Kernel density estimation for circular data: a Fourier series-based plug-in approach for bandwidth selection*

Carlos Tenreiro†

CMUC, Department of Mathematics, University of Coimbra, Coimbra, Portugal

13 April 2022

Abstract

In this paper we derive asymptotic expressions for the mean integrated squared error of a class of delta sequence density estimators for circular data. This class includes the class of kernel density estimators usually considered in the literature, as well as a new class which is closer in spirit to the class of Parzen–Rosenblatt estimators for linear data. For these two classes of kernel density estimators, a Fourier series-based direct plug-in approach for bandwidth selection is presented. The proposed bandwidth selector has a $n^{-1/2}$ relative convergence rate whenever the underlying density is smooth enough and the simulation results testify that it presents a very good finite sample performance against other bandwidth selectors in the literature.

KEYWORDS: Circular data; Kernel density estimation; Bandwidth selection; Plug-in rule; Fourier series-based estimators.

AMS 2010 SUBJECT CLASSIFICATIONS: 62G07, 62G20

*The Version of Record of this manuscript has been published and is available in Journal of Nonparametric Statistics (Vol. 34 (2), 2022, 377–406). https://doi.org/10.1080/10485252.2022.2057974

†ORCID ID: 0000-0002-5495-6644. E-mail: tenreiro@mat.uc.pt URL: http://www.mat.uc.pt/∼tenreiro/ Postal address: CMUC, Department of Mathematics, University of Coimbra, Apartado 3008, 3001–501 Coimbra, Portugal.

1 Introduction

Given an independent and identically distributed sample of angles $X_1, \ldots, X_n \in [0, 2\pi[$ from some absolutely continuous random variable X with unknown probability density function f, a delta sequence estimator is an estimator of f that takes the form

$$\hat{f}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \delta_n(\theta - X_i), \tag{1}$$

where $\theta \in [0, 2\pi[$ and $\delta_n : \mathbb{R} \to [0, \infty[$, for $n \in \mathbb{N}$, is a sequence of periodic functions with period 2π, called delta function sequence, which satisfies the conditions

(Δ.1) $\displaystyle\int_{-\pi}^{\pi} \delta_n(y)dy = 1$, for all n;

(Δ.2) $\displaystyle\sup_{\lambda<|y|\leq\pi} \delta_n(y) \to 0$, as $n \to +\infty$, for all $0 < \lambda < \pi$;

(Δ.3) $\displaystyle\int_{-\pi}^{\pi} \delta_n(y)^2 dy < \infty$, for all n.

The reader is referred to Watson and Leadbetter (1964) for the concept of delta function sequence in the context of linear data. This class of density estimators includes some density estimators already considered in the literature. If the delta function sequence is symmetric, that is $\delta_n(-y) = \delta_n(y)$, for all $y \in \mathbb{R}$ and $n \in \mathbb{N}$, the previous class is closely related with the subclass of nonnegative circular estimators (of second sin-order) considered in Di Marzio, Panzera, and Taylor (2009) (see also Di Marzio, Panzera, and Taylor 2011, Definitions 1 and 2, pp. 2157–2158). Conditions (ii) and (iii) of Definition 1 in Di Marzio et al. (2009, p. 2067) are trivially fulfilled under assumptions (Δ.1) and (Δ.2), being this last condition stronger than condition (iii). However, under the previous conditions the delta function sequence does not necessarily admit a pointwise convergent Fourier series representation as assumed in condition (i) of Definition 1 in Di Marzio et al. (2009). Nevertheless, from assumption (Δ.3) we know that the delta function sequence admits the $L_2([-\pi, \pi])$ Fourier series representation

$$\delta_n(y) = \frac{1}{2\pi}\left(1 + 2\sum_{k=1}^{\infty} \left\{a_k(\delta_n)\cos(ky) + b_k(\delta_n)\sin(ky)\right\}\right), \tag{2}$$

where

$$a_k(\delta_n) = \int_{-\pi}^{\pi} \delta_n(y)\cos(ky)dy \quad \text{and} \quad b_k(\delta_n) = \int_{-\pi}^{\pi} \delta_n(y)\sin(ky)dy, \tag{3}$$

are, up to a constant, the real Fourier coefficients of δ_n (Butzer and Nessel 1971, Definition 1.2.1 and Proposition 4.2.3, pp. 40, 175).

Kernel type methods for estimating densities of q-dimensional unit spheres, for $q \geq 1$, were initially studied in Beran (1979), Hall, Watson, and Cabrera (1987), Bai, Rao, and Zhao (1988) and Klemelä (2000), this last work being restricted to $q \geq 2$. The class of delta sequence estimators includes the class of kernel estimators considered in the first three of these works in the specific case of circular data. For $\theta \in [0, 2\pi[$, they are defined by

$$\tilde{f}_L(\theta; h) = \frac{c_h(L)}{n} \sum_{i=1}^{n} L\left(\frac{1 - \cos(\theta - X_i)}{h^2}\right), \tag{4}$$

where $L : [0, \infty[\to \mathbb{R}$ is a bounded function satisfying some additional conditions, $h = h_n$ is a sequence of positive numbers such that $h_n \to 0$, as $n \to \infty$, and $c_h(L)$, depending on the kernel L and the bandwidth h, is chosen so that $\tilde{f}_L(\cdot; h)$ integrates to unity. If L is the so-called von Mises kernel $L(t) = e^{-t}$, then (4) is the density estimator considered in Taylor (2008) and Oliveira, Crujeiras, and Rodríguez-Casal (2012) denoted henceforth by $\tilde{f}_{\mathrm{vM}}$. 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 h^{-2} as it takes the form

$$\tilde{f}_{\mathrm{vM}}(\theta; h) = \frac{1}{n} \sum_{i=1}^{n} f_{\mathrm{vM}}(\theta; X_i, h^{-2}), \tag{5}$$

where

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

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

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

The class of delta sequence estimators also comprises an estimator that is closer in spirit to the Parzen–Rosenblatt estimator for linear data (Rosenblatt 1956; Parzen 1962). For $\theta \in [0, 2\pi[$, it is defined by

$$\check{f}_K(\theta; g) = \frac{d_g(K)}{n} \sum_{i=1}^{n} K_g(\theta - X_i), \tag{6}$$

where K_g is a real-valued periodic function on $\mathbb{R}$, with period 2π, such that $K_g(\theta) = K(\theta/g)/g$, for $\theta \in [-\pi, \pi[$, with $K : \mathbb{R} \to \mathbb{R}$ a bounded function satisfying some additional conditions, $g = g_n > 0$ is the bandwidth, and $d_g(K)$ is a normalising constant depending on the kernel K and the bandwidth g which is chosen so that $\check{f}_K(\cdot; g)$ integrates to unity. Of course, for $K(u) = L(u^2)$ and $g = \sqrt{2}\, h$ the estimators $\tilde{f}_L$ and $\check{f}_K$ are closely related and it is expected that there will be no significant differences between them. The results presented in this paper will support this statement. If $L(t) = e^{-t}$ we get $K(u) = e^{-u^2}$, in which case $\check{f}_K$ is close to a kernel density estimator for linear data based on the normal or Gaussian kernel. Although the rationale behind its construction can be found in Silverman (1986, pp. 29–32), to the best of our knowledge it is the first time that this estimator is explicitly proposed and studied in the literature.

This work has two complementary purposes. The first one, which is addressed in Sections 2 and 3, comprises the study of the consistency of the delta sequence estimator $\hat{f}_n$ defined at (1) as an estimator of f and, under some additional assumptions on the delta function sequence, the derivation of an asymptotic expansion for the mean integrated squared error (MISE) of $\hat{f}_n$ defined by

$$\mathrm{MISE}(f;\hat{f}_n,n) := \mathrm{E}\big(\mathrm{ISE}(f;\hat{f}_n,n)\big) = \mathrm{E}\int_0^{2\pi}\big\{\hat{f}_n(\theta)-f(\theta)\big\}^2 d\theta. \tag{7}$$

Such an expansion generalises the one obtained in Di Marzio et al. (2009, Theorem 1, p. 2068). The specific cases of kernel estimators $\tilde{f}_L$, $\tilde{f}_{\mathrm{vM}}$ and $\check{f}_K$, as well as the delta sequence estimator $\hat{f}_{\mathrm{wC}}$ based on the wrapped Cauchy kernel considered in Tsuruta and Sagae (2017a), are discussed in detail. As in kernel estimation for linear data, the asymptotic expansions obtained for the mean integrated squared error of estimators $\tilde{f}_L$ and $\check{f}_K$, besides permitting the derivation of explicit expressions for their asymptotic optimal bandwidths, enable us to compare the two estimators regarding their mean integrated squared error asymptotic performance, and also to identify the optimal kernels for each of these classes of estimators. Additionally, the efficiencies of other kernels with respect to the optimal ones can also be quantified. As in optimal kernel theory for linear data, we conclude that one loses very little when suboptimal kernels are used.

The second purpose of this work is the automatic selection of the smoothing parameter for kernel estimators $\tilde{f}_L$ and $\check{f}_K$. In addition to the seminal paper of Hall et al. (1987) where two cross-validation methods for bandwidth selection are considered, some more recent works where several plug-in bandwidth selectors are proposed, some of them restricted to the case of circular data, comprise the papers of Taylor (2008), Di Marzio et al. (2009), Oliveira et al. (2012), García-Portugués, Crujeiras, and González-Manteiga (2013) and García-Portugués (2013). Following the strategy of Tenreiro (2011), we propose in Section 4 an alternative Fourier series-based direct plug-in approach for selecting the bandwidths of kernel estimators $\tilde{f}_L$ and $\check{f}_K$, and we prove that the proposed selectors achieve the relative convergence rate $n^{-1/2}$ whenever the underlying density is smooth enough (in the context of linear data, see also Tenreiro 2020). These theoretical properties, which are not shared by other existing plug-in methods, but principally because of the very good finite sample performance they possess, provide very strong evidence that the proposed selectors might present a good overall behaviour for a wide range of circular density features.

The remaining sections of this work are organised as follows. In Section 5, the finite-sample behaviour of the new Fourier series-based direct plug-in bandwidth selector for estimator $\tilde{f}_{\mathrm{vM}}$ is illustrated by means of a Monte Carlo study, and in Section 6, it is used in two real data sets. Finally, in Section 7, we draw some overall conclusions. For the convenience of exposition all the proofs are deferred to Section 8 and some of the simulation results are relegated to the online supplementary material. The simulations and plots in this paper were performed using programs written in the R language (R Development Core Team 2019).

2 Bias and variance

As X is a circular random variable that takes values on $[0, 2\pi[$, the probability density function f of X is a nonnegative valued function defined on the interval $[0, 2\pi[$ such that $\int_0^{2\pi} f(\theta)d\theta = 1$. For the sake of simplicity we will 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 $k \in \mathbb{Z}$.

2.1 Asymptotic behaviour

In the following result we establish the consistency in mean squared error of the delta sequence estimator (1) as estimator of f, for all density f continuous on $[0, 2\pi]$. As the established consistency is uniform in $\theta \in [0, 2\pi[$, from this result we also deduce the convergence to zero of the mean integrated squared error of the estimator.

Theorem 1. *Under assumptions* (Δ.1)–(Δ.3)*, if f is continuous on* $[0, 2\pi]$ *we have:*

a)

$$\sup_{\theta\in[0,2\pi[} |\mathrm{E}\hat{f}_n(\theta) - f(\theta)| \to 0.$$

b)

$$\sup_{\theta\in[0,2\pi[} \left| n\alpha(\delta_n)^{-1}\, \mathrm{Var}\hat{f}_n(\theta) - f(\theta)\right| \to 0,$$

where

$$\alpha(\delta_n) = \int_{-\pi}^{\pi} \delta_n(y)^2 dy \to +\infty. \tag{8}$$

Moreover, if $n\alpha(\delta_n)^{-1} \to +\infty$ *we have*

$$\sup_{\theta\in[0,2\pi[} \mathrm{E}\big(\hat{f}_n(\theta) - f(\theta)\big)^2 \to 0.$$

A simple example of a delta function sequence satisfying assumptions (Δ.1)–(Δ.3) is given by the so-called wrapped Cauchy kernel defined, for $y \in \mathbb{R}$, by

$$\delta_n(y) = \frac{1}{2\pi}\left(1 + 2\sum_{k=1}^{\infty} \rho^k \cos(ky)\right) = \frac{1}{2\pi}\frac{1-\rho^2}{1+\rho^2 - 2\rho\cos y}, \tag{9}$$

where the concentration parameter $\rho = \rho_n$ is such that $0 < \rho < 1$ and $\rho \to 1$, as n tends to infinity (see Mardia and Jupp 2000, p. 51, and Tsuruta and Sagae 2017a, sec. 3). Taking into account that

$$\alpha(\delta_n) = \frac{1}{2\pi}\frac{1+\rho^2}{1-\rho^2} = \frac{1}{2\pi h}(1 + o(1)),$$

where $h = 1 - \rho$, from Theorem 1 we conclude that the delta sequence estimator $\hat{f}_{\mathrm{wC}}$ based on the wrapped Cauchy density is a consistent estimator for all density continuous on $[0, 2\pi]$, whenever $h \to 0$ and $nh \to +\infty$, as $n \to +\infty$.

