Pré-Publicações do Departamento de Matemática
Universidade de Coimbra
Preprint Number 23–10

A PARZEN–ROSENBLATT TYPE DENSITY ESTIMATOR FOR CIRCULAR DATA: EXACT AND ASYMPTOTIC OPTIMAL BANDWIDTHS

CARLOS TENREIRO

Abstract: For the Parzen–Rosenblatt type density estimator for circular data we prove the existence of a minimiser, $h_{\mathrm{MISE}}(f;K,n)$, of its exact mean integrated squared error. Under mild conditions we show that it is asymptotically equivalent to the bandwidth $h_{\mathrm{AMISE}}(f;K,n)$ that minimises the leading terms of the mean integrated squared error expansion, and we obtain the order of convergence of the relative error $h_{\mathrm{AMISE}}(f;K,n)/h_{\mathrm{MISE}}(f;K,n)-1$. Some small- and moderate-sample-size comparisons between the two bandwidths are also presented when the underlying density is a mixture of von Mises densities.

Keywords: Parzen–Rosenblatt type density estimator; circular data; exact and asymptotic optimal bandwidths.
AMS Subject Classification (2020): 62H11, 62G07.

1. Introduction

Given an independent and identically distributed sample of angles $X_1,\ldots,X_n \in [0,2\pi[$ from some absolutely continuous circular random variable X with unknown probability density function f, we are interested in this paper in the Parzen–Rosenblatt type estimator of f (hereafter, PR-type estimator, in short; Rosenblatt, 1956, Parzen, 1962) defined, for $\theta \in [0,2\pi[$, by

$$\hat{f}_n(\theta;h) = \frac{d_h(K)}{n}\sum_{i=1}^{n} K_h(\theta - X_i), \tag{1}$$

where $h = h_n > 0$ is a sequence of strictly positive real numbers converging to zero as n tends to infinity, K_h is a real-valued periodic function on $\mathbb{R}$, with period 2π, such that $K_h(\theta) = K(\theta/h)/h$, for $\theta \in [-\pi,\pi[$, with K a kernel on $\mathbb{R}$, that is, an integrable real-valued function on $\mathbb{R}$ with $\int_{\mathbb{R}} K(u)du > 0$ (not necessarily a probability density function), and $d_h(K)$ is a normalizing constant depending on the kernel K and the bandwidth h which is chosen so that $\hat{f}_n(\cdot;h)$ integrates to unity. Despite being structurally closed to the PR-estimator for linear data, the periodicity imposed to K_h makes estimator

Received April 19, 2023.

$\hat{f}_n(\cdot; h)$ well adapted to deal with circular data by automatically correcting the potential boundary problems that may occur at the extreme points of the distribution support when the standard PR-estimator is used to estimate f. Although the rationale behind its definition can be found in Silverman (1986, pp. 29–32), to the best of our knowledge estimator (1) was for the first time explicitly proposed and studied in Tenreiro (2022). Note that the PR-type estimator (1) is closely related with the standard kernel density estimator for circular data defined, for $\theta \in [0, 2\pi[$, by

$$\tilde{f}_n(\theta; g) = \frac{c_g(L)}{n} \sum_{i=1}^{n} L\left(\frac{1 - \cos(\theta - X_i)}{g^2}\right), \tag{2}$$

where $L : [0, +\infty[\to \mathbb{R}$ is such that $L(t) = K(\sqrt{t})$, $g = h/\sqrt{2}$ and $c_g(L)$, depending on the kernel L and the bandwidth g, is chosen so that $\tilde{f}_n(\cdot; g)$ integrates to unity. Kernel estimators of this form for estimating densities of q-dimensional unit spheres, for $q \geq 1$, were initially studied in Beran (1979), Hall et al. (1987), Bai et al. (1988) and Klemela (2000), the last work being restricted to $q \geq 2$, and more recently by García-Portugués et al. (2013) and García-Portugués (2013). If we take for K the Gaussian kernel $K(u) = e^{-u^2}$, L is the so-called von Mises kernel $L(t) = e^{-t}$ and (2) is the density estimator considered in Taylor (2008) and Oliveira et al. (2012). In this case the estimator is a combination of circular normal or von Mises densities with mean directions X_i and concentration parameters equal to g^{-2} as it takes the form

$$\tilde{f}_n(\theta; g) = \frac{1}{n} \sum_{i=1}^{n} f_{\mathrm{vM}}(\theta; X_i, g^{-2}),$$

where

$$f_{\mathrm{vM}}(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\big(\kappa \cos(\theta - \mu)\big), \tag{3}$$

is the von Mises density with mean direction $\mu \in [0, 2\pi[$ and concentration parameter $\kappa \geq 0$, and $I_r(\nu)$ is, for $\nu \geq 0$ and $r \geq 0$, the modified Bessel function of order r defined by

$$I_r(\nu) = \frac{1}{2\pi} \int_0^{2\pi} \cos(r\theta) \exp(\nu \cos\theta) d\theta.$$

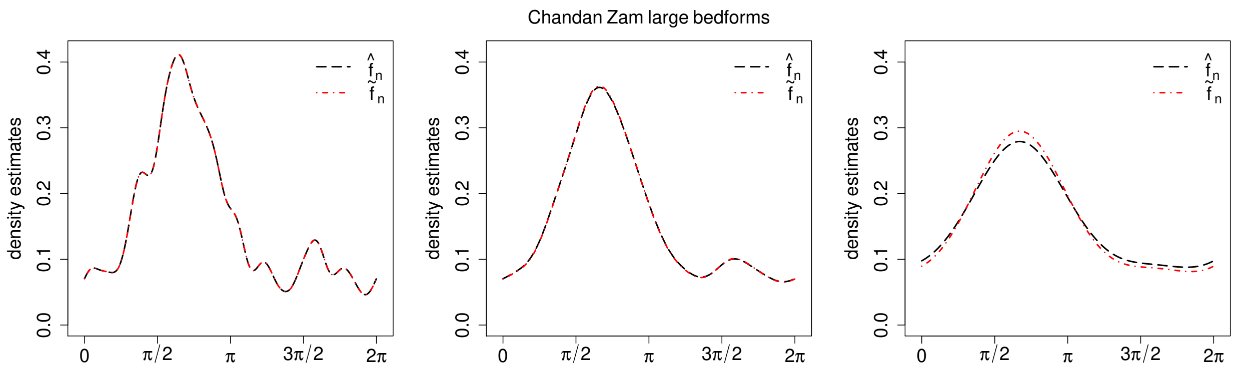

FIGURE 1. *Estimates for the cross-bed density by using estimators $\hat{f}_n(\cdot;h)$ and $\tilde{f}_n(\cdot;g)$ where $K(u) = e^{-u^2}$, $L(t) = e^{-t}$, $h = \sqrt{2}g$ and $g = 0.175$ (left), $g = 0.35$ (centre) and $g = 0.7$ (right).*

The close connection between estimators (1) and (2) is illustrated in Figure 1 where estimates for the cross-bed density based on 104 cross-bed measurements from the Himalayan molasse in Pakistan (see Fisher, 1993, Measurements of Chaudan Zam large bedforms, pp. 250–251) are produced for three different bandwidths.

Unlike the selection of the kernel, the choice of the bandwidth is crucial to the performance of each one of the previous estimators (see Tenreiro, 2022, pp. 387–388). As for any density estimator, the closeness of the PR-type estimator (1) to its target density f can be measured through the mean integrated squared error (MISE) which is defined by

$$\mathrm{MISE}(f;h,n) := \mathrm{E}\big(\mathrm{ISE}(f;h,n)\big) = \mathrm{E}\int_0^{2\pi} \big\{\hat{f}_n(\theta;h) - f(\theta)\big\}^2 d\theta. \tag{4}$$

According to this error criterion a natural choice for h is the minimiser of the real-valued function $h \mapsto \mathrm{MISE}(f;h,n)$, defined for $h \in]0,+\infty[$, whenever such a minimiser exists. The existence of such a minimiser, we denote henceforth by $h_{\mathrm{MISE}}(f;K,n)$, is the first question we address in this paper. Unlike the case of linear data considered in Chacón et al. (2007), who proved that under general conditions on the kernel K, there exists an exact optimal bandwidth for all square integrable density f and all sample sizes, we will see that the same does not occur in the context of circular data. Nevertheless, an exact optimal bandwidth $h_{\mathrm{MISE}}(f;K,n)$ always exists for large enough sample sizes, and we provide a sufficient condition on K and f for the existence of $h_{\mathrm{MISE}}(f;K,n)$ for all sample sizes. To the best of our knowledge similar results are not available in the literature for the standard kernel estimator for circular data (2).

