: Local polynomial regression for circular predictors. *Stat. Probab. Lett.* **79**(19), 2066–2075 (2009). <https://doi.org/10.1016/j.spl.2009.06.014>

Di Marzio, M., Panzera, A., Taylor, C.C.: Kernel density estimation on the torus. *J. Stat. Plan. Inference* **141**(6), 2156–2173 (2011). <https://doi.org/10.1016/j.jspi.2011.01.002>

García-Portugués, E.: Exact risk improvement of bandwidth selectors for kernel density estimation with directional data. *Electron. J. Stat.* **7**, 1655–1685 (2013). <https://doi.org/10.1214/13-ejs821>

Hall, P., Watson, G.S., Cabrera, J.: Kernel density estimation with spherical data. *Biometrika* **74**(4), 751–762 (1987). <https://doi.org/10.1093/biomet/74.4.751>

Hornik, K., Grün, B.: movMF: an R package for fitting mixtures of von Mises-Fisher distributions. *J. Stat. Softw.* **58**(10), 1–31. (2014). <https://doi.org/10.18637/jss.v058.i10>

Jones, M.C., Marron, J.S., Sheather, S.J.: A brief survey of bandwidth selection for density estimation. *J. Am. Stat. Assoc.* **91**(433), 401–407 (1996). <https://doi.org/10.1080/01621459.1996.10476701>

Klemelä, J.: Estimation of densities and derivatives of densities with directional data. *J. Multivar. Anal.* **73**(1), 18–40 (2000). <https://doi.org/10.1006/jmva.1999.1861>

Ley, C., Verdebout, T.: (eds.) *Applied Directional Statistics*. Chapman & Hall/CRC Interdisciplinary Statistics Series. CRC Press, Boca Raton (2018)

Mardia, K.V., Jupp, P.E.: *Directional Statistics*. Wiley Series in Probability and Statistics. Chichester: Wiley (1999)

Muller, H.G.: Smooth optimum kernel estimators of densities, regression curves and modes. *Ann. Stat.* **12**(2), 766–774 (1984). <https://doi.org/10.1214/aos/1176346523>

Oliveira, M., Crujeiras, R.M., Rodríguez-Casal, A.: A plug-in rule for bandwidth selection in circular density estimation. *Comput. Stat. Data Anal.* **56**(12), 3898–3908 (2012). <https://doi.org/10.1016/j.csda.2012.05.021>

Oliveira, M., Crujeiras, R.M., Rodríguez-Casal, A.: NPCirc: an R package for nonparametric circular methods. *J. Stat. Softw.* **61**(9), 1–26. <https://doi.org/10.18637/jss.v061.i09>

Pewsey, A.: The wrapped skew-normal distribution on the circle. *Commun. Stat. Theory Methods* **29**(11), 2459–2472 (2000). <https://doi.org/10.1080/03610920008832616>

R Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2022)

SenGupta, A., Arnold, B.C.: (eds.) *Directional Statistics for Innovative Applications*. Forum for Interdisciplinary Mathematics. Springer, Singapore (2022)

Sheather, S.J., Jones, M.C.: A reliable data-based bandwidth selection method for kernel density estimation. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **53**(3): 683–690. <https://doi.org/10.1111/j.2517-6161.1991.tb01857.x> (1991)

Taylor, C.C.: Automatic bandwidth selection for circular density estimation. *Comput. Stat. Data Anal.* **52**(7), 3493–3500 (2008). <https://doi.org/10.1016/j.csda.2007.11.003>

Tenreiro, C.: Kernel density estimation for circular data: a fourier series-based plug-in approach for bandwidth selection. *J. Nonparametr. Stat.* **34**(2), 377–406 (2022). <https://doi.org/10.1080/10485252.2022.2057974>

Tsuruta, Y., Sagae, M.: Asymptotic property of wrapped Cauchy kernel density estimation on the circle. *Bull. Inform. Cybernet.* **49**, 1–10 (2017). <https://doi.org/10.5109/2232318>

Wand, M.P., Jones, M.C.: *Kernel Smoothing*. Chapman and Hall, Great Britain (1995)

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.