In the case of estimator $\tilde{f}_L$ given by (4), conditions (Δ.1)–(Δ.3) are fulfilled whenever the smoothing parameter h converges to zero as n tends to infinity, and the kernel $L : [0, \infty[\to \mathbb{R}$, assumed to be nonnegative and bounded, satisfies the additional conditions:

(L.1) $\lim_{t \to +\infty} t^{1/2} L(t) = 0$;

(L.2) $0 < \int_0^\infty t^{-1/2} L(t) dt < \infty$.

Moreover, if $h \to 0$, as $n \to +\infty$, we have

$$\alpha(\delta_n) = h^{-1} 2^{-1/2} \int_0^\infty t^{-1/2} L(t)^2 dt \left(\int_0^\infty t^{-1/2} L(t) dt \right)^{-2} (1 + o(1)). \tag{10}$$

Concerning estimator $\check{f}_K$ given by (6), conditions (Δ.1)–(Δ.3) are fulfilled whenever the smoothing parameter g converges to zero as n tends to infinity, and the kernel $K : \mathbb{R} \to \mathbb{R}$, assumed to be nonnegative, bounded and symmetric, satisfies the conditions:

(K.1) $\lim_{u \to +\infty} uK(u) = 0$;

(K.2) $0 < \int_{-\infty}^\infty K(u) du < \infty$.

In this case, if $g \to 0$, as $n \to +\infty$, we have

$$\alpha(\delta_n) = g^{-1} \int_{-\infty}^\infty K(u)^2 du \left(\int_{-\infty}^\infty K(u) du \right)^{-2} (1 + o(1)). \tag{11}$$

From Theorem 1 and equality (10) we conclude that under conditions (L.1) and (L.2), $\tilde{f}_L$ is a consistent estimator of f, for all density f continuous on $[0, 2\pi]$, whenever the smoothing parameter satisfies the classical conditions $h \to 0$, $nh \to +\infty$, as n tends to infinity. Taking into account (11) a similar result holds for estimator $\check{f}_K$ whenever K satisfies conditions (K.1) and (K.2) and g is such that $g \to 0$, $ng \to +\infty$, as $n \to +\infty$. Note that the kernel L satisfies conditions (L.1) and (L.2) iff the kernel K defined by $K(u) = L(u^2)$ satisfies conditions (K.1) and (K.2). In this case the main terms of the asymptotic expansions (10) and (11) coincide whenever $g = \sqrt{2}\, h$.

2.2 Asymptotic expansions

From Theorem 1, a uniform asymptotic expansion for the variance of the estimator $\hat{f}_n$ is given by

$$\sup_{\theta \in [0, 2\pi[} \left| \mathrm{Var} \hat{f}_n(\theta) - n^{-1} \alpha(\delta_n) f(\theta) \right| = o\big(n^{-1} \alpha(\delta_n)\big). \tag{12}$$

In order to obtain an equally useful asymptotic expansion for the bias of the estimator, the following additional assumptions on the delta function sequence need to be imposed:

(Δ.4) $\int_{-\pi}^{\pi} y \delta_n(y) dy = 0$, for all n;

(Δ.5) $\displaystyle\int_{-\pi}^{\pi} |y|^{2+\gamma}\delta_n(y)dy = o\big(\beta(\delta_n)\big)$, for all $\gamma \in]0,1]$, where $\displaystyle\beta(\delta_n) := \int_{-\pi}^{\pi} y^2\delta_n(y)dy$.

Taking into account their symmetry, the delta function sequences of the estimators $\tilde{f}_L$ and $\check{f}_K$ trivially fulfil the first of the previous assumptions. The second condition holds for estimator $\tilde{f}_L$ if the kernel L is such that

(L.3) $\displaystyle\int_0^{\infty} tL(t)\,dt < \infty$.

In this case we have

$$\beta(\delta_n) = 2h^2 \int_0^{\infty} t^{1/2}L(t)dt \left(\int_0^{\infty} t^{-1/2}L(t)dt\right)^{-1} (1+o(1)). \tag{13}$$

With respect to estimator $\check{f}_K$, condition (Δ.5) holds if the kernel K is such that

(K.3) $\displaystyle\int_{-\infty}^{\infty} |u|^3K(u)du < \infty$.

Moreover, we have

$$\beta(\delta_n) = g^2 \int_{-\infty}^{\infty} u^2K(u)du \left(\int_{-\infty}^{\infty} K(u)du\right)^{-1} (1+o(1)). \tag{14}$$

Kernel L fulfils condition (L.3) iff the kernel K defined by $K(u) = L(u^2)$ satisfies condition (K.3), and the sequences (13) and (14) are asymptotically equivalent whenever $g = \sqrt{2}\,h$.

Theorem 2. *Under assumptions (Δ.1)–(Δ.5), assume that f is twice differentiable on $[0,2\pi]$ and that f'' satisfies the Lipschitz condition*

$$\big|f''(x) - f''(y)\big| \le C|x-y|^{\alpha},\ x,y \in [0,2\pi], \tag{15}$$

for some $\alpha \in]0,1]$ and $C > 0$. We have

$$\sup_{\theta\in[0,2\pi[} \big|\mathrm{E}\hat{f}_n(\theta) - f(\theta) - \frac{1}{2}\beta(\delta_n)f''(\theta)\big| = o\big(\beta(\delta_n)\big), \tag{16}$$

where

$$\beta(\delta_n) = \int_{-\pi}^{\pi} y^2\delta_n(y)dy \to 0. \tag{17}$$

It is possible to improve the asymptotic rate of convergence to zero of the bias of $\hat{f}_n$ by allowing the delta function sequence (δ_n) to take negative values (see Wand and Jones 1995, pp. 32–35, for linear data, and Tsuruta and Sagae 2017b, for circular data). Nevertheless, this issue is not pursued here.

3 MISE expansions and asymptotic optimal bandwidths

The mean integrated squared error defined at (7) is a widely used global measure of the performance of a density estimator $\hat{f}_n$. Under the conditions of Theorem 2 with $n\alpha(\delta_n)^{-1} \to +\infty$, as $n \to +\infty$, from expansions (12) and (16) we deduce that the integrated variance and the integrated squared bias of $\hat{f}_n$ can be expressed as

$$\mathrm{IV}(f; \hat{f}_n, n) := \int_0^{2\pi} \mathrm{Var}\hat{f}_n(\theta) d\theta = n^{-1}\alpha(\delta_n) + o\big(\alpha(\delta_n)\big) \tag{18}$$

and

$$\mathrm{ISB}(f; \hat{f}_n, n) := \int_0^{2\pi} \big\{\mathrm{E}\hat{f}_n(\theta) - f(\theta)\big\}^2 d\theta = \frac{1}{4}\beta(\delta_n)^2\theta_2(f) + o\big(\beta(\delta_n)^2\big), \tag{19}$$

where $\alpha(\delta_n)$ and $\beta(\delta_n)$ are given by (8) and (17), respectively, and $\theta_2(f)$ denotes the quadratic funcional

$$\theta_2(f) = \int_0^{2\pi} f''(\theta)^2 d\theta.$$

In the next result we present two asymptotic expansions for the mean integrated squared error of the delta sequence estimator (1). The first one follows directly from expansions (18) and (19). The second one is valid whenever the delta function sequence is symmetric.

Theorem 3. *Under assumptions* (Δ.1)–(Δ.5)*, assume that f is twice differentiable on $[0, 2\pi]$ and that f'' satisfies the Lipschitz condition (15). If $n\alpha(\delta_n)^{-1} \to +\infty$ then*

$$\mathrm{MISE}(f; \hat{f}_n, n) = n^{-1}\alpha(\delta_n) + \frac{1}{4}\beta(\delta_n)^2\theta_2(f) + o\big(n^{-1}\alpha(\delta_n) + \beta(\delta_n)^2\big). \tag{20}$$

Moreover, if the delta function sequence is symmetric we have

$$\mathrm{MISE}(f; \hat{f}_n, n) = n^{-1}\alpha(\delta_n) + \frac{1}{16}(1 - a_2(\delta_n))^2\theta_2(f) + o\big(n^{-1}\alpha(\delta_n) + (1 - a_2(\delta_n))^2\big), \tag{21}$$

where $a_2(\delta_n)$ is given by (3).

Several results that appear in the literature on circular density estimation follow from this general result. When the delta function sequence (δ_n) is symmetric, from the $L_2([-\pi, \pi])$ representation (2) we have $\alpha(\delta_n) = \frac{1}{2\pi}\big(1 + 2\sum_{k=1}^{\infty} a_k(\delta_n)^2\big)$, and the approximation of the mean integrated squared error is given by

$$\mathrm{AMISE}(f; \hat{f}_n, n) = n^{-1}\alpha(\delta_n) + \frac{1}{16}(1 - a_2(\delta_n))^2\theta_2(f), \tag{22}$$

agrees with the one presented in Di Marzio et al. (2009, Theorem 1, p. 2068). However, from Theorem 3 we can further conclude that this approximation is asymptotically equivalent to the mean integrated squared error of $\hat{f}_n$, a fact that is not established in the work of Di Marzio et al. (2009).

In the particular case of estimator $\tilde{f}_{\mathrm{vM}}$ defined at (5), the delta function sequence, also known as von Mises kernel, is given by $\delta_n(y) = \exp(h^{-2}\cos y)/(2\pi I_0(h^{-2}))$, for $y \in \mathbb{R}$. Therefore, from (22) we get the approximation of the mean integrated squared error given by

$$\mathrm{AMISE}_1(f; \tilde{f}_{\mathrm{vM}}, h, n) = \frac{I_0(2h^{-2})}{2\pi n I_0\big(h^{-2}\big)^2} + \frac{1}{16}\left(1 - \frac{I_2\big(h^{-2}\big)}{I_0\big(h^{-2}\big)}\right)^2 \theta_2(f), \tag{23}$$

when $h \to 0$ and $nh \to \infty$, as

$$\alpha(\delta_n) = \frac{1}{4\pi^2 I_0(h^{-2})^2} \int_{-\pi}^{\pi} \exp\left(2h^{-2}\cos y\right) dy = \frac{I_0(2h^{-2})}{2\pi I_0(h^{-2})^2}$$

and

$$a_2(\delta_n) = \frac{1}{2\pi I_0(h^{-2})} \int_{-\pi}^{\pi} \cos(2y)\exp\left(h^{-2}\cos y\right) dy = \frac{I_2\big(h^{-2}\big)}{I_0\big(h^{-2}\big)}.$$

This approximation, which we now know to be asymptotically equivalent to the mean integrated squared error of $\tilde{f}_{\mathrm{vM}}$, is considered in Oliveira et al. (2012, p. 3899) to define new plug-in bandwidth selectors.

The delta sequence estimator $\hat{f}_{\mathrm{wC}}$ based on the wrapped Cauchy kernel (9) considered before, is an example of an estimator that does not fulfil assumption (Δ.5). In fact, for the wrapped Cauchy kernel we have

$$\int_{-\pi}^{\pi} |y|^{2+\gamma}\delta_n(y) dy = \frac{h}{2\pi}\int_{-\pi}^{\pi} \frac{|y|^{2+\gamma}}{1-\cos y}\, dy\, (1 + o(1)),$$

for all $\gamma \geq 0$, when $h = 1 - \rho_n$ tends to zero, as n tends to infinity. Therefore, although we may conclude from Theorem 1 that $n^{-1}\alpha(\delta_n) = \frac{1}{2\pi nh}(1 + o(1))$ is asymptotically equivalent to the integrated variance of $\hat{f}_{\mathrm{wC}}$, from Theorem 3 we cannot deduce that one of the terms $\frac{1}{4}\beta(\delta_n)^2\theta_2(f)$ or $\frac{1}{16}\big(1 - a_2(\delta_n)\big)^2\theta_2(f)$ is asymptotically equivalent to the integrated square bias of $\hat{f}_{\mathrm{wC}}$. This is clear from the expansion for the mean integrated squared error of $\hat{f}_{\mathrm{wC}}$ we give in the next result which corrects the one presented in Tsuruta and Sagae (2017a, Theorem 2, p. 4).

Theorem 4. *If f continuously differentiable on $[0, 2\pi]$ and $h = 1 - \rho$ is such that $h \to 0$ and $nh \to +\infty$, we have*

$$\mathrm{MISE}(f; \hat{f}_{\mathrm{wC}}, h, n) = \frac{1}{2\pi nh} + h^2\theta_1(f) + o\left(\frac{1}{nh} + h^2\right),$$

where $\theta_1(f)$ denotes the quadratic funcional

$$\theta_1(f) = \int_0^{2\pi} f'(\theta)^2 d\theta.$$

If f is not the circular uniform distribution, the asymptotic optimal bandwidth, that is, the bandwidth that minimises the most significant terms of the mean integrated squared error asymptotic expansion, usually called the asymptotic mean integrated squared error, is given by

$$h^* = (4\pi)^{-1/3}\theta_1(f)^{-1/3}n^{-1/3}.$$

This expression also corrects the expression for the asymptotic optimal bandwidth of $\hat{f}_{\mathrm{wC}}$ given in Tsuruta and Sagae (2017a, eq. 8, p. 4).