Under some regularity conditions on f and moment conditions on K, it has been proved in Tenreiro (2022, Theorem 3.4) that the MISE (4) admits the asymptotic expansion

$$\mathrm{MISE}(f;h,n) = (nh)^{-1}\boldsymbol{d}_1(K) + \frac{h^4}{4}\boldsymbol{d}_2(K)\int_0^{2\pi} f''(\theta)^2 d\theta + o\big((nh)^{-1} + h^4\big), \quad (5)$$

whenever the bandwidth satisfies the classical conditions $h \to 0$ and $nh \to +\infty$, as $n \to +\infty$, where

$$\boldsymbol{d}_1(K) = \int_{\mathbb{R}} K(u)^2 du \left(\int_{\mathbb{R}} K(u)du\right)^{-2} \quad (6)$$

and

$$\boldsymbol{d}_2(K) = \left(\int_{\mathbb{R}} u^2 K(u)du\right)^2 \left(\int_{\mathbb{R}} K(u)du\right)^{-2}. \quad (7)$$

Then, we deduce that the asymptotic optimal bandwidth for estimator (1), that is, the bandwidth that minimises the leading terms of (5), called asymptotic MISE,

$$\mathrm{AMISE}(f;h,n) = (nh)^{-1}\boldsymbol{d}_1(K) + \frac{h^4}{4}\boldsymbol{d}_2(K)\int_0^{2\pi} f''(\theta)^2 d\theta, \quad (8)$$

is given by

$$h_{\mathrm{AMISE}}(f;K,n) = \boldsymbol{d}(K)\left(\int_0^{2\pi} f''(\theta)^2 d\theta\right)^{-1/5} n^{-1/5}, \quad (9)$$

where

$$\boldsymbol{d}(K) = \left(\int_{\mathbb{R}} K(u)^2 du\right)^{1/5} \left(\int_{\mathbb{R}} u^2 K(u)du\right)^{-2/5},$$

whenever f is not the circular uniform distribution. The fact that $h_{\mathrm{AMISE}}(f; K,n)$ minimises the leading terms of (5) does not enable us to conclude by itself that the two bandwidths $h_{\mathrm{AMISE}}(f;K,n)$ and $h_{\mathrm{MISE}}(f;K,n)$ are asymptotically equivalent. This is the second question we address in this paper. Under some additional assumptions on f and K we establish such an asymptotic equivalence and we obtain the order of convergence of the relative error $h_{\mathrm{AMISE}}(f;K,n)/h_{\mathrm{MISE}}(f;K,n) - 1$, which agrees with the corresponding order of convergence for the case of linear data. As a by-product of this result we conclude that any plug-in bandwidth selector, such as the Fourier series-based plug-in bandwidth selector introduced in Tenreiro (2022), which is a consistent estimator of $h_{\mathrm{AMISE}}(f;K,n)$, is also a consistent estimator of the exact optimal bandwidth $h_{\mathrm{MISE}}(f;K,n)$. As the previous results are asymptotic in nature, it is relevant to assess the quality of h_{AMISE} as a surrogate for h_{MISE} for small- and moderate-sample sizes. This is the third question we address in this paper. For that we consider some of the circular density models introduced in Oliveira et al. (2012). When the underlying density presents complex distributional characteristics, we conclude that only for large sample sizes the asymptotic optimal bandwidth h_{AMISE} is a suitable surrogate for h_{MISE}. This is quite a striking conclusion since h_{AMISE} is the usual target bandwidth for plug-in bandwidth selection methods for the estimation of f.

The rest of the paper is as follows. In Section 2 we derive an expansion for the exact MISE of the estimator (1) which enables to prove that the asymptotic expansion (5) is valid for a broader class of kernels than that considered in Tenreiro (2022, Theorem 3.4). Such an exact MISE expansion is used in Section 3 to discuss the existence of an optimal bandwidth for a fixed sample size n. The limit behaviour of the optimal bandwidth as n tends to infinity is studied in Section 4, where we also establish the order of convergence of $h_{\mathrm{AMISE}}/h_{\mathrm{MISE}}-1$. To assess the quality of h_{AMISE} as a surrogate for h_{MISE}, we present in Section 5 some small- and moderate-sample-size comparisons between the two bandwidths when the underlying density function is a mixture of von Mises densities. Finally, in Section 6 we draw some overall conclusions. For convenience of exposition all the proofs are deferred to Section 7. The plots shown in this paper were carried out using the R software (R Development Core Team, 2019).

2. Exact and asymptotic expansions for the MISE

In this section we derive one exact and two asymptotic expansions for the MISE of the PR-type estimator (1) that generalize expansion (5) obtained in Tenreiro (2022, Theorem 3.4). Here and henceforth in the paper, a kernel on $\mathbb{R}$ is an integrable real-valued function on $\mathbb{R}$ with $\int_{\mathbb{R}} K(u)du > 0$. First recall that the MISE (4) can be written as

$$\mathrm{MISE}(f;h,n) = \mathrm{IV}(f;h,n) + \mathrm{ISB}(f;h),$$

where

$$\mathrm{IV}(f;h,n) := \int_0^{2\pi} \mathrm{Var}\hat{f}_n(\theta;h)\,d\theta$$

and

$$\mathrm{ISB}(f;h) := \int_0^{2\pi} \big\{\mathrm{E}\hat{f}_n(\theta;h) - f(\theta)\big\}^2 d\theta,$$

are, respectively, the integrated variance and integrated squared bias of $\hat{f}_n$. Similarly to the PR-estimator for linear data, if K is square integrable on $\mathbb{R}$ ($K \in L_2(\mathbb{R})$) and f is square integrable on $[0,2\pi[$ ($f \in L_2([0,2\pi[)$), the bias and variance of $\hat{f}_n$ at a point θ can be expressed in terms of the convolution between K_h and the circular density f. Recall that if α and β are real-valued functions with period 2π defined on $\mathbb{R}$, the convolution of α and β is defined, for $x \in \mathbb{R}$, by

$$(\alpha * \beta)(x) = \int_0^{2\pi} \alpha(x-y)\beta(y)dy,$$

whenever this integral exists. As the integrand is periodic with period 2π, the previous definition does not depend on the considered interval of integration with length 2π. The convolution $(\alpha * \beta)(x)$ exists for almost every $x \in \mathbb{R}$ whenever α and β are integrable on $[0,2\pi[$, and it exists and is continuous for every $x \in \mathbb{R}$, whenever α and β are square integrable on $[0,2\pi[$. Obviously, the convolution is a periodic function if it exists (see Butzer and Nessel, 1971, §0.4). Therefore, for all $\theta \in [0,2\pi[$, the mean of $\hat{f}_n$ is given by

$$\mathrm{E}\hat{f}_n(\theta;h) = d_h(K)\int_0^{2\pi} K_h(\theta-u)f(u)du = d_h(K)(K_h * f)(\theta), \tag{10}$$

and, for almost all $\theta \in [0,2\pi[$, the variance of $\hat{f}_n$ is given by

$$\mathrm{Var}\hat{f}_n(\theta;h) = n^{-1}d_h(K)^2\left(\int_0^{2\pi} K_h(\theta-u)^2 f(u)du - (K_h * f)(\theta)^2\right), \tag{11}$$

where for the sake of simplicity we also denote by f the periodic extension of f to $\mathbb{R}$ given by $f(\theta) = f(\theta - 2k\pi)$, whenever $\theta \in [2k\pi, 2(k+1)\pi[$, for some $k \in \mathbb{Z}$. From these equations, it turns out that the mean and variance of $\hat{f}_n$ has the same form as the mean and variance of the ordinary kernel estimator for linear data (see Wand and Jones, 1995, formulas (2.4) and (2.5)). Hence, we can easily deduce exact expansions for the integrated variance (IV) and integrated squared bias (ISB) of the estimator (1) which are at the basis of the main results presented in this paper. If g is a real-valued function on $\mathbb{R}$, we denote by $\bar{g}$ the function defined by $\bar{g}(u) = g(-u)$, for $u \in \mathbb{R}$.

Theorem 1. *If $f \in L_2([0, 2\pi[)$ and $K \in L_2(\mathbb{R})$ is a kernel, then for all $n \in \mathbb{N}$ and $h > 0$ with $\int_{-\pi/h}^{\pi/h} K(u)du \neq 0$, we have*

$$\mathrm{IV}(f; h, n) = (nh)^{-1} d_h(K)^2 d_h(K^2)^{-1} - n^{-1} d_h(K)^2 C(f; h), \tag{12}$$

and

$$\mathrm{ISB}(f; h) = d_h(K)^2 C(f; h) - 2d_h(K)D(f; h) + \int_0^{2\pi} f(\theta)^2 d\theta, \tag{13}$$

where

$$d_h(K)^{-1} := \int_{-\pi}^{\pi} K_h(\theta)d\theta = \int_{-\pi/h}^{\pi/h} K(u)du,$$

$$C(f; h) := \int_0^{2\pi} (K_h * f)(\theta)^2 d\theta = \int_{-\pi/h}^{\pi/h}\int_{-\pi/h}^{\pi/h} K(u)K(v)(\bar{f} * f)(h(u - v))dudv$$

and

$$D(f; h) := \int_0^{2\pi} (K_h * f)(\theta)f(\theta)d\theta = \int_{-\pi/h}^{\pi/h} K(u)(\bar{f} * f)(hu)du.$$

Combining equations (12) and (13) we obtain an exact formula for the MISE of the estimator (1) given by

$$\begin{aligned} \mathrm{MISE}(f; h, n) = (nh)^{-1} d_h(K)^2 d_h(K^2)^{-1} + \big(1 - n^{-1}\big) d_h(K)^2 C(f; h) \\ - 2d_h(K)D(f; h) + \int_0^{2\pi} f(\theta)^2 d\theta, \end{aligned} \tag{14}$$

which is the analogue of formula (2.2) in Marron and Wand (1992) (see also Wand and Jones, 1995, formula (2.8)) for the kernel density estimator of linear data. This exact formula will be useful to explore the existence and limit behaviour of the optimal bandwidth we discuss in Sections 3 and 4. Before that, we show that the asymptotic expansion (5) can be established under slightly weaker conditions than those considered in Theorem 3.4 of Tenreiro (2022).

Theorem 2. *Let us assume that $f \in L_2([0, 2\pi[)$ and $K \in L_2(\mathbb{R})$ is a kernel. If $h \to 0$ and $nh \to \infty$, as $n \to \infty$, then $\mathrm{MISE}(f; h, n) \to 0$. Moreover, (5) is valid whenever K is symmetric with $\int_{\mathbb{R}} u^2 |K(u)|du < \infty$, and f is such that f' is absolutely continuous on $[0, 2\pi[$ with $f'' \in L_2([0, 2\pi[)$.*

Similarly to the case of linear data, the order of convergence to zero of the integrated squared bias of the PR-type estimator (1) can be improved if a higher-order kernel is taken for K. Next we assume that K is a kernel of order k with bounded support, for some integer $k \geq 2$, that is, $K : \mathbb{R} \to \mathbb{R}$ is an integrable function with bounded support such that $\int_{\mathbb{R}} |u|^k |K(u)| du < \infty$, with $\int_{\mathbb{R}} K(u)du > 0$, $\int_{\mathbb{R}} u^j K(u)du = 0$, for $j = 1, \ldots, k-1$, and $\int_{\mathbb{R}} u^k K(u)du \neq 0$.

Theorem 3. *For some integer $k \geq 2$ let us assume that $f^{(k-1)}$ is absolutely continuous on $[0, 2\pi[$ with $f^{(k)} \in L_2([0, 2\pi[)$ and $K \in L_2(\mathbb{R})$ is a kernel of order k with bounded support. If $h \to 0$ and $nh \to \infty$, as $n \to \infty$, then*

$$\mathrm{MISE}(f; h, n) = (nh)^{-1} \boldsymbol{d}_1(K) + \frac{h^{2k}}{(k!)^2} \boldsymbol{d}_k(K) \int_0^{2\pi} f^{(k)}(\theta)^2 d\theta + o\big((nh)^{-1} + h^{2k}\big),$$

where $\boldsymbol{d}_1(K)$ is given by (6) and $\boldsymbol{d}_k(K)$ is defined by (7) with $u^k K(u)$ instead of $u^2 K(u)$.

The previous formulas for the MISE of the PR-type estimator for circular data are the analogue of the well known formulas for the MISE of the PR-estimator for linear data (see Wand and Jones, 1995, pp. 21, 33) showing the close relationship between these two kernel density estimators.

3. Existence of an exact optimal bandwidth

In the context of linear data, Chacón et al. (2007, Theorem 1) proved that under mild conditions on the kernel K there exists an optimal bandwidth minimising the MISE for all square integrable density f and all sample sizes. As suggested by the graphics displayed in Figure 2 about the MISE of the PR-type estimator with $K(u) = e^{-u^2}$ for the circular density model M14 considered in Oliveira et al. (2012), which is a mixture of four von Mises densities, such an exact optimal bandwidth may not exist in the context of circular data when the sample size is small (the same behaviour was observed for the parabolic kernel $K(u) = (1 - u^2)I(|u| \leq 1)$). Nevertheless, as established in the following result, an exact optimal bandwidth always exists whenever the sample size is large enough and K satisfies the following assumptions:

(K.1) K is such that $\int_{-\lambda}^{\lambda} K(u)du \neq 0$, for all $\lambda > 0$;

(K.2) K is continuous at zero with $K(0) > 0$.

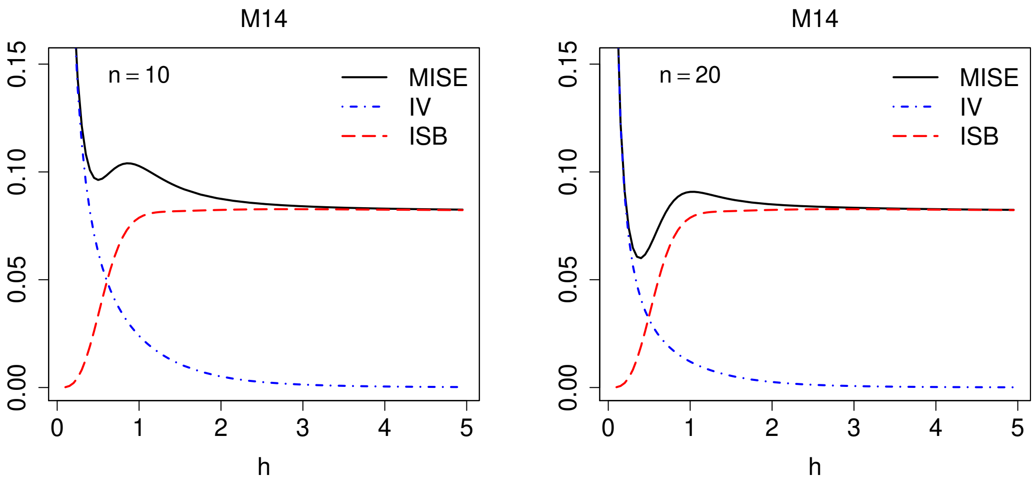

FIGURE 2. MISE *behaviour as a function of* h *for the circular density model M14 (von Mises mixture with four components) with* $n = 10$ *(left) and* $n = 20$ *(right) when* $K(u) = e^{-u^2}$.