In the remaining of this section we discuss some additional consequences of Theorem 3 in the case of the kernel estimators $\tilde{f}_L$ and $\check{f}_K$ defined by (4) and (6), respectively. Especially, we will see that the optimal convergence rate $n^{-2/3}$ we get for the MISE of estimator $\hat{f}_{\mathrm{wC}}$ is overcome by the corresponding optimal convergence rate $n^{-4/5}$ obtained for $\tilde{f}_L$ and $\check{f}_K$.

We consider first the case of the estimator $\tilde{f}_L$ defined at (4).

Theorem 5. *Let L be a nonnegative and bounded kernel satisfying conditions (L.1)–(L.3), and assume that f is under the conditions of Theorem 2. If h is such that $h \to 0$ and $nh \to +\infty$, as $n \to +\infty$, we have*

$$\mathrm{MISE}(f; \tilde{f}_L, h, n) = \frac{1}{nh}\boldsymbol{c}_1(L) + h^4 \boldsymbol{c}_2(L)\,\theta_2(f) + o\Big(\frac{1}{nh} + h^4\Big),$$

where

$$\boldsymbol{c}_1(L) = 2^{-1/2}\int_0^\infty t^{-1/2}L(t)^2 dt \left(\int_0^\infty t^{-1/2}L(t)dt\right)^{-2}$$

and

$$\boldsymbol{c}_2(L) = \left(\int_0^\infty t^{1/2}L(t)dt\right)^2 \left(\int_0^\infty t^{-1/2}L(t)dt\right)^{-2}.$$

If f is not the circular uniform distribution, the asymptotic optimal bandwidth for estimator $\tilde{f}_L$ is given by

$$h^* = \boldsymbol{c}(L)\theta_2(f)^{-1/5}n^{-1/5},$$

where

$$\boldsymbol{c}(L) = 2^{-1/2}\left(\int_0^\infty t^{-1/2}L(t)^2 dt\right)^{1/5}\left(\int_0^\infty t^{1/2}L(t)dt\right)^{-2/5}. \tag{24}$$

Although the previous asymptotic expansion for the mean integrated squared error of $\tilde{f}_L$ is not established in the works of Taylor (2008), Di Marzio et al. (2009) and Oliveira et al. (2012), which address the estimation of densities in the unit circle, it agrees with the one we can deduce from Proposition 1 of García-Portugués (2013), which is valid for the general kernel density estimator on the q-dimensional unit sphere Ω_q with $q \geq 1$ (see also García-Portugués et al. 2013, sec. 4). In fact, under slightly different assumptions on f and L, if we use Proposition 1 of García-Portugués (2013) in the specific case of circular data together with the asymptotic approximation given in equation (2) of García-Portugués et al. (2013), we get for $\mathrm{MISE}(f; \tilde{f}_L, h, n)$ an expansion that is identical to the one given in Theorem 5 but with $\theta_2(f)$ replaced by $\int_{\Omega_1}\Psi(\bar{g},x)^2\omega_1(dx)$, where ω_1 denotes the Lebesgue measure in the unit circle Ω_1, $\bar{g}$ is the real-valued function defined by $\bar{g}(x) = g(x/||x||)$, for $x \in \mathbb{R}^2\backslash\{(0,0)\}$, where $||x||$ denotes the Euclidean norm, and g is the function defined on Ω_1 by $f(\theta) = g(x(\theta))$, with $x(\theta) = (\cos\theta, \sin\theta)$, for $\theta \in [0, 2\pi[$, and, finally, $\Psi(\bar{g},x) = -x^T\nabla\bar{g}(x) + \nabla^2\bar{g}(x) - x^T H_{\bar{g}}(x)x$, where $\nabla\bar{g}(x)$, $\nabla^2\bar{g}(x)$ and $H_{\bar{g}}(x)$ denote, respectively, the gradient vector, the Laplacian and the Hessian matrix of $\bar{g}$ at x. However, as

$f''(\theta) = \Psi(\bar{g}, x(\theta))$, from which we deduce that $\int_{\Omega_1} \Psi(\bar{g}, x)^2 \omega_1(dx) = \int_0^{2\pi} f''(\theta)^2 d\theta = \theta_2(f)$, we conclude that the expansion for $\mathrm{MISE}(f; \tilde{f}_L, h, n)$ we get from Proposition 1 of García-Portugués (2013) in the specific case of circular data is exactly the one presented in Theorem 5.

As seen previously, when $\tilde{f}_{\mathrm{vM}}$ is the estimator given by (5) the MISE approximation given at (23) is asymptotically equivalent to the MISE of $\tilde{f}_{\mathrm{vM}}$. Taking for L the von Mises kernel, from Theorem 5 we get an alternative asymptotic approximation to the MISE of $\tilde{f}_{\mathrm{vM}}$ given by

$$\mathrm{AMISE}_2(f; \tilde{f}_{\mathrm{vM}}, h, n) = \frac{1}{2\sqrt{\pi}\, nh} + \frac{h^4}{4}\, \theta_2(f). \tag{25}$$

This approximation is used in García-Portugués (2013) to define new plug-in bandwidth selectors (we will return later to this point). For small values of h the MISE approximations (23) and (25) are similar as revealed by the quotients between the corresponding integrated variance and integrated squared bias terms which are in an interval centred at one with radius not greater than 0.05 for $0 < h \leq 0.2$.

Assuming that the true density f is a von Mises density with mean direction $\mu \in [0, 2\pi[$ and concentration parameter $\kappa \geq 0$, we denote by $f_{\mathrm{vM}}(\cdot; \mu, \kappa)$, we know that

$$\theta_2(f_{\mathrm{vM}}(\cdot; \mu, \kappa)) = \frac{3\kappa^2 I_0(2\kappa) - \kappa I_1(2\kappa)}{8\pi I_0(\kappa)^2}, \tag{26}$$

which equals the curvature term given in García-Portugués (2013, Proposition 1). Replacing $\theta_2(f)$ by (26) in (25), we get an approximation to the MISE of $\tilde{f}_{\mathrm{vM}}$ when the true density is a von Mises density that is interesting to compare with the approximation considered in Taylor (2008, p. 3495), which can be rewritten as

$$\widetilde{\mathrm{AMISE}}(f; \tilde{f}_{\mathrm{vM}}, h, n) = \frac{1}{2\sqrt{\pi}\, nh} + \frac{3h^4\big(\kappa^2 I_0(2\kappa) - \kappa I_1(2\kappa)\big)}{32\pi I_0(\kappa)^2}, \tag{27}$$

by observing that $I_2(2\kappa) = I_0(2\kappa) - I_1(2\kappa)/\kappa$, due to the properties of the Bessel functions. We see that integrated variance terms agree in both expressions (25) and (27), but the same does not happen with respect to the integrated squared bias terms which are not asymptotically equivalent. Therefore, the MISE approximation (27) considered in Taylor (2008, p. 3495) to define his von Mises-scale plug-in bandwidth selector is not asymptotically equivalent to the MISE of the estimator when the true density f is the von Mises density. The derivation of the approximation (27) in Taylor (2008) was not accurate enough to allow the identification of the most significant terms of the integrated squared bias expansion of estimator $\tilde{f}_{\mathrm{vM}}$.

We turn next to the estimator $\check{f}_K$ defined at (6).

Theorem 6. *Let K be a nonnegative, bounded and symmetric kernel satisfying assumptions (K.1)–(K.3), and assume that f is under the conditions of Theorem 2. If g is such that $g \to 0$ and $ng \to +\infty$, as $n \to +\infty$, we have*

$$\mathrm{MISE}(f; \check{f}_K, g, n) = \frac{1}{ng} \boldsymbol{d}_1(K) + \frac{g^4}{4} \boldsymbol{d}_2(K)\, \theta_2(f) + o\bigg(\frac{1}{ng} + g^4\bigg),$$

where

$$\boldsymbol{d}_1(K) = \int_{-\infty}^{\infty} K(u)^2 du \Big(\int_{-\infty}^{\infty} K(u) du \Big)^{-2}$$

and

$$\boldsymbol{d}_2(K) = \Big(\int_{-\infty}^{\infty} u^2 K(u) du \Big)^2 \Big(\int_{-\infty}^{\infty} K(u) du \Big)^{-2}.$$

If f is not the circular uniform distribution, the asymptotic optimal bandwidth for estimator $\check{f}_K$ is given by

$$g^* = \boldsymbol{d}(K)\theta_2(f)^{-1/5} n^{-1/5},$$

where

$$\boldsymbol{d}(K) = \Big(\int_{-\infty}^{\infty} K(u)^2 du \Big)^{1/5} \Big(\int_{-\infty}^{\infty} u^2 K(u) du \Big)^{-2/5}. \tag{28}$$

The previous formulas agree with the well-known formulas for the mean integrated squared error and the asymptotic optimal bandwidth of the Parzen–Rosenblatt estimator for linear data (see Wand and Jones 1995, p. 21). Moreover, the estimator $\tilde{f}_L$ with kernel L and bandwidth h and the estimator $\check{f}_K$ with kernel $K(u) = L(u^2)$ and bandwidth $g = \sqrt{2}\,h$ share the same first-order asymptotic terms for the corresponding mean integrated squared errors and therefore the same asymptotic optimal bandwidth. This is one more piece of evidence that supports the previously mentioned close relationship between these two kernel density estimators.

The special case of the circular uniform distribution, for which $\theta_2(f) = 0$, is not covered by the previous optimal bandwidth asymptotic theory. For this distribution the bias of the delta sequence estimator is equal to zero and its exact mean integrated squared error is simply given by $\mathrm{MISE}(f; \hat{f}_n, n) = \frac{1}{2\pi n}\big(\alpha(\delta_n) - \frac{1}{2\pi}\big)$, for every delta function sequence satisfying conditions (Δ.1) and (Δ.3). In the particular case of estimator $\tilde{f}_L$, under very general conditions on the kernel L we have $\mathrm{MISE}(f; \tilde{f}_L, h, n) = o(1)$ even when the smoothing parameter does not converge to zero as n tends to infinity. More precisely, if $h \to \lambda \in [0, +\infty]$, as $n \to +\infty$, the fastest rate of convergence is obtained when $\lambda = +\infty$, in which case we get $\mathrm{MISE}(f; \tilde{f}_L, h, n) = o(n^{-1})$. A similar result is valid for estimator $\check{f}_K$.

The asymptotic comparison between estimators $\tilde{f}_L$ and $\check{f}_K$, or between two estimators from one of these classes that use different kernel functions, can be based on the previous asymptotic expansions for the mean integrated squared error. If deterministic smoothing parameters $h = \boldsymbol{c}(L)\gamma\, n^{-1/5}$ and $g = \boldsymbol{d}(K)\gamma\, n^{-1/5}$, with $\gamma > 0$, are respectively used in estimators $\tilde{f}_L$ and $\check{f}_K$, from Theorems 5 and 6 we know that their mean integrated squared errors are such that

$$\mathrm{MISE}(f; \tilde{f}_L, h, n) = \varphi(f;\gamma)\phi(L)\, n^{-4/5}(1 + o(1))$$

and

$$\mathrm{MISE}(f; \check{f}_K, g, n) = \varphi(f;\gamma)\psi(K)\, n^{-4/5}(1 + o(1)),$$

where

$$\varphi(f;\gamma) = \frac{1}{\gamma} + \frac{\gamma^4}{4}\theta_2(f),$$

$L(t)$	$K(u)$	$\text{eff}(L) = \text{eff}(K)$
$I(t \leq 1)$	$I(\lvert u\rvert \leq 1)$	0.9295
$(1-t)I(0 \leq t \leq 1)$	$(1-u^2)I(\lvert u\rvert \leq 1)$	1
$(1-t)^2 I(t \leq 1)$	$(1-u^2)^2 I(\lvert u\rvert \leq 1)$	0.9939
$(1-t)^3 I(t \leq 1)$	$(1-u^2)^3 I(\lvert u\rvert \leq 1)$	0.9867
e^{-t}	e^{-u^2}	0.9512

Table 1: *Efficiencies of kernels L and K with respect to optimal kernels L^* and K^*.*

$$\phi(L) = \left(\int_0^\infty t^{-1/2} L(t)^2 dt\right)^{4/5} \left(\int_0^\infty t^{1/2} L(t) dt\right)^{2/5} \left(\int_0^\infty t^{-1/2} L(t) dt\right)^{-2}$$

and

$$\psi(K) = \left(\int_{-\infty}^\infty K(u)^2 du\right)^{4/5} \left(\int_{-\infty}^\infty u^2 K(u) du\right)^{2/5} \left(\int_{-\infty}^\infty K(u) du\right)^{-2}.$$

This last functional is well-known in the context of kernel estimation for linear data. We know that the parabolic kernel $K^*(u) = (1-u^2)I(|u| \leq 1)$ minimises $\psi(K)$ among all the nonnegative, bounded and symmetric kernels K satisfying conditions (K.1)–(K.3) (see Epanechnikov 1969, Bosq and Lecoutre 1987, pp. 82–83, and Wand and Jones 1995, p. 30). As $\psi(K) = \phi(L)$ for $L(t) = K(\sqrt{t}\,)$, we also deduce that the half-triangular kernel $L^*(t) = (1-t)I(t \leq 1)$ minimises the functional $\phi(L)$ among all the nonnegative and bounded kernels L satisfying conditions (L.1)–(L.3) (see also Hall et al. 1987, p. 758). Therefore, the kernels L^* and K^* are optimal for each one of the classes of estimators $\tilde{f}_L$ and $\check{f}_K$, whenever the considered bandwidths are, respectively, given by $h = \boldsymbol{c}(L)\gamma\, n^{-1/5}$ and $g = \boldsymbol{d}(K)\gamma\, n^{-1/5}$, for some positive value γ. The efficiencies of other kernels with respect to these optimal kernels can be deduced from the previous asymptotic expansions for the mean integrated squared errors. They are given by the ratios $\text{eff}(L) := \big(\phi(L^*)\big/\phi(L)\big)^{5/4}$ and $\text{eff}(K) := \big(\psi(K^*)\big/\psi(K)\big)^{5/4}$. For some kernels, these efficiencies are reported in Table 1. As in kernel estimation for linear data framework (see Wand and Jones 1995, Table 2.1, p. 31), we see that suboptimal kernels may be almost as efficient as the optimal ones.

4 A Fourier series-based plug-in bandwidth selector

When a nonnegative and bounded kernel L satisfying conditions (L.1)–(L.3) is used in (4), or when a nonnegative, symmetric and bounded kernel K satisfying conditions (K.1)–(K.3) is used in (6), under some smoothness assumptions on f we have seen that the asymptotic optimal bandwidths for estimators (4) and (6) are respectively given by

$$h^* = \boldsymbol{c}(L)\,\theta_2(f)^{-1/5}n^{-1/5} \;\text{ and }\; g^* = \boldsymbol{d}(K)\,\theta_2(f)^{-1/5}n^{-1/5}, \tag{29}$$

when $\theta_2(f) = \int_0^{2\pi} f''(\theta)^2 d\theta \neq 0$, and $\boldsymbol{c}(L)$ and $\boldsymbol{d}(K)$ are given by (24) and (28), respectively. As the only unknown quantity in (29) is the density curvature $\theta_2(f)$, the problem of providing data-dependent bandwidth selectors through the estimation of h^* and g^*, is reduced to that of estimating $\theta_2(f)$, this being the rationale of the direct plug-in approach to bandwidth selection. For classical references on the direct plug-in method for linear data the reader is referred to Woodroofe (1970), Nadaraya (1974) and Deheuvels and Hominal (1980).

The approaches developed in the literature in a circular data context to deal with the fact that the quadratic functional $\theta_2(f)$ appearing in the previous theoretical bandwidths is unknown, assume that f is a member of some parametric family of densities, such as the von Mises family, or more generally, the family of von Mises mixtures (cf. García-Portugués 2013; see also Taylor 2008 and Oliveira et al. 2012). After obtaining a suitable approximation $\hat{f}$ of f from the considered reference density family, the automatic bandwidth selectors are given by $\hat{h}^* = \boldsymbol{c}(L)\,\theta_2(\hat{f})^{-1/5}n^{-1/5}$ and $\hat{g}^* = \boldsymbol{d}(K)\,\theta_2(\hat{f})^{-1/5}n^{-1/5}$, where the quantity $\theta_2(\hat{f})$ takes the place of $\theta_2(f)$ in (29). Although $\theta_2(\hat{f})$ can be a good approximation for $\theta_2(f)$, especially when the considered reference density model is able to capture the curvature of the underlying density f, from a theoretical point of view we cannot assure that the data-dependent bandwidths $\hat{h}^*$ and $\hat{g}^*$ are asymptotically equivalent to h^* and g^*, respectively, when f belongs to some large set of circular densities. Wanting to define automatic plug-in bandwidth selectors that possess such a consistency property, we follow in this section the approach of Tenreiro (2011) where the Fourier series-based estimators studied in Laurent (1997) are used to estimate the quadratic functional $\theta_2(f)$.

The Fourier series-based or projection estimator of $\theta_2(f)$ is motivated by the representation

$$\theta_2(f) = \frac{1}{\pi}\sum_{k=1}^{\infty} k^4\big(a_k(f)^2 + b_k(f)^2\big),$$

where $a_k(f)$ and $b_k(f)$ are, up to a constant, the real Fourier coefficients of f given by $a_k(f) = \int_0^{2\pi} f(\theta)\cos(k\theta)d\theta$ and $b_k(f) = \int_0^{2\pi} f(\theta)\sin(k\theta)d\theta$ (also known as the trigonometric moments of X). Based on this representation, which is valid whenever f and f' are absolutely continuous on $[0, 2\pi]$ and f'' is square integrable on $[0, 2\pi]$ (see Butzer and Nessel 1971, Propositions 4.1.8 and 4.2.2, pp. 172, 175), the Fourier series-based estimator of $\theta_2(f)$ is defined by

$$\hat{\hat{\theta}}_{2,m} = \frac{1}{\pi}\sum_{k=1}^{m} k^4\,\hat{\hat{c}}_k, \tag{30}$$

where $\hat{\hat{c}}_k$ is the unbiased estimator of $a_k(f)^2 + b_k(f)^2$ given by

$$\hat{\hat{c}}_k = \frac{2}{n(n-1)}\sum_{1\le i<j\le n} \cos(k(X_i - X_j)), \tag{31}$$

and $m = m(n)$ is a sequence on integers converging to infinity (for related estimators of $\theta_2(f)$ in a linear data context, see Chiu 1991, p. 1891, and Wu 1995, p. 1476). As shown in Laurent (1997), the previous estimator achieves the $n^{-1/2}$ rate of convergence, whenever f is smooth enough and it is efficient. Moreover, when the $n^{-1/2}$ rate is not achievable they achieve the optimal rate of convergence. A closely related alternative positive estimator of $\theta_2(f)$ is

$$\hat{\hat{\theta}}_{2,m} = \theta_2(\tilde{f}_m) = \frac{1}{\pi}\sum_{k=1}^{m} k^4\left(\hat{a}_k^2 + \hat{b}_k^2\right), \tag{32}$$

where

$$\tilde{f}_m(x) = \frac{1}{2\pi}\left(1 + 2\sum_{k=1}^{m}\{\hat{a}_k\cos(kx) + \hat{b}_k\sin(kx)\}\right), \tag{33}$$

is the Fourier series-based estimator of f studied in Kronmal and Tarter (1968), and $\hat{a}_k$ and $\hat{b}_k$ are unbiased estimators of $a_k(f)$ and $b_k(f)$ given by

$$\hat{a}_k = \frac{1}{n}\sum_{i=1}^{n}\cos(kX_i) \quad \text{and} \quad \hat{b}_k = \frac{1}{n}\sum_{i=1}^{n}\sin(kX_i).$$

The number m of Fourier terms plays the role of smoothing parameter and makes the trade-off between the variance and the bias of these estimators. A large value of m implies a small bias but a large variance, whereas a small value of m implies a large bias but a small variance. As in practical situations the choice of m should be based on the observations, this is, $m = \hat{m}(X_1, \ldots, X_n)$, we consider the automatic estimators $\hat{\bar{\theta}}_{2,\hat{m}}$ and $\hat{\hat{\theta}}_{2,\hat{m}}$ of $\theta_2(f)$, whose asymptotic behaviour, established in Tenreiro (2011, Lemma 1, pp. 543–544), enables us to describe the corresponding behaviour of the relative errors of the plug-in bandwidth selector defined by

$$\hat{h}^*_{\hat{m}} = \boldsymbol{c}(L)\,\hat{\theta}_{2,\hat{m}}^{-1/5} n^{-1/5}, \tag{34}$$

where $\hat{\theta}_{2,m}$ denotes either $\hat{\bar{\theta}}_{2,m}$ or $\hat{\hat{\theta}}_{2,m}$ defined by (30) and (32), respectively. Of course, the same asymptotic behaviour can be established for the relative error of the plug-in bandwidth selector defined by

$$\hat{g}^*_{\hat{m}} = \boldsymbol{d}(K)\,\hat{\theta}_{2,\hat{m}}^{-1/5} n^{-1/5}. \tag{35}$$

Theorem 7. *Let L be a nonnegative and bounded kernel satisfying conditions (L.1)–(L.3). For f different from the circular uniform distribution, and $s = p + \alpha > 2$, with $p \in \mathbb{N}$ and $\alpha \in]0,1]$, let us assume that f is p-times differentiable on $[0, 2\pi]$ and that $f^{(p)}$ satisfies the Lipschitz condition*

$$\left|f^{(p)}(x) - f^{(p)}(y)\right| \le C|x-y|^\alpha,\ x, y \in [0, 2\pi],$$

for some $\alpha \in]0,1]$ and $C > 0$.

a) Consistency. *If $\hat{m}$ is such that $\hat{m} \xrightarrow{p} +\infty$ and $n^{-1}\hat{m}^5 \xrightarrow{p} 0$ then*

$$\frac{\hat{h}^*_{\hat{m}}}{h^*} \xrightarrow{p} 1.$$

b) Rates of convergence. *If $\hat{m}$ satisfies*

$$\mathrm{P}\big(C_1\, n^{\xi_1} \le \hat{m} \le C_2\, n^{\xi_2}\big) \to 1, \tag{36}$$

where C_1, C_2, ξ_1, ξ_2 are strictly positive constants with

$$0 < \xi_1 \le \xi_2 < \frac{1}{5},$$

then

$$\frac{\hat{h}^*_{\hat{m}}}{h^*} - 1 = O_p\left(n^{-\min\{1/2, 1-5\xi_2, 2\xi_1(s-2)\}}\right).$$

c) Asymptotic normality. *If $s > 4 + 1/2$ and $\hat{m}$ satisfies (36) with*

$$\frac{1}{4(s-2)} < \xi_1 \le \xi_2 < \frac{1}{10},$$

then

$$\sqrt{n}\left(\frac{\hat{h}^*_{\hat{m}}}{h^*} - 1\right) \xrightarrow{d} N\big(0, \sigma^2(f)\big),$$

with

$$\sigma^2(f) = \frac{4}{25}\bigg(\frac{\mathrm{E}(f^{(4)}(X_1)^2)}{\mathrm{E}^2(f^{(4)}(X_1))} - 1\bigg).$$

The practical implementation of the proposed plug-in bandwidths depends on the data-dependent method for selecting m we consider. As in Tenreiro (2011) we will take m in such a way that f can be well approximated, in the sense of the mean integrated squared error, by the Fourier series-based estimator $\tilde{f}_m$ given at (33). For a squared integrable density function f with support contained within the interval $[0, 2\pi]$, Hart (1985) proves that the mean integrated square error of $\tilde{f}_m$ can be expressed as

$$\mathrm{MISE}(\tilde{f}_m) = \frac{1}{\pi}\bigg(H(m) + \sum_{k=1}^{\infty}\big(a_k(f)^2 + b_k(f)^2\big)\bigg),$$

where

$$H(m) = \frac{m}{n} - \frac{n+1}{n}\sum_{k=1}^{m}\big(a_k(f)^2 + b_k(f)^2\big),$$

with $H(0) = 0$. Therefore, the data-dependent method for selecting m we consider is defined by the first integer $\hat{m}_{H_\gamma}$ satisfying

$$\hat{m}_\gamma = \arg\min_{m \in \mathscr{M}_n} \hat{H}_\gamma(m), \tag{37}$$

where

$$\hat{H}_\gamma(m) = \frac{m}{n} - \gamma\frac{n+1}{n}\sum_{k=1}^{m}\hat{\hat{c}}_k,$$

and $\hat{H}_\gamma(0) = 0$, with $\mathscr{M}_n = \{L_n, L_n+1, \ldots, U_n\}$, $L_n < U_n$ are deterministic sequences of non-negative integers, $0 < \gamma \le 1$ needs to be chosen by the user, and $\hat{\hat{c}}_k$ is given by (31). The value

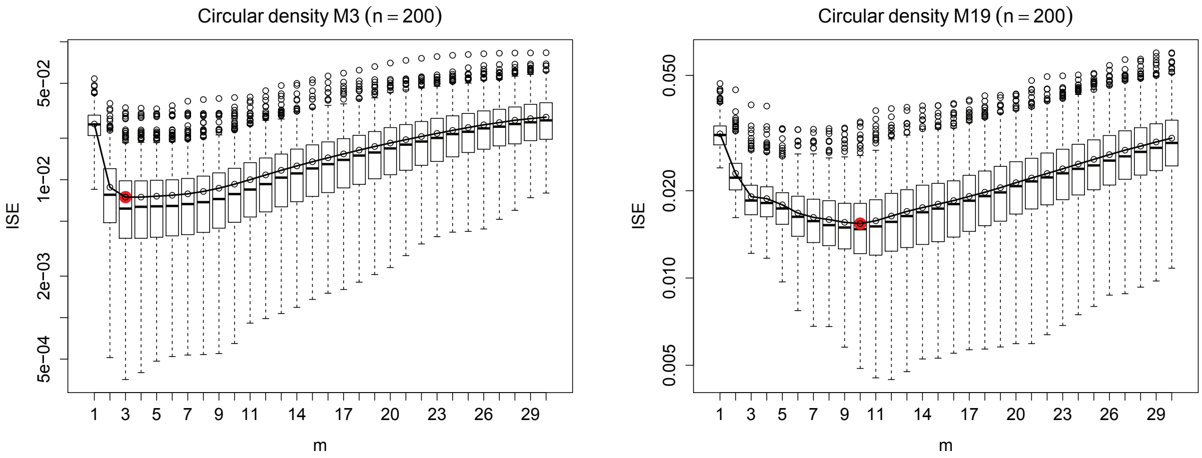

Figure 1: *Empirical distribution of* $\mathrm{ISE}(f;\tilde{f}_{\mathrm{vM}},\hat{h}^*_m,n)$ *depending on* m *for models M3 and M19* ($n=200$) *from the Oliveira et al. (2012) set of circular density models. The number of replications is* 500.