Taking into account that $\lambda \mapsto \int_{-\lambda}^{\lambda} K(u)du$ is continuous on $]0, +\infty[$ with $\lim_{\lambda\to\infty} \int_{-\lambda}^{\lambda} K(u)du = \int_{\mathbb{R}} K(u)du > 0$, from (K.1) we have $\int_{-\lambda}^{\lambda} K(u)du > 0$, for all $\lambda > 0$. Moreover, if K is continuous at zero, we necessarily have $K(0) \geq 0$.

Theorem 4. *Let us assume that* $f \in L_2([0, 2\pi[)$ *is not the uniform density and* $K \in L_2(\mathbb{R})$ *is a kernel satisfying assumptions (K.1) and (K.2). Then there exists* $m \in \mathbb{N}$ *such that, for all* $n \geq m$ *there exists* $h_{\text{MISE}} = h_{\text{MISE}}(f; n) > 0$ *such that*

$$\text{MISE}(f; h_{\text{MISE}}, n) \leq \text{MISE}(f; h, n), \text{ for all } h > 0.$$

Notice that the previous result says nothing about the uniqueness of the optimal bandwidth. Likely, as in Marron and Wand (1992), it could be possible to find an example where the optimal bandwidth is not unique. However we do not pursue this further in this paper.

The case of the circular uniform distribution is a special one. For this distribution the integrated squared bias is equal to zero and the exact MISE is simply given by $\text{MISE}(f; h, n) = n^{-1}\big(d_h(K)^2 d_h(K^2)^{-1} h^{-1} - 1/(2\pi)\big)$. Therefore, we have $\text{MISE}(f; h, n) = o(1)$ even when the smoothing parameter does not converge to zero as n tends to infinity. More precisely, if $h \to \lambda \in [0, +\infty]$, the fastest rate of convergence is obtained when $\lambda = +\infty$, in which case we get $\text{MISE}(f; h, n) = o(n^{-1})$. From the previous expression for the MISE we also conclude that an exact optimal bandwidth does not exist for the circular uniform distribution.

Below we present a sufficient condition on K and f for the existence of an exact optimal bandwidth for all $n \in \mathbb{N}$. For that we need the following additional assumption on K:

(K.3) For some $p > 0$, the limit

$$\lim_{h\to\infty} h^p\big(K(u/h) - K(0)\big) =: \ell_K(u),$$

exists and is finite for all $u \in [-\pi, \pi[$, where ℓ_K is not zero almost everywhere (with respect to the Lebesgue measure), and for some $H \geq 0$ we have

$$\sup_{h>H} \sup_{u\in[-\pi,\pi[} h^p\big|K(u/h) - K(0)\big| < \infty.$$

Note that if K satisfies assumption $(K.3)$ then K is necessarily continuous at zero. However, no differentiability at zero is imposed on K. For example, the triangular kernel $K(u) = (1-|u|)I(|u| \leq 1)$ fulfills assumption $(K.3)$ with $\ell_K(u) = -|u|$. Nevertheless, if K is continuously differentiable up to order $p \in \mathbb{N}$ in a neighbourhood of the origin with $K^{(j)}(0) = 0$, for $j = 1, \ldots, p-1$, and $K^{(p)}(0) \neq 0$, then K satisfies assumption (K.3) with $\ell_K(u) = \frac{1}{p!}K^{(p)}(0)u^p$. Therefore, for the Gaussian kernel $K(u) = e^{-u^2}$, as well as for the parabolic kernel $K(u) = (1 - u^2)I(|u| \leq 1)$, we have $\ell_K(u) = -u^2$.

Theorem 5. *Let us assume that $f \in L_2([0, 2\pi[)$ and that $K \in L_2(\mathbb{R})$ satisfies assumptions (K.1), (K.2) and (K.3). If*

$$\int_{-\pi}^{\pi} \ell_K(x)\Big(\bar{f} * f(x) - \frac{1}{2\pi}\Big)dx > 0, \tag{15}$$

then for all $n \in \mathbb{N}$ there exists $h_{\mathrm{MISE}} = h_{\mathrm{MISE}}(f; K, n) > 0$ such that

$$\mathrm{MISE}(f; h_{\mathrm{MISE}}, n) \leq \mathrm{MISE}(f; h, n), \text{ for all } h > 0.$$

If the kernel K is such that $\ell_K(u) = -u^2$, the condition (15) is satisfied by all the models considered in Oliveira et al. (2012) with the exception of models M1 (circular uniform), M7, M14 and M20. As it follows from Figure 2, an exact optimal bandwidth does not necessarily exist for model M14 when the sample size is small. As illustrated in Figure 3, the same is true for models M7 and M20.

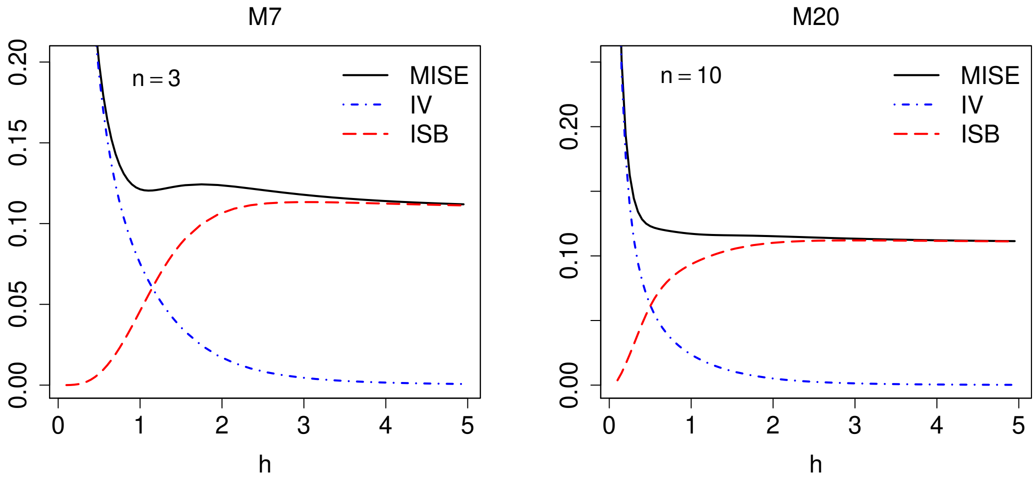

FIGURE 3. MISE *behaviour as a function of* h *for the circular density models M7 (von Mises mixture with two components) with* $n = 3$ *(left) and M20 (mixture of two wrapped skew-Normal and two wrapped Cauchy) with* $n = 10$ *(right), when* $K(u) = e^{-u^2}$.

4. Asymptotic behaviour of the exact optimal bandwidth

In this section we study the limit behaviour of the exact optimal bandwidth $h_{\mathrm{MISE}}(f; K, n)$, which existence has been established in Theorem 4. We first prove that it satisfies the classical conditions $h_{\mathrm{MISE}}(f; K, n) \to 0$ and $nh_{\mathrm{MISE}}(f; K, n) \to +\infty$, as $n \to +\infty$. For that we follow the approach of Chacón et al. (2007).

Theorem 6. *If* $f \in L_2([0, 2\pi[)$ *is not the uniform density and* $K \in L_2(\mathbb{R})$ *satisfies assumptions (K.1) and (K.2), then*

$$\lim_{n \to +\infty} nh_{\mathrm{MISE}}(f; K, n) = +\infty.$$

For the study of the limit behaviour of $h_{\mathrm{MISE}}(f; K, n)$, as $n \to +\infty$, we will restrict our attention to the important case where K is a symmetric and nonnegative kernel. Thus, from now on the estimator (1) is a circular probability density function for each sample $X_1, \ldots, X_n$, that is, $\hat{f}_n$ is a proper circular density estimator.

Theorem 7. *If* $f \in L_2([0, 2\pi[)$ *is not the uniform density and* $K \in L_2(\mathbb{R})$ *is a nonnegative and symmetric kernel, satisfying assumption (K.2), then*

$$\lim_{n \to +\infty} h_{\mathrm{MISE}}(f; K, n) = 0.$$

In the following result we prove that $h_{\text{MISE}}(f;K,n)$ is asymptotically equivalent to the asymptotic optimal bandwidth $h_{\text{AMISE}}(f;K,n)$ given by (9), and we obtain the order of convergence of the relative error $h_{\text{AMISE}}(f;K,n)$ $/h_{\text{MISE}}(f;K,n)-1$. This order of convergence is equal to the corresponding order of convergence for the standard PR-estimator (see Hall and Marron, 1991, p. 160).

Theorem 8. *Under the conditions of Theorems 2 and 7, we have*

$$0<\liminf_{n\to+\infty} n^{1/5}h_{\text{MISE}}(f;K,n)\le \limsup_{n\to+\infty} n^{1/5}h_{\text{MISE}}(f;K,n)<\infty$$

and

$$\lim_{n\to\infty}\frac{h_{\text{AMISE}}(f;K,n)}{h_{\text{MISE}}(f;K,n)}=1.$$

Moreover, if f is three-times continuously differentiable on $[0,2\pi[$ and K has bounded support, then

$$\frac{h_{\text{AMISE}}(f;K,n)}{h_{\text{MISE}}(f;K,n)}-1=O\big(n^{-2/5}\big).$$

Note that although the kernel $K(u)=e^{-u^2}$ has an unbounded support, it can be proved that the second part of Theorem 8 also applies to this kernel. The derivation of such result is considerably more complicated than when the kernel has bounded support, but its proof follows closely that of the bounded support case.

5. Some finite sample comparisons

Our goal in this section is to compare the finite sample performance of the two bandwidths, h_{MISE} and h_{AMISE}. To this end we work with the exact MISE formula (14) within the class of von Mises mixture densities, that is, the class of densities f that can be written as $f(\theta)=\sum_{\ell=1}^{k} w_\ell f_{\text{vM}}(\theta;\mu_\ell,\kappa_\ell)$, where $\sum_{\ell=1}^{k} w_\ell=1$ with $w_\ell>0$, and $f_{\text{vM}}(\cdot;\mu,\kappa)$ denotes the von Mises density with mean direction $\mu\in[0,2\pi[$ and concentration parameter $\kappa\ge 0$ given by (3). This set of densities is very rich, containing densities with a wide variety of distribution features such as multimodality, skewness and/or peakedness. The formula (14), where we always take for K the Gaussian kernel $K(u)=e^{-u^2}$, enables us to compare the exact MISE and its minimiser h_{MISE} with AMISE and h_{AMISE}, given by (8) and (9), respectively. To analyse their finite sample performance, we use the circular densities models M2, M8, M13

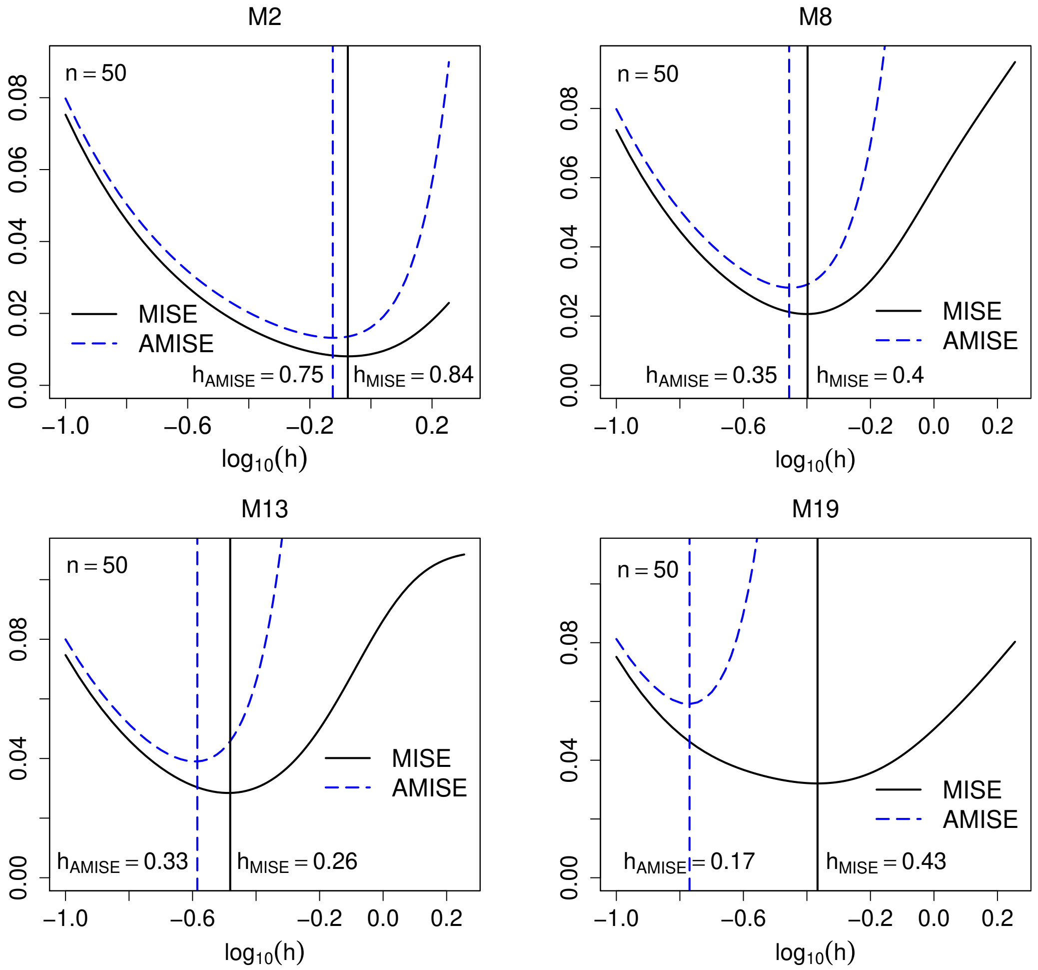

FIGURE 4. *Comparisons of* MISE(h) *(solid curve) and* AMISE(h) *(dashed curve), plotted on the* $\log_{10}(h)$ *scale with kernel* $K(u) = e^{-u^2}$*, for the circular density models M2, M8, M13 and M19 (von Mises mixtures with 1, 2, 3 and 5 components, respectively).*

and M19 introduced in Oliveira et al. (2012), which are von Mises mixture densities with 1, 2, 3 and 5 components, respectively, therefore presenting an increasing distributional complexity from the simplest M2 model to the more complex M19 model. For a careful description of these models and the plots of the corresponding circular densities, see Oliveira et al. (2012, pp. 3901, 3902, 3907).