$\hat{m}_\gamma$ depends on $\mathscr{M}_n$ through the sequences L_n and U_n that need also to be chosen by the user. If they are taken equal to $L_n = \lfloor C_1 n^{\xi_1} \rfloor + 1$ and $U_n = \lfloor C_2 n^{\xi_2} \rfloor$, where $\lfloor x \rfloor$ is the integral part of x and C_1, C_2, ξ_1, ξ_2 are strictly positive constants satisfying the conditions of Theorem 7, we know that the data-dependent bandwidths $\hat{h}^*$ and $\hat{g}^*$ will possess good asymptotic properties. Assuming for ease of explanation that $s \geq 5$ in Theorem 7, we deduce that the best orders of convergence for the relative errors of each one of the bandwidths $\hat{h}^*$ and $\hat{g}^*$ will take place by choosing $\xi_1 = \xi_2 = 1/11$. In this case, since the power $n^{1/11}$ remains small for very large sample sizes, the sequences L_n and U_n are dominated by the size of the constants C_1 and C_2. If we want to deal with a wide set of distributional characteristics of the underlying density function f the sequences L_n and U_n should be chosen such that the set $\mathscr{M}_n$ contains very small and moderately large values of m.

This is illustrated in Figure 1 where we show 30 boxplots describing the empirical distribution of the integrated squared error $\mathrm{ISE}(f;\tilde{f}_{\mathrm{vM}},h,n) = \int_0^{2\pi}\{\tilde{f}_{\mathrm{vM}}(\theta;h) - f(\theta)\}^2 d\theta$, based on 500 samples from the circular densities M3 and M19 considered in Oliveira et al. (2012), where $h = \hat{h}^*_m$ for $m \in \{1,2,\ldots,30\}$. We include a polygonal line going through the sample mean values of these distributions, thus giving an approximation of $\mathrm{EISE}(m) := \mathrm{E}\big(\mathrm{ISE}(f;\tilde{f}_{\mathrm{vM}},\hat{h}^*_m,n)\big)$. The solid red circle is used to point out the optimal value of m in the sense of minimising the approximation of the EISE function. The integrals are evaluated numerically by using a grid of equally spaced 1501 points and the composite Simpson's rule. As for other densities that present simple distributional features, for the wrapped normal density M3 with mean direction $\mu = 0$ and mean resultant length $\rho = 0.9$, a small value of m seems to be the best choice. A different situation occurs for densities that present more complex distributional features. This is the case of density M19 that is a mixture of five von Mises densities with mixture proportions $\alpha = \big(4, \frac{5}{36}, \frac{5}{36}, \frac{5}{36}, \frac{5}{36}\big)$, mean directions $\mu = (2, 4, 3.5, 4, 4.5)$, and concentration parameters $\kappa = (3, 3, 50, 50, 50)$. For these densities, using a large value of m seems to be highly advisable. In the following we take $C_1 = 0.25$ and $C_2 = 25$ which leads to $L_n = 1$ and $30 \leq U_n \leq 87$ for $10 \leq n \leq 10^6$. Some simulation experiments reveal that the previous method for selecting m is quite robust against the choice of C_2 and its performance is not affected if larger values for C_2 are taken.

The inclusion of the correction parameter γ in the previous criterion function is crucial for the good performance of the method. To the best of our knowledge, a similar idea was for the first time suggested by Hart (1985) for selecting the number of terms to be used in a Fourier series-based density estimator. As the considered set $\mathscr{M}_n$ of possible values of m includes large values of m, some simulation experiments reveal that taking $\gamma = 1$, in which case $\hat{H}_\gamma(m)$ is an unbiased estimator of $H(m)$, does not prevent us from getting excessively large values of m, which leads to very poor results especially for densities whose Fourier coefficients converge quickly to zero. In fact, excessively large values of m might lead to an overestimation of the quadratic functional θ_2, and therefore to an underestimation of the asymptotic optimal bandwidths h^* or g^*. Taking into account that the function $\gamma \mapsto \hat{m}_{H_\gamma}$ is nondecreasing with probability one, we may expect to soften the above-mentioned problems by including a correction parameter strictly less than one in the considered criterion function. As suggested by this property, the simulation results support the idea that small values of γ generally improve Hart's method for distributions whose Fourier coefficients converge quickly to zero, and large values of γ are more appropriate for distributions with Fourier coefficients converging slowly to zero. In order to find a compromise between these two extreme situations, we decide to follow the suggestion of Tenreiro (2011) and taking $\gamma = 0.5$.

5 Simulation study

We present in this section the results of a simulation study carried out to analyse the finite sample behaviour of the Fourier series-based direct plug-in bandwidth selectors introduced in the previous section. However, as the results obtained by the plug-in bandwidths $\hat{h}^*_{\hat{m}}$ and $\hat{g}^*_{\hat{m}}$ defined, respectively, by (34) and (35), with $\hat{m}$ given by (37), were very similar, we will restrict our attention to the kernel estimator (4) for the first of these bandwidths. Moreover, as the estimator $\hat{\hat{\theta}}_{2,\hat{m}}$ of $\theta_2(f)$ defined by (30) may occasionally produce poor, sometimes negative, estimates of $\theta_2(f)$ when the size of the sample is small, and it performs similarly to $\hat{\tilde{\theta}}_{2,\hat{m}}$ defined by (32) when the sample size is moderate or large, the data-dependent bandwidth based on $\hat{\hat{\theta}}_{2,\hat{m}}$ is not considered hereafter. Finally, as the estimator $\tilde{f}_{\mathrm{vM}}$ given at (5) is used in the already mentioned papers of Taylor (2008), Di Marzio et al. (2009), Oliveira et al. (2012) and García-Portugués (2013), that address the automatic selection of the smoothing parameter, from now on we take for L the von Mises kernel $L(t) = e^{-t}, t \geq 0$. Therefore, as $\boldsymbol{c}(L) = (4\pi)^{-1/10}$ for this kernel, the asymptotic optimal bandwidth h^* in (29) is given by

$$h^* = (4\pi)^{-1/10} \theta_2(f)^{-1/5} n^{-1/5}, \tag{38}$$

and the Fourier series-based plug-in bandwidth we consider henceforth is defined by

$$\hat{h}_{\mathrm{FO}} := (4\pi)^{-1/10}\hat{\hat{\theta}}_{2,\hat{m}}^{-1/5} n^{-1/5} = (4\pi)^{-1/10}\theta_2(\tilde{f}_{\hat{m}})^{-1/5} n^{-1/5},$$

where $\hat{\hat{\theta}}_{2,m}$ is given by (32), $\hat{m} = \hat{m}_\gamma$ is given by (37) with L_n, U_n and γ chosen as explained in the previous section, and $\tilde{f}_m$ is the Fourier series-based estimator of f given at (33).

Some other plug-in bandwidths existing in the literature are included in this study. The simplest one, firstly considered in García-Portugués (2013, eq. 6, p. 1661), is an adaptation of the method proposed in a kernel density estimation for linear data context by Deheuvels (1977, p. 36) and Deheuvels and Hominal (1980, pp. 28–29), and made popular by Silverman (1986, pp. 45–48). The idea is to estimate $\theta_2(f)$ by making a parametric hypothesis on f. Assuming that f is a von Mises density with mean direction μ and concentration parameter κ, from (26) and (38) we can define the von Mises reference distribution bandwidth selector by

$$\hat{h}_{\mathrm{vM}} = (4\pi)^{-1/10}\left(\frac{3\hat{\kappa}^2 I_0(2\hat{\kappa}) - \hat{\kappa} I_1(2\hat{\kappa})}{8\pi I_0(\hat{\kappa})^2}\right)^{-1/5} n^{-1/5},$$

where we take for $(\hat{\mu}, \hat{\kappa})$ the maximum likelihood estimator of (μ, κ) (under the von Mises model) given by the equations

$$\frac{1}{n}\sum_{i=1}^{n}\sin(X_i - \hat{\mu}) = 0, \ \frac{1}{n}\sum_{i=1}^{n}\cos(X_i - \hat{\mu}) = \frac{I_1(\hat{\kappa})}{I_0(\hat{\kappa})}.$$

This bandwidth selector is implemented by the function bw_dir_rot of the R package 'DirStats' (García-Portugués 2020). A similar proposal leading to an alternative "von Mises-scale plug-in rule" was made by Taylor (2008, p. 3495). However, as this plug-in bandwidth is based on approximation (27) which is not asymptotically equivalent to the mean integrated squared error of $\tilde{f}_{\mathrm{vM}}$, it is not considered in what follows.

A more flexible reference distribution family is proposed in Oliveira et al. (2012). These authors assume that f is a mixture of M von Mises densities with mean directions μ_j and concentration parameters κ_j, for $j = 1, \ldots, M$, that is, f takes the form

$$f_M(\theta) = \sum_{j=1}^{M} p_i f_{\mathrm{vM}}(\theta; \mu_j, \kappa_j), \quad \sum_{j=1}^{M} p_j = 1, \quad p_j \geq 0.$$

For each one of the considered mixtures the associated $3M$ parameters of the model are estimated by using maximum likelihood estimation via an EM algorithm performed with the function movMF of the R package 'movMF' (Hornik and Grün 2014), and the selection of the number of mixture components is performed by using the Akaike Information Criterion (AIC) (see Oliveira et al. 2012, p. 3900). The AIC is computed for mixtures of $M = 2, 3, 4, 5$ von Mises distributions and the selected number of mixtures $\hat{M}$ for the reference distribution is the one minimising the AIC. As described in Oliveira et al. (2012, p. 3906) some computational problems may arrive in practice in the implementation of the EM algorithm. When no result can be obtained for the different values of M, one takes $\hat{M} = 1$ in which case the bandwidth selected is the von Mises reference distribution bandwidth as the distribution parameters are estimated via maximum likelihood. Denoting by $f_{\hat{M}}$ the density selected from the considered reference density family where the parameters are estimated and $\hat{M}$ is obtained from the sample, we consider in our study the bandwidth selector given by

$$\hat{h}_{\mathrm{OLI}} = (4\pi)^{-1/10}\theta_2(f_{\hat{M}})^{-1/5}n^{-1/5}.$$

This is not the bandwidth selector originally proposed in Oliveira et al. (2012, p. 3900), which is defined as the minimiser over $h > 0$ of the mean integrated squared error approximation given at (23) after replacing $\theta_2(f)$ by $\theta_2(f_{\hat{M}})$, implemented by the function (bw.pi)$^{-1/2}$ of the R package 'NPCirc' (Oliveira, Crujeiras, and Rodríguez-Casal 2015). However, as the two selectors performed similarly, we decided to show here only the results of the selector based on the asymptotic optimal bandwidth.

An alternative procedure for selecting the number M of mixture components is proposed in García-Portugués (2013, Section 4.1). For each one of the considered mixtures the parameters of the model are estimated as in Oliveira et al. (2012) but the selection of M is performed by using a different strategy and the Bayesian Information Criteriun (BIC). The details on the considered procedure are given in García-Portugués (2013, Algorithm 3, p. 1666). Denoting by $f_{\tilde{M}}$ the density selected from the considered reference density family, we include in our study the AMI and EMI selectors proposed in García-Portugués (2013, Algorithm 1, p. 1664, and Algorithm 2, p. 1665). They are defined by

$$\hat{h}_{\mathrm{AMI}} = (4\pi)^{-1/10}\theta_2(f_{\tilde{M}})^{-1/5}n^{-1/5}$$

and

$$\hat{h}_{\mathrm{EMI}} = \underset{h>0}{\operatorname{argmin}}\, \mathrm{MISE}(f_{\tilde{M}}; \tilde{f}_{\mathrm{vM}}, h, n),$$

where the exact mean integrated squared error is evaluated by using the closed expression for $\mathrm{MISE}(f; \tilde{f}_{\mathrm{vM}}, h, n)$ derived in García-Portugués et al. (2013, Proposition 4, p. 159) when f is a mixture of von Mises densities. These bandwidth selectors are implemented by the functions bw_dir_ami and bw_dir_emi of the R package 'DirStats' (García-Portugués 2020).