Figure 4 shows the graphs of MISE(h) and AMISE(h) plotted on the $\log_{10}(h)$ scale, for a sample size $n = 50$. Note that the approximation of these

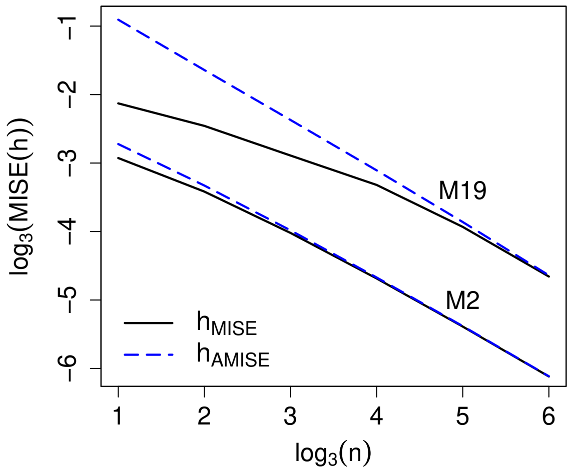

FIGURE 5. $\log_3(\mathrm{MISE}(h_{\mathrm{MISE}}))$ *(solid curve) and* $\log_3(\mathrm{MISE}(h_{\mathrm{AMISE}}))$ *(dashed curve) as a function of* $\log_3(n)$ *for the circular density models M2 and M19 (von Mises mixtures with 1 and 5 components, respectively) with kernel* $K(u) = e^{-u^2}$.

curves is good for small h, but poor for large h. For the simpler circular models, h_{AMISE} provides a quite reliable approximation to h_{MISE}. However, the quality of h_{AMISE} as surrogate for h_{MISE} deteriorates as the complexity of the underlying model increases. In the case of model M19, which is a von Mises mixture density with 5 components, h_{AMISE} is a very poor approximation to h_{MISE}. These findings are stressed in Figure 5 where we plot the values $\log_3(\mathrm{MISE}(h_{\mathrm{MISE}}))$ and $\log_3(\mathrm{MISE}(h_{\mathrm{AMISE}}))$ versus $\log_3(n)$ for the circular models M2 and M19. The $\mathrm{MISE}(h_{\mathrm{AMISE}})$ curves are almost linear for moderate and large sample sizes both having slope $-4/5$, but with much different intercepts because the quadratic functional $\int_0^{2\pi} f''(\theta)^2 d\theta$, which magnitude can be taken as a measure of how difficult a circular density is to estimate (Wand and Jones, 1995, pp. 36–39), is much bigger for density M19 than for density M2. For the circular model M19, which density presents complex distributional characteristics, only for large sample sizes the asymptotic optimal bandwidth h_{AMISE} is a good surrogate for h_{MISE}. Similarly to the case of linear data (see Marron and Wand, 1992, pp. 719–725), this is quite a striking conclusion since h_{AMISE} is the usual target bandwidth for plug-in bandwidth selection methods for the estimation of f (see Tenreiro, 2022).

6. Conclusions

For the PR-type density estimator for circular data we establish in this paper the existence of a minimiser of its exact MISE, called exact optimal bandwidth. Under mild conditions we show that the exact and the asymptotic optimal bandwidths are asymptotically equivalent, and we obtain the order of convergence of the corresponding relative error. As a by-product of these results we deduce that any plug-in bandwidth selector based on the asymptotic optimal bandwidth expression, which is a consistent estimator of this target bandwidth, is also a consistent estimator of the exact optimal bandwidth. Some small- and moderate-sample-size comparisons between the two bandwidths are also presented in this paper when the underlying density is a mixture of von Mises densities. They enable us to conclude that the asymptotic optimal bandwidth might not be a suitable surrogate for exact optimal bandwidth when the underlying density presents complex distributional characteristics and the sample size is not large. This is quite a striking conclusion since the asymptotic optimal bandwidth is the usual target bandwidth for plug-in bandwidth selection methods for estimating the underlying probability density function.

7. Proofs

Proof of Theorem 1: From (10) and (11) we have

$$n\mathrm{IV}(f;h,n) = d_h(K)^2 \left(\int_0^{2\pi}\int_0^{2\pi} K_h(\theta-u)^2 f(u)dud\theta - C(f;h) \right),$$

and

$$\mathrm{ISB}(f;h) = d_h(K)^2 C(f;h) - 2d_h(K)D(f;h) + \int_0^{2\pi} f(\theta)^2 d\theta,$$

where, by using the 2π-periodicity of K_h and f, we have

$$d_h(K)^{-1} = \int_{-\pi}^{\pi} K_h(\theta)d\theta \tag{16}$$

$$= \int_{-\pi/h}^{\pi/h} K(u)du, \tag{17}$$

$$\begin{aligned}C(f;h) &= \int_0^{2\pi} (K_h * f)(\theta)^2 d\theta \\ &= \int_0^{2\pi}\int_0^{2\pi}\int_0^{2\pi} K_h(\theta-x)K_h(\theta-y)f(x)f(y)dxdyd\theta \\ &= \int_0^{2\pi}\int_{\theta-2\pi}^{\theta}\int_{\theta-2\pi}^{\theta} K_h(u)K_h(v)f(\theta-u)f(\theta-v)dudvd\theta\end{aligned}$$

$$= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} K_h(u)K_h(v)(\bar{f} * f)(u-v)dudv \tag{18}$$

$$= \int_{-\pi/h}^{\pi/h}\int_{-\pi/h}^{\pi/h} K(u)K(v)(\bar{f} * f)(h(u-v))dudv, \tag{19}$$

and

$$\begin{aligned}D(f;h) &= \int_0^{2\pi} K_h * f(\theta)f(\theta)d\theta \\ &= \int_0^{2\pi}\int_0^{2\pi} K_h(\theta-x)f(x)f(\theta)dxd\theta \\ &= \int_0^{2\pi}\int_{\theta-2\pi}^{\theta} K_h(u)f(\theta-u)f(\theta)dud\theta\end{aligned}$$

$$= \int_{-\pi}^{\pi} K_h(u)(\bar{f} * f)(u)du \tag{20}$$

$$= \int_{-\pi/h}^{\pi/h} K(u)(\bar{f} * f)(hu)du. \tag{21}$$

Finally, it is enough to see that

$$\int_0^{2\pi}\int_0^{2\pi} K_h(\theta-u)^2 f(u)dud\theta = \int_0^{2\pi}\int_{-\pi}^{\pi} K_h(v)^2 f(u)dudv = h^{-1}d_h(K^2)^{-1}. \quad \blacksquare$$

In the following lemmas we summarise the behaviour of the functions $h \mapsto d_h(K)$, $h \mapsto C(f;h)$ and $h \mapsto D(f;h)$. The continuity of these functions and their limit behaviour as $h \to 0$ follow from (17), (19) and (21) by using the integrability of K and the continuity of $\bar{f} * f$. Their limit behaviour as $h \to +\infty$ follows from (16), (18) and (20).

Lemma 1. *If K is a kernel on $\mathbb{R}$, then*

$$\lim_{h\to 0} d_h(K)^{-1} = \int_{\mathbb{R}} K(u)du, \tag{22}$$

and the function $h \mapsto d_h(K)$ is continuous on $]0,+\infty[$ whenever $\int_{-\pi/h}^{\pi/h} K(u)du \neq 0$, for all $h > 0$. Moreover, if K is continuous at zero then

$$\lim_{h\to +\infty} h d_h(K)^{-1} = 2\pi K(0).$$

Lemma 2. *If $f \in L_2([0,2\pi[)$ and K is a kernel on $\mathbb{R}$, then the functions $h \mapsto C(f;h)$ and $h \mapsto D(f;h)$ are continuous on $]0,+\infty[$ with*

$$\lim_{h\to 0} C(f;h) = \left(\int_{\mathbb{R}} K(u)du\right)^2 \int_0^{2\pi} f(\theta)^2 d\theta \tag{23}$$

and

$$\lim_{h\to 0} D(f;h) = \int_{\mathbb{R}} K(u)du \int_0^{2\pi} f(\theta)^2 d\theta. \tag{24}$$

Moreover, if K is continuous at zero we have

$$\lim_{h\to +\infty} h^2 C(f;h) = 2\pi K(0)^2,$$

and

$$\lim_{h\to +\infty} hD(f;h) = K(0).$$

Taking into account Theorem 1 and the limit behaviour, as $h \to 0$, of the functions $h \mapsto d_h(K)$, $h \mapsto C(f;h)$ and $h \mapsto D(f;h)$, stated in Lemmas 1 and 2, the proofs of Theorems 2 and 3 follow from standard arguments. They are presented here for completeness reasons.