Two other data-driven procedures for selecting the bandwidth proposed by Hall et al. (1987) are also included in our study. They are the least-square cross-validation and the Kullback-Leibler or likelihood cross-validation methods. Denoting by $\tilde{f}_{\mathrm{vM},-i}$ the kernel density estimator (5) by leaving out the i-th observation, the least-square cross-validation bandwidth $\hat{h}_{\mathrm{LSCV}}$ is obtained by minimising the classic least-square cross-validation criterion function given by $\mathrm{LSCV}(h) = \int_0^{2\pi} \tilde{f}_{\mathrm{vM}}(\theta; h)d\theta - 2n^{-1}\sum_{i=1}^n \tilde{f}_{\mathrm{vM},-i}(X_i; h)$, whereas the likelihood cross-validation bandwidth $\hat{h}_{\mathrm{LCV}}$ is obtained by maximising the likelihood cross-validation criterion function defined by $\mathrm{LCV}(h) = \prod_{i=1}^n \tilde{f}_{\mathrm{vM},-i}(X_i; h)$. These methods are implemented by the function (bw.CV)$^{-1/2}$ (methods "LSCV" and "LCV") from the R package 'NPCirc' (Oliveira et al. 2015).

The set of 20 circular distributions considered in Oliveira et al. (2012), which includes the von Mises distribution, the cardioid distribution, various wrapped distributions and mixtures of them, is used to analyse the effectiveness of the proposed plug-in bandwidth $\hat{h}_{\mathrm{FO}}$ and to compare it with the data-driven bandwidths $\hat{h}_{\mathrm{vM}}$, $\hat{h}_{\mathrm{OLI}}$, $\hat{h}_{\mathrm{AMI}}$, $\hat{h}_{\mathrm{EMI}}$, $\hat{h}_{\mathrm{LSCV}}$ and $\hat{h}_{\mathrm{LCV}}$. This set of densities is very rich, containing densities with a wide variety of distribution features such as multimodality, skewness and/or peakedness. For a careful description of the different models and the plots of the corresponding circular densities see Oliveira et al. (2012, pp. 3901, 3902, 3907). Although we have used self-programmed code written in R and functions from the 'circular' package in R (Lund and Agostinelli 2017) for generating data from the previous models, this can also be done by using the function rcircmix from the above-mentioned 'NPCirc' package.

For different sample sizes and for each one of the 20 test distributions the quality of each one of the considered bandwidths $\hat{h}$ is analysed through the measure of stochastic performance defined by

$$L_2\text{–norm of } \mathrm{ISE}(f;\tilde{f}_{\mathrm{vM}},\hat{h},n) = \sqrt{\mathrm{Var}(\mathrm{ISE}(f;\tilde{f}_{\mathrm{vM}},\hat{h},n)) + \mathrm{E}^2(\mathrm{ISE}(f;\tilde{f}_{\mathrm{vM}},\hat{h},n))}.$$

As expressed by the last equality, this performance measure takes into account not only the mean of the $\mathrm{ISE}(f;\tilde{f}_{\mathrm{vM}},\hat{h},n)$ distribution, but also its variability. As the least-square cross-validation bandwidth showed an inferior global performance compared to the likelihood cross-validation bandwidth, only the results obtained by the bandwidths $\hat{h}_{\mathrm{FO}}$, $\hat{h}_{\mathrm{vM}}$, $\hat{h}_{\mathrm{OLI}}$, $\hat{h}_{\mathrm{AMI}}$, $\hat{h}_{\mathrm{EMI}}$ and $\hat{h}_{\mathrm{LCV}}$ are reported in Figures 2, 3, 4 and 5. In these figures the empirical L_2–norm of $\mathrm{ISE}(f;\tilde{f}_{\mathrm{vM}},\hat{h},n)$, based on 500 replications, is shown for sample sizes $n = 25 \cdot 2^k$, $k = 0, 1, \ldots, 7$. As before, the integrals $\int_0^{2\pi}\{\tilde{f}_{\mathrm{vM}}(\theta;\hat{h}) - f(\theta)\}^2 d\theta$ are evaluated by using a grid of equally spaced 1501 points and the composite Simpson's rule.

As we can see from the graphics, the bandwidth $\hat{h}_{\mathrm{vM}}$ is suitable when the underlying density has a distributional structure that is close to a von Mises distribution. This situation occurs with models 1, 2, 3, 4 and 9. However, its performance is very poor for circular distributions that present more complex features. Some extreme situations where this poor behaviour is observed for all sample sizes are models 7, 11, 13, 14, 16 and 20. In all these cases the sampling distribution of the considered concentration parameter estimator $\hat{\kappa}$ is distributed around zero leading to large bandwidths that provide uniform estimates for the underlying circular density (on this situation, see Oliveira et al. 2012, p. 3906). With respect to the bandwidth $\hat{h}_{\mathrm{OLI}}$, we can see that it shows a poor behaviour for almost all the considered scenarios when the sample size is small. This behaviour can be improved if we take $\hat{M} = 2$ for small samples, which is in line with the observation of Oliveira et al. (2012, p. 3903) that consider such a choice a fair one when the sample size is small. This strategy produced better results (see the online supplementary material) for almost all the considered models when $n \leq 200$, with the exception of models 11, 14, 16, 17, 18 and 20, where worse results were observed for at least one of the sample sizes $n = 100, 200$. The performance of $\hat{h}_{\mathrm{OLI}}$ improves significantly for moderate or large sample sizes, where it is among the best of the considered bandwidth selectors. The bandwidths $\hat{h}_{\mathrm{AMI}}$ and $\hat{h}_{\mathrm{EMI}}$ show a

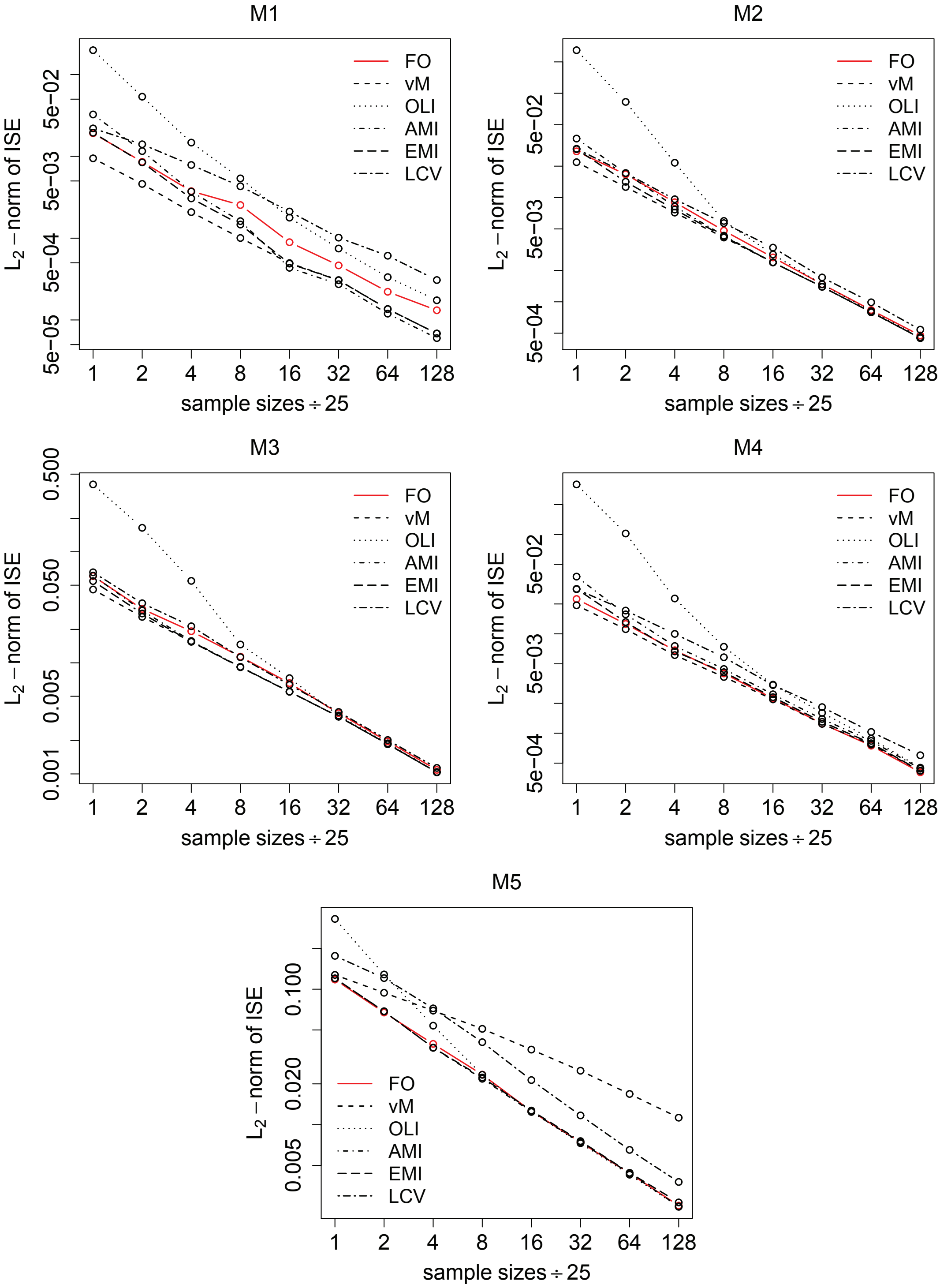

Figure 2: *Empirical L_2–norm of* $\mathrm{ISE}(f; \tilde{f}_{\mathrm{vM}}, \hat{h}, n)$ *(log scale) for the bandwidths* $\hat{h}_{\mathrm{FO}}$, $\hat{h}_{\mathrm{vM}}$, $\hat{h}_{\mathrm{OLI}}$, $\hat{h}_{\mathrm{AMI}}$, $\hat{h}_{\mathrm{EMI}}$ *and* $\hat{h}_{\mathrm{LCV}}$, *and circular density models* 1 *to* 5. *The number of replications is* 500.

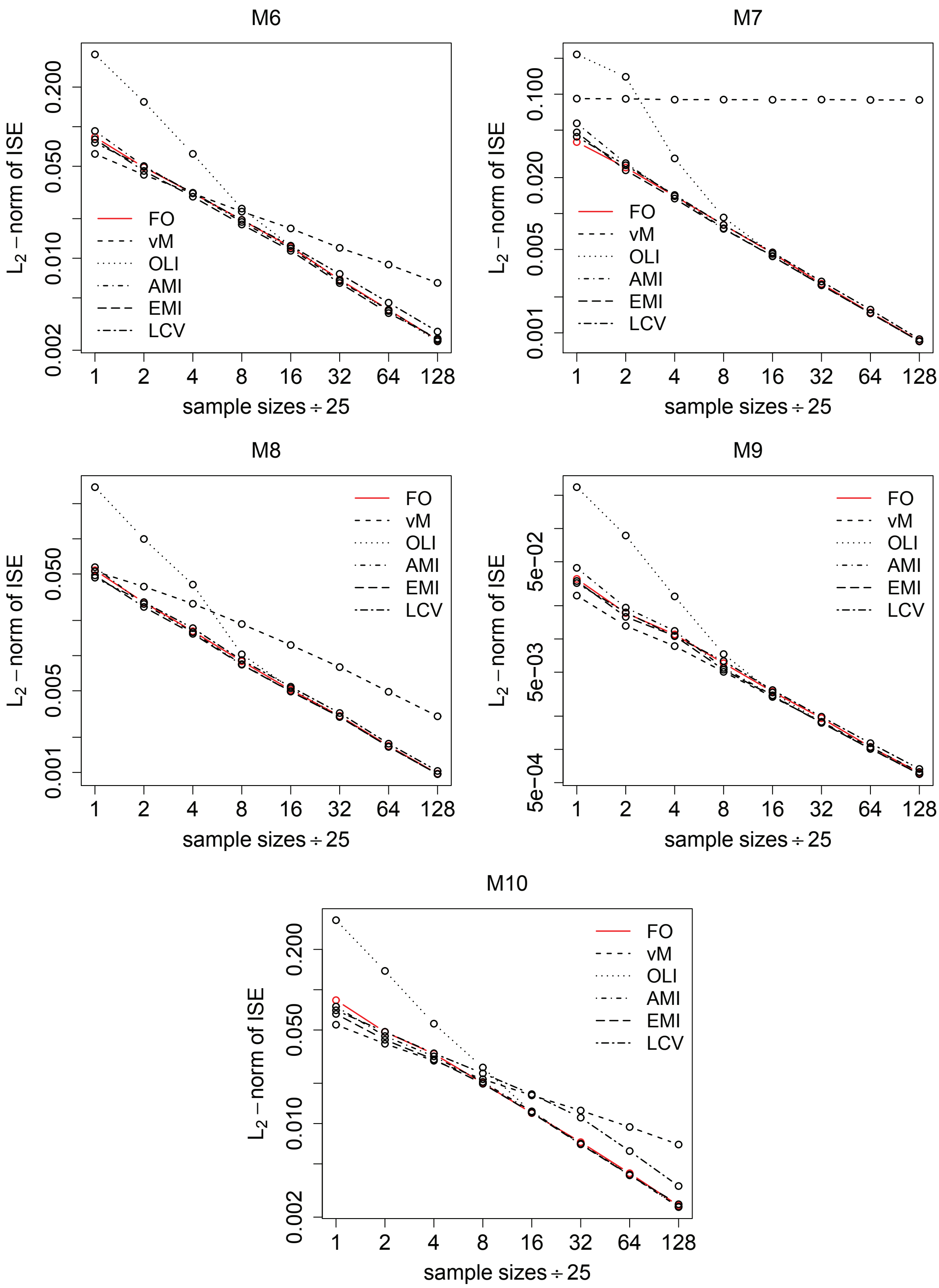

Figure 3: *Empirical L_2–norm of* $\mathrm{ISE}(f; \tilde{f}_{\mathrm{vM}}, \hat{h}, n)$ *(log scale) for the bandwidths* $\hat{h}_{\mathrm{FO}}$, $\hat{h}_{\mathrm{vM}}$, $\hat{h}_{\mathrm{OLI}}$, $\hat{h}_{\mathrm{AMI}}$, $\hat{h}_{\mathrm{EMI}}$ *and* $\hat{h}_{\mathrm{LCV}}$*, and circular density models* 6 *to* 10*. The number of replications is* 500.

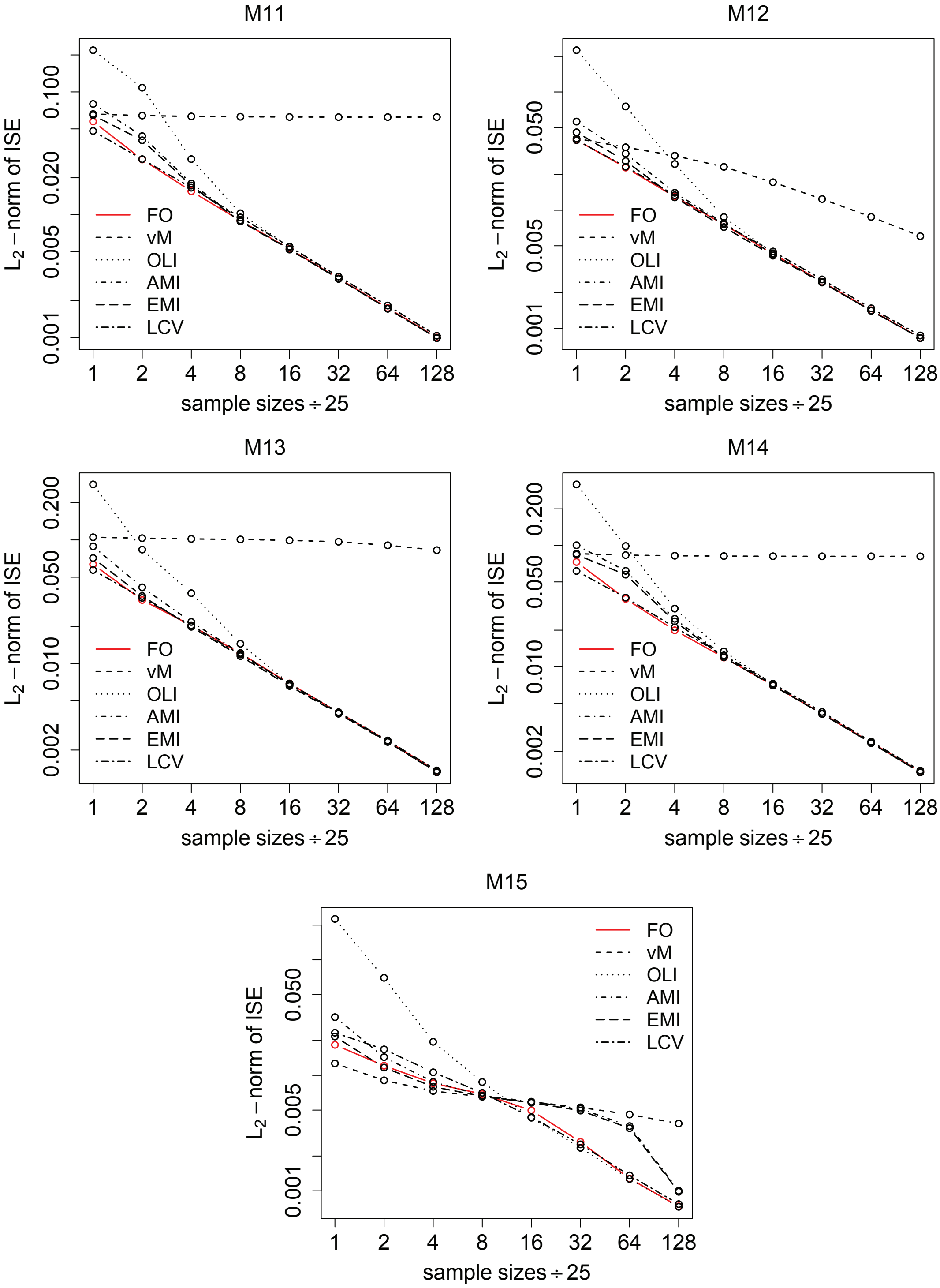

Figure 4: *Empirical* L_2*–norm of* $\mathrm{ISE}(f; \tilde{f}_{\mathrm{vM}}, \hat{h}, n)$ *(log scale) for the bandwidths* $\hat{h}_{\mathrm{FO}}$, $\hat{h}_{\mathrm{vM}}$, $\hat{h}_{\mathrm{OLI}}$, $\hat{h}_{\mathrm{AMI}}$, $\hat{h}_{\mathrm{EMI}}$ *and* $\hat{h}_{\mathrm{LCV}}$*, and circular density models* 11 *to* 15*. The number of replications is* 500*.*

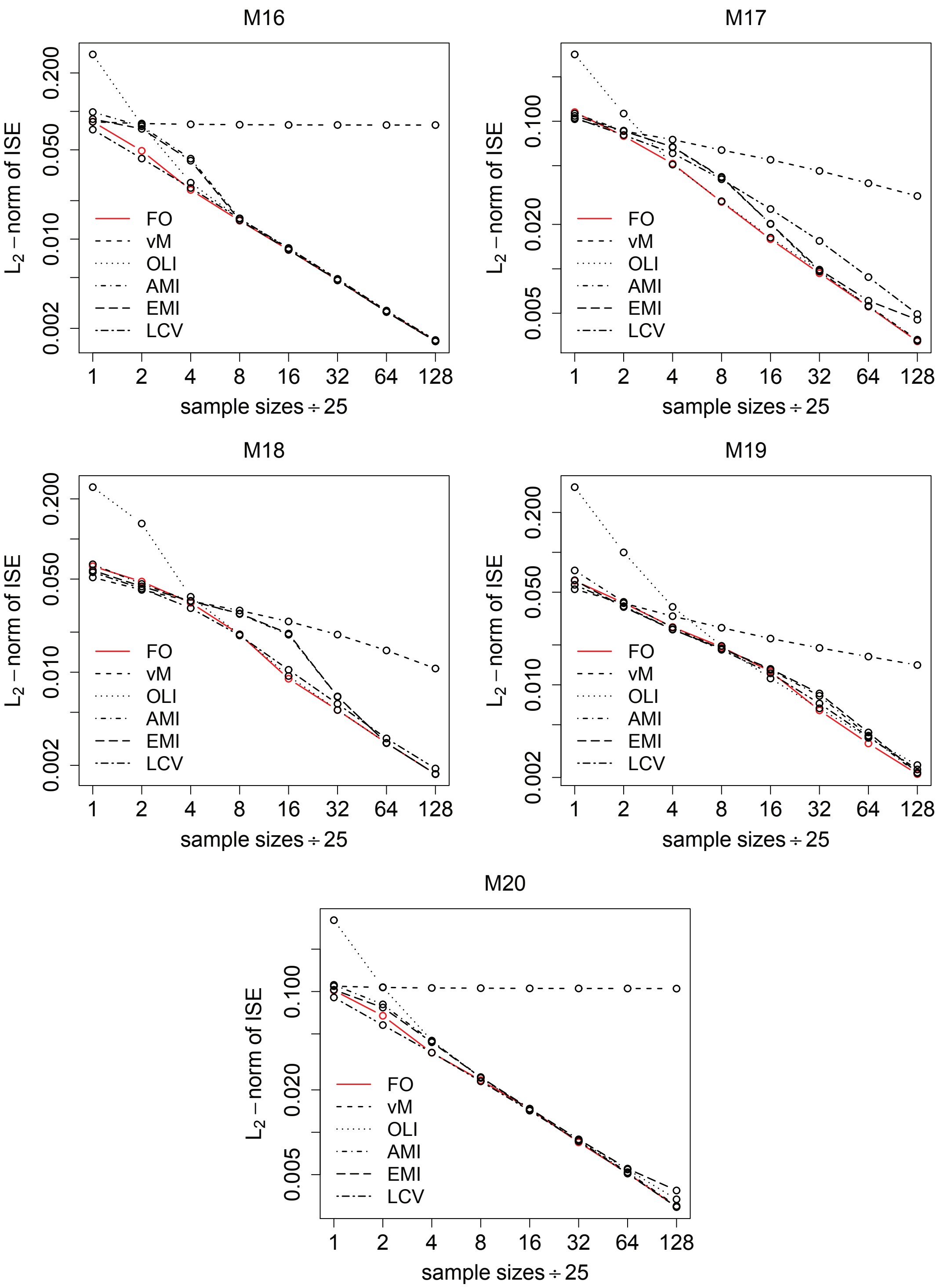

Figure 5: *Empirical L_2–norm of* $\mathrm{ISE}(f; \tilde{f}_{\mathrm{vM}}, \hat{h}, n)$ *(log scale) for the bandwidths* $\hat{h}_{\mathrm{FO}}$*,* $\hat{h}_{\mathrm{vM}}$*,* $\hat{h}_{\mathrm{OLI}}$*,* $\hat{h}_{\mathrm{AMI}}$*,* $\hat{h}_{\mathrm{EMI}}$ *and* $\hat{h}_{\mathrm{LCV}}$*, and circular density models* 16 *to* 20*. The number of replications is* 500*.*

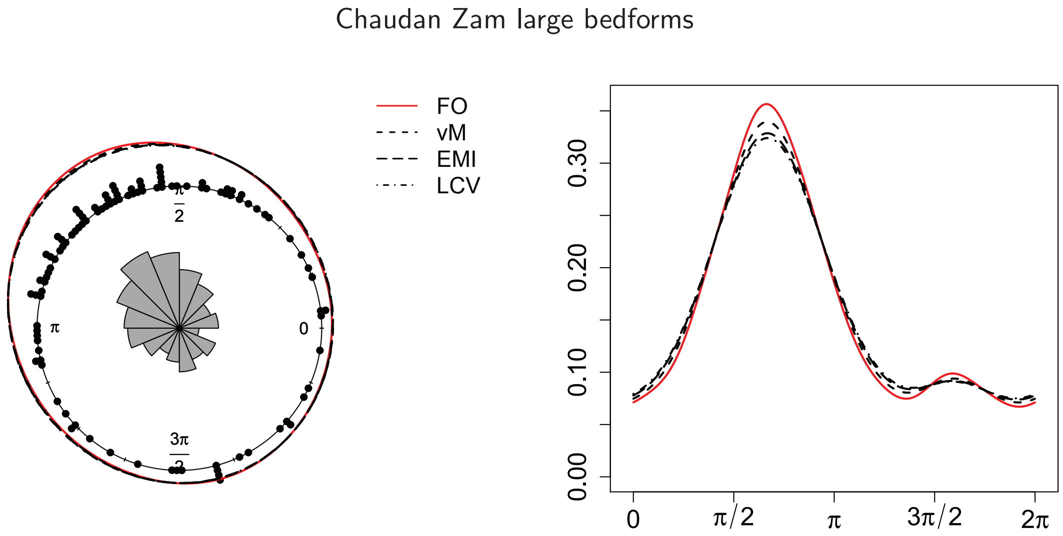

Figure 6: *Estimates for the cross-bed density by using the kernel estimator* $\tilde{f}_{\mathrm{vM}}(\cdot;\hat{h})$ *and the bandwidth selectors* $\hat{h}_{\mathrm{FO}}$*,* $\hat{h}_{\mathrm{vM}}$*,* $\hat{h}_{\mathrm{EMI}}$ *and* $\hat{h}_{\mathrm{LCV}}$*.*

very good global behaviour with the exception of models 15, 17, 18 and 19 where for some of the considered sample sizes they present and inferior performance with respect to other bandwidth selectors. With the exceptions of models 5, 10 and 17, whose densities present a single or several strong peaks, the likelihood cross-validation bandwidth $\hat{h}_{\mathrm{LCV}}$ shows a very good behaviour for all the test densities and sample sizes. However, a better global behaviour is shown by the Fourier series-based plug-in bandwidth $\hat{h}_{\mathrm{FO}}$. This bandwidth is quite competitive against the von Mises reference distribution bandwidth selector for simple distribution models, the exception being model 1, and, at the same time, presents a good performance for all the considered circular density models and sample sizes. It is the best or is among the best of the considered bandwidth selectors for all the considered models and sample sizes. This analysis suggests that $\hat{h}_{\mathrm{FO}}$ is always a good choice for selecting the bandwidth in kernel density estimation for circular data.