Proof of Theorem 2: Taking into account (22), (23) and (24), from (12) and (13) we get

$$\mathrm{IV}(f;h,n) = (nh)^{-1} \boldsymbol{d}_1(K) + o\big((nh)^{-1}\big), \tag{25}$$

and $\mathrm{ISB}(f;h,n) = o(1)$, as $h \to 0$, where $\boldsymbol{d}_1(K)$ is given by (6). The stated convergence to zero of $\mathrm{MISE}(f;h,n)$, as n tends to infinity, follows from the assumption $nh \to +\infty$, as $n \to +\infty$. Moreover, if K is symmetric with $\int_{\mathbb{R}} u^2|K(u)|du < \infty$, from (10) and standard reasoning based on the Taylor expansion of f we have

$$\mathrm{E}\hat{f}_n(\theta;h) - f(\theta) = d_h(K)h^2 \int_{-\pi/h}^{\pi/h} v^2 K(v) \int_0^1 (1-t) f''(\theta - thv) dt dv,$$

from which we get

$$\begin{aligned}
&\mathrm{ISB}(f;h)\\
&= d_h(K)^2 h^4 \int_{-\pi/h}^{\pi/h}\int_{-\pi/h}^{\pi/h}\int_0^1\int_0^1 u^2K(u)v^2K(v)(1-s)(1-t)\\
&\qquad\qquad\qquad\qquad\qquad\qquad \times (\bar{f}''\ast f'')(h(su-tv))dsdtdudv\\
&= \frac{h^4}{4}\boldsymbol{d}_2(K)\int_0^{2\pi} f''(\theta)^2 d\theta + o\big(h^4\big),\ \text{ as } h\to 0, \qquad (26)
\end{aligned}$$

by using the continuity of $\bar{f}''\ast f''$ at the origin, that follows from the square integrability of f'' on $[0,2\pi[$, and $\boldsymbol{d}_2(K)$ is given by (7). The stated expansion for $\mathrm{MISE}(f;h,n)$ follows from (25) and (26). ■

Proof of Theorem 3: Assuming that K has support $[-M,M]$, for some $M>0$, we have

$$K_h\ast f(\theta)=\int_{-\pi/h}^{\pi/h}K(v)f(\theta-vh)dv=\int_{\mathbb{R}}K(v)f(\theta-vh)dv,$$

for $0<h\leq \pi/M$. Therefore, for small values of h the convolution $K_h\ast f$ has the same form as the corresponding convolution for the kernel estimator of linear data. Again, using the standard arguments based on the Taylor expansion of f we have

$$\mathrm{E}\hat{f}_n(\theta;h)-f(\theta)=d_h(K)\frac{1}{(k-1)!}h^k\int_{\mathbb{R}}\int_0^1 u^kK(u)(1-t)^{k-1}f^{(k)}(\theta-tuh)dtdu,$$

from which we deduce the expansion

$$\mathrm{ISB}(f;h)=\frac{h^{2k}}{(k!)^2}\boldsymbol{d}_k(K)\int_0^{2\pi} f^{(k)}(\theta)^2d\theta+o\big(h^{2k}\big),\ \text{ as } h\to 0, \qquad (27)$$

by using the continuity of $\overline{f^{(k)}}\ast f^{(k)}$ at the origin. The stated expansion for $\mathrm{MISE}(f;h,n)$ follows from (25) and (27). ■

Proof of Theorem 4: As the kernel K satisfies (K.1) and (K.2), from Lemmas 1 and 2, (12) and (13) we deduce that for each $n\in\mathbb{N}$ the functions $h\mapsto \mathrm{IV}(f;h,n)$ and $h\mapsto \mathrm{ISB}(f;h)$, and therefore $h\mapsto \mathrm{MISE}(f;h,n)$, are continuous on $]0,+\infty[$ with

$$\lim_{h\to 0}\mathrm{IV}(f;h,n)=+\infty,\qquad \lim_{h\to+\infty}\mathrm{IV}(f;h,n)=0,$$

$$\lim_{h\to 0}\mathrm{ISB}(f;h,n)=0,\qquad \lim_{h\to+\infty}\mathrm{ISB}(f;h,n)=\int_0^{2\pi}\Big(f(\theta)-\frac{1}{2\pi}\Big)^2d\theta \tag{28}$$

$$\lim_{h\to 0}\mathrm{MISE}(f;h,n)=+\infty \text{ and } \lim_{h\to+\infty}\mathrm{MISE}(f;h,n)=\int_0^{2\pi}\Big(f(\theta)-\frac{1}{2\pi}\Big)^2d\theta>0. \tag{29}$$

Taking into account that $\mathrm{IV}(f;h,n)=n^{-1}\mathrm{IV}^*(f;h)$, where $\mathrm{IV}^*(f;h)$ is independent of n, from (28) and the continuity of $h\mapsto \mathrm{ISB}(f;h)$ we conclude that there exist $h'>0$ and $m\in\mathbb{N}$ such that for all $n\geq m$ we have

$$\mathrm{MISE}(f;h',n)=n^{-1}\mathrm{IV}^*(f;h')+\mathrm{ISB}(f;h')<\int_0^{2\pi}\Big(f(\theta)-\frac{1}{2\pi}\Big)^2d\theta.$$

Therefore, from (29) and the continuity of $h\mapsto\mathrm{MISE}(f;h,n)$ on $]0,+\infty[$, we conclude that for all $n\geq m$ there exists $h_{\mathrm{MISE}}(f;K,n)>0$ such that $\mathrm{MISE}(f;h_{\mathrm{MISE}}(f;K,n),n)\leq\mathrm{MISE}(f;h,n)$, for all $h>0$. ■

Proof of Theorem 5: Using the fact that K fulfills assumption (K.3), from (16), (18) and (20) we have

$$\lim_{h\to+\infty}h^p\big(hd_h(K)^{-1}-2\pi K(0)\big)=\int_{-\pi}^{\pi}\ell_K(x)dx,$$

$$\lim_{h\to+\infty}h^p\big(h^2C(f;h)-2\pi K(0)^2\big)=2K(0)\int_{-\pi}^{\pi}\ell_K(x)dx$$

and

$$\lim_{h\to+\infty}h^p\big(hD(f;h)-K(0)\big)=\int_{-\pi}^{\pi}\ell_K(x)\bar f*f(x)dx.$$

Therefore, from (12) and (13) we get

$$\lim_{h\to\infty}h^p\left(\mathrm{MISE}(f;h)-\int_0^{2\pi}\Big(f(\theta)-\frac{1}{2\pi}\Big)^2d\theta\right)$$
$$=-\frac{1}{\pi K(0)}\int_{-\pi}^{\pi}\ell_K(x)\Big(\bar f*f(x)-\frac{1}{2\pi}\Big)dx<0.$$

Hence, for all $n\in\mathbb{N}$ there exists $h'=h'(n)>0$ large enough such that $\mathrm{MISE}(f;h',n)<\int_0^{2\pi}\big(f(\theta)-\frac{1}{2\pi}\big)^2d\theta$, which concludes the proof. ■

Proof of Theorem 6: We omit this proof because it is similar to the proof of Theorem 2 in Chacón et al. (2007). We only mention the fact, used below, that it is based on the convergence

$$\lim_{n\to+\infty} \mathrm{MISE}(f; h_{\mathrm{MISE}}(f;K,n), n) = 0, \tag{30}$$

where $h_{\mathrm{MISE}}(f;K,n) > 0$ is the exact optimal bandwidth. ■

Proof of Theorem 7: We begin by proving that $\mathrm{ISB}(f;h) > 0$, for all $h > 0$. For that, let us write $\mathrm{ISB}(f;h) = \int_{-\pi}^{\pi} \{\dot{K}_h * f(\theta) - f(\theta)\}^2 d\theta$, where $\dot{K}_h = d_h(K)K_h$ is a symmetric circular density for all $h > 0$. As $\dot{K}_h * f$ and f are square integrable on $[-\pi, \pi[$, from Parseval's identity (see Butzer and Nessel, 1971, Proposition 4.2.2) we have

$$\mathrm{ISB}(f;h) = \frac{1}{2\pi} \sum_{k=-\infty}^{+\infty} |\varphi_{\dot{K}_h * f - f}(k)|^2 = \frac{1}{\pi} \sum_{k=1}^{+\infty} |\varphi_{\dot{K}_h}(k) - 1|^2 |\varphi_f(k)|^2,$$

where $\varphi_g(k) = \int_{-\pi}^{\pi} g(u) e^{iku} du$, for $k \in \mathbb{Z}$, is the Fourier transform of $g \in L_2([-\pi,\pi[)$ (see Butzer and Nessel, 1971, §4.2.1). If $\mathrm{ISB}(f;h) = 0$, for some $h > 0$, we necessarily have $|\varphi_f(k)| \neq 0$, for some $k \in \mathbb{N}$, as f is not the uniform density. Thus, we must have $\varphi_{\dot{K}_h}(k) = 1$ which contradicts the fact that $\dot{K}_h$ is a symmetric circular density.

Let us now prove that $h_{\mathrm{MISE}}(n) := h_{\mathrm{MISE}}(f;K,n) \to 0$, as $n \to \infty$. Reasoning again by contradiction, suppose that $h_{\mathrm{MISE}}(n) \not\to 0$, as n tends to infinity. Thus there exists a subsequence $(h_{\mathrm{MISE}}(n_k))$ of $(h_{\mathrm{MISE}}(n))$ such that $h_{\mathrm{MISE}}(n_k) \to \lambda \in]0, +\infty]$. From the continuity of $h \mapsto \mathrm{ISB}(f;h)$ on $]0,+\infty[$ and (28), we have $\lim_{k\to+\infty} \mathrm{ISB}(f; h_0(n_k)) = \mathrm{ISB}(f;\lambda) > 0$, with $\mathrm{ISB}(f;\infty) := \int_0^{2\pi} \big(f(\theta) - \frac{1}{2\pi}\big)^2 d\theta$, which contradicts (30). ■

Proof of Theorem 8: This proof will be not fully detailed here because it follows from the proof for the PR-estimator for linear data. The asymptotic equivalence between the bandwidths $h_{\mathrm{AMISE}} := h_{\mathrm{MISE}}(f;K,n)$ and $h_{\mathrm{MISE}} := h_{\mathrm{MISE}}(f;K,n)$ follows directly from the asymptotic expansion (5) established in Theorem 2 and the fact that h_{MISE} satisfies the limit conditions established in Theorems 6 and 7 (for the details, see Chacón, 2004, Teorema 2.33, pp. 44–46). Assuming that the kernel K has support $[-M, M]$, for some $M > 0$, from (14) we deduce that the MISE of estimator (1) can be written as

$$\begin{aligned}
&\mathrm{MISE}(f;h,n) \\
&= (nh)^{-1} \boldsymbol{d}_1(K) + k_0^{-2}\big(1 - n^{-1}\big) C(f;h) - 2k_0^{-1} D(f;h) + \int_0^{2\pi} f(\theta)^2 d\theta,
\end{aligned}$$

for $0 < h \leq \pi/M$, where $k_0 = \int_{\mathbb{R}} K(u)du$, $C(f;h) = \int_{\mathbb{R}} \varphi(u)(\bar{f} * f)(hu)du$, with $\varphi(u) = \int_{\mathbb{R}} K(u+v)K(v)dv$, and $D(f;h) = \int_{\mathbb{R}} K(u)(\bar{f} * f)(hu)du$. Taking into account that f is three-times continuously differentiable on $[0, 2\pi[$, we conclude that the function $h \mapsto \mathrm{MISE}(f;h,n)$ is twice continuously differentiable on $]0, \pi/M[$, with

$$\begin{aligned}&\mathrm{MISE}'(f;h,n)\\ &= -(nh^2)^{-1}\boldsymbol{d}_1(K) + h^3 k_0^{-2} k_2^2 \theta_2(f) - \frac{h^5}{4} k_0^{-2} k_2 k_4 \theta_3(f) + o\big(h^5\big) + O\big(n^{-1}h\big)\end{aligned}$$

and

$$\mathrm{MISE}''(f;h,n) = 2(nh^3)^{-1}\boldsymbol{d}_1(K) - 3h^2 k_0^{-2} k_2^2 \theta_2(f) + o\big(h^2\big) + O\big(n^{-1}\big),$$

as $h \to 0$, where $k_r = \int_{\mathbb{R}} u^r K(u)du$ and $\theta_r(f) = \int_0^{2\pi} f^{(r)}(\theta)^2 d\theta$. Thus, by the Taylor expansion of $\mathrm{MISE}'(f;h,n)$ about $\mathrm{MISE}'(f;h_{\mathrm{MISE}},n) = 0$ we get

$$h_{\mathrm{AMISE}}/h_{\mathrm{MISE}} - 1 = h_{\mathrm{MISE}}^{-1} \mathrm{MISE}'(f;h_{\mathrm{AMISE}},n)\mathrm{MISE}''(f;\tilde{h},n)^{-1},$$

with $\tilde{h}$ between h_{MISE} and h_{AMISE}. This enables us to conclude as $h_{\mathrm{MISE}} = O(n^{-1/5})$, $\mathrm{MISE}'(f;\, h_{\mathrm{AMISE}}, n) = O(n^{-1})$ and $\mathrm{MISE}''(f;\tilde{h},n) = O(n^{-2/5})$. ∎

Funding

Research partially supported by the Centre for Mathematics of the University of Coimbra – UID/MAT/00324/2020, funded by the Portuguese Government through FCT/MCTES.

References

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

Beran, R. (1979). Exponential models for directional data. *Ann. Statist.* 7, 1162–1178.

Butzer, P.L., Nessel, R.J. (1971). *Fourier analysis and approximation. One-dimensional theory.* Basel: Birkhauser.

Chacón, J.E. (2004). *Estimación de densidades: algunos resultados exactos y asintóticos.* PhD Thesis. Universidad de Extremadura, Spain.

Chacón, J.E., Montanero, J., Nogales, A.G., Pérez, P. (2007). On the existence and limit behavior of the optimal bandwidth in kernel density estimation. *Statist. Sinica* 17, 289–300.

Fisher, N.I. (1993). *Statistical analysis of circular data.* Cambridge: Cambridge University Press.

García-Portugués, E. (2013). Exact risk improvement of bandwidth selectors for kernel density estimation with directional data. *Electron. J. Stat.* 7, 1655–1685.

García-Portugués, E., Crujeiras, R.M., González-Manteiga, W. (2013). Kernel density estimation for directional-linear data. *J. Multivariate Anal.* 121, 152–175.

Hall, P., Watson, G.S., Cabrera, J. (1987). Kernel density estimation with spherical data. *Biometrika* 74, 751–762.

Hall, P., Marron, J.S. (1991). Lower bounds for bandwidth selection in density estimation. *Probab. Theory Related Fields* 90, 149–173.

Klemela, J. (2000). Estimation of densities and derivatives of densities with directional data. *J. Multivariate Anal.* 73, 18–40.

Marron, J.S., Wand, M.P. (1992). Exact mean integrated squared error. *Ann. Statist.* 20, 712–736.

Oliveira, M., Crujeiras, R.M., Rodríguez-Casal, A. (2012). A plug-in rule for bandwidth selection in circular density estimation. *Comp. Stat. Data Anal.* 56, 3898–3908.

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* 33, 1065–1076.

R Development Core Team (2019). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org

Rosenblatt, M. (1956). Remarks on some non-parametric estimates of a density function. *Ann. Math. Statist.* 27, 832–837.

Silverman, B.W. (1986). *Density estimation for statistics and data analysis*. New York: Chapman & Hall.

Taylor, C.C. (2008). Automatic bandwidth selection for circular density estimation. *Comput. Statist. Data Anal.* 52, 3493–3500.

Tenreiro, C. (2022). Kernel density estimation for circular data: a Fourier series-based plug-in approach for bandwidth selection. *J. Nonparametr. Stat.* 34, 377–406.

Wand, M.P., Jones, M.C. (1995). *Kernel smoothing*. New York: Chapman & Hall.

Carlos Tenreiro
CMUC, Department of Mathematics, University of Coimbra, 3000–143 Coimbra, Portugal
E-mail address: tenreiro@mat.uc.pt