6 Two real-data examples

In this section we consider two real-data sets analysed in Oliveira et al. (2012) and available through the R package 'NPCirc' (Oliveira et al. 2015). For each of them, the data-driven bandwidth selectors considered in the previous section, namely $\hat{h}_{\mathrm{FO}}$, $\hat{h}_{\mathrm{vM}}$, $\hat{h}_{\mathrm{OLI}}$, $\hat{h}_{\mathrm{AMI}}$, $\hat{h}_{\mathrm{EMI}}$ and $\hat{h}_{\mathrm{LCV}}$, are used.

The first data set consists of 104 cross-bed measurements from the Himalayan molasse in Pakistan presented in Fisher (1993, Measurements of Chaudan Zam large bedforms, pp. 250–251). The smoothing parameter selectors $\hat{h}_{\mathrm{FO}} = 0.370$ and $\hat{h}_{\mathrm{OLI}} = 0.362$ ($\hat{M} = 2$) yield identical bandwidths, while larger bandwidths are produced by $\hat{h}_{\mathrm{vM}} = 0.442$, $\hat{h}_{\mathrm{AMI}} = 0.484$, $\hat{h}_{\mathrm{EMI}} = 0.488$ ($\tilde{M} = 1$) and $\hat{h}_{\mathrm{LCV}} = 0.507$. As we can see from Figure 6 the different smoothing parameters

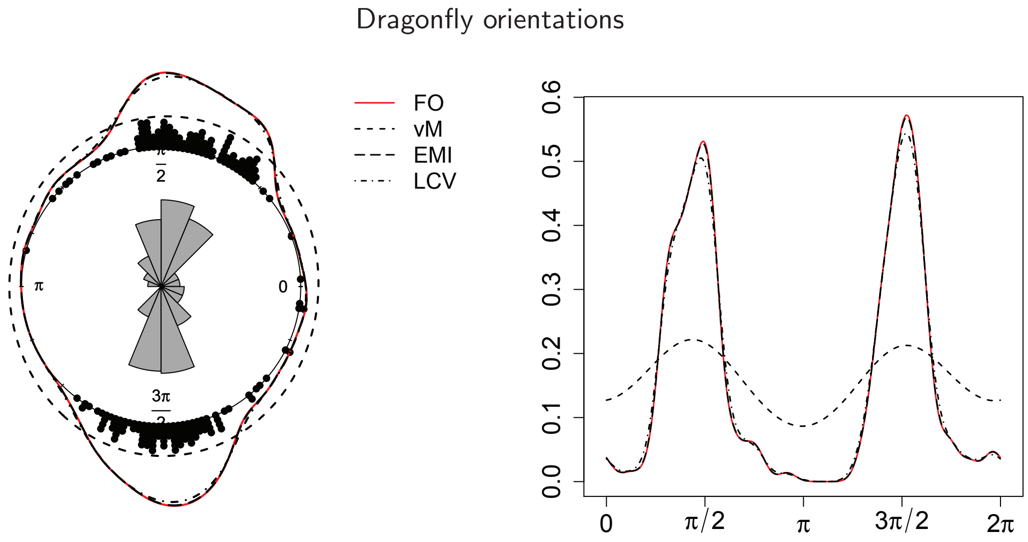

Figure 7: *Estimates for the dragonflies orientation density by using the kernel estimator* $\tilde{f}_{\mathrm{vM}}(\cdot;\hat{h})$ *and the bandwidth selectors* $\hat{h}_{\mathrm{FO}}$, $\hat{h}_{\mathrm{vM}}$, $\hat{h}_{\mathrm{EMI}}$ *and* $\hat{h}_{\mathrm{LCV}}$.

provide similar density estimates (only the estimates for the bandwidths $\hat{h}_{\mathrm{FO}}$, $\hat{h}_{\mathrm{vM}}$, $\hat{h}_{\mathrm{EMI}}$ and $\hat{h}_{\mathrm{LCV}}$ are displayed). Other than the main mode distribution, the linear plot seems to reveal the presence of a second less important mode distribution in an opposite direction to the main one.

The second data set, presented in Batschelet (1981, pp. 23–24), consists of the orientation of 214 dragonflies with respect to the azimuth of the sun. As most dragonflies have chosen a direction of approximately 90° either to the right or to the left of the sun's rays, the underlying circular density should be bimodal. For this data set the bandwidth selectors $\hat{h}_{\mathrm{FO}} = 0.136$, $\hat{h}_{\mathrm{OLI}} = 0.126$ ($\hat{M} = 4$), $\hat{h}_{\mathrm{AMI}} = 0.130$ and $\hat{h}_{\mathrm{EMI}} = 0.140$ ($\tilde{M} = 3$) yield identical bandwidths. Although a slightly larger bandwidth is produced by the likelihood cross-validation selector $\hat{h}_{\mathrm{LCV}} = 0.168$, the different kernel density estimates are similar, revealing a clear bimodal circular distribution as shown in Figure 7. A bimodal distribution structure is also revealed by the larger bandwidth produced by the von Mises reference distribution bandwidth selector $\hat{h}_{\mathrm{vM}} = 0.778$. However, based on the simulation results obtained in the previous section for the circular density with opposite modes M7 (see Figure 3), we expect a poor behaviour for this last bandwidth selector.

7 Conclusions

The asymptotic expansions for the mean integrated squared error of a general class of delta sequence estimators for circular data presented in this paper improve and correct similar ones existing in the literature, and enabled us to derive, in a unified way, explicit expressions for the asymptotic optimal bandwidths of kernel estimators $\tilde{f}_L$ and $\check{f}_K$ given at (4) and (6), respectively. Based on these expressions a Fourier series-based direct plug-in approach for bandwidth selection is proposed. The theoretical properties established for the new bandwidth selector method, not shared by other existing plug-in methods, but principally because of the very good finite sample performance it possesses, provides very strong evidence that it might present a good overall behaviour for a wide range of circular density features.

8 Proofs

As mentioned at the beginning of Section 2, we denote by f not only the probability density function of the observed circular random variables, but also its periodic extension to the real line with period 2π. In this section all the limits are understood to be taken as $n \to +\infty$.

Proof of Theorem 1: From the periodicity of δ_n and f we have

$$\mathrm{E}\hat{f}_n(\theta) = \int_0^{2\pi} \delta_n(\theta - x) f(x) dx = \int_{-\pi}^{\pi} \delta_n(y) f(\theta - y) dy, \tag{39}$$

for $\theta \in [0, 2\pi[$. The uniform convergence stated in a) follows now from standard arguments as f is uniformly continuous on $\mathbb{R}$ and the sequence (δ_n) satisfies conditions $(\Delta.1)$ and $(\Delta.2)$ (see Watson and Leadbetter 1964, Proof of Lemma 3, p. 104). Similar arguments can be used to establish b). For that we start by noting that $\alpha(\delta_n) \to +\infty$ as the sequence (δ_n) satisfies condition $(\Delta.3)$ (cf. Watson and Leadbetter 1964, Lemma 1, p. 103). In order to conclude, it suffices to use the equality

$$n\alpha(\delta_n)^{-1} \,\mathrm{Var}\hat{f}_n(\theta) = \int_{-\pi}^{\pi} \varphi_n(y) f(\theta - y) dy - \alpha(\delta_n)^{-1} \big(\mathrm{E}\hat{f}_n(\theta)\big)^2,$$

where $\theta \in [0, 2\pi[$ and $\varphi_n(y) = \delta_n(y)^2 / \int_{-\pi}^{\pi} \delta_n(y)^2 dy$, and the fact that (φ_n) also satisfies conditions $(\Delta.1)$ and $(\Delta.2)$. ■

Proof of Theorem 2: We start by proving that $\int_{-\pi}^{\pi} |y|^\beta \delta_n(y) dy \to 0$, for all $\beta > 0$. In fact, from assumptions $(\Delta.1)$ and $(\Delta.2)$, for any $0 < \lambda < \pi$ and n large enough, we have

$$\int_{-\pi}^{\pi} |y|^\beta \delta_n(y) dy \le \lambda^\beta + 2\pi^{\beta+1} \sup_{\lambda < |y| \le \pi} \delta_n(y).$$

Using again assumption $(\Delta.2)$, we get the stated convergence. Therefore, for $\beta = 2$ we get $\beta(\delta_n) \to 0$. Using now assumptions $(\Delta.1)$ and $(\Delta.4)$, and the fact that f'' satisfies Lipschitz condition (15), from classic arguments we get

$$\sup_{\theta \in [0, 2\pi[} \big|\mathrm{E}\hat{f}_n(\theta) - f(\theta) - \frac{1}{2}\beta(\delta_n) f''(\theta)\big| \le \frac{C}{2} \int_{-\pi}^{\pi} |y|^{2+\alpha} \delta_n(y) dy.$$

The stated result follows now from assumption $(\Delta.5)$. ■

Proof of Theorem 3: Expansion (20) follows easily from (18) and (19). Moreover, taking into account that $y = \sin y + \frac{1}{2} y^3 \int_0^1 (1-t)^2 \cos(ty) dt$, from assumption $(\Delta.5)$ we have

$$\beta(\delta_n) = \int_{-\pi}^{\pi} (\sin y)^2 \delta_n(y) dy + o(\beta(\delta_n)).$$

Finally, from representation (2) we get

$$\beta(\delta_n) = \frac{1}{2}\big(1 - a_2(\delta_n)\big) + o\big(\beta(\delta_n)\big),$$

which, together with (20), leads to the alternative expansion (21) for mean integrated squared error of the delta sequence estimator. ■

Proof of Theorem 4: As f is squared integrable in $[0, 2\pi]$, we know that f admits the $L_2([0, 2\pi])$ representation $f(\theta) = \frac{1}{2\pi}\big(1 + 2\sum_{k=1}^{\infty}\{a_k(f)\cos(k\theta) + b_k(f)\sin(k\theta)\}\big)$, where $a_k(f)$ and $b_k(f)$ are given by $a_k(f) = \int_0^{2\pi} f(\theta)\cos(k\theta)d\theta$ and $b_k(f) = \int_0^{2\pi} f(\theta)\sin(k\theta)d\theta$, for $k \in \mathbb{N}$ (Butzer and Nessel 1971, Proposition 4.2.3). Therefore, from (39) and representation (2) we get the following expression for the integrated squared bias of the delta sequence estimator based on any symmetric delta function sequence satisfying (Δ.1) and (Δ.3):

$$\mathrm{ISB}(f; \hat{f}_n, n) = \frac{1}{\pi}\sum_{k=1}^{\infty}(1 - a_k(\delta_n))^2\big(a_k(f)^2 + b_k(f)^2\big).$$

Moreover, as f is absolutely continuous on $[0, 2\pi]$ and f' is square integrable on $[0, 2\pi]$ (these are in fact the minimal conditions on f under which Theorem 4 is valid), we can use Propositions 4.1.8 and 4.2.2 of Butzer and Nessel (1971, pp. 172, 175) to conclude that $\pi^{-1}\sum_{k=1}^{\infty} k^2\big(a_k(f)^2 + b_k(f)^2\big) = \theta_1(f) < \infty$. Therefore, as for the wrapped Cauchy kernel we have $\frac{1-a_k(\delta_n)}{1-\rho} = \frac{1-\rho^k}{1-\rho} \to k$, from the dominated convergence theorem we deduce that

$$\mathrm{ISB}(f; \hat{f}_{\mathrm{wC}}, n) = (1-\rho)^2\pi^{-1}\sum_{k=1}^{\infty} k^2\big(a_k(f)^2 + b_k(f)^2\big)(1 + o(1)) = (1-\rho)^2\theta_1(f)(1 + o(1)),$$

which concludes the proof. ■

Proof of Theorems 5 and 6: From (10) and (13) we have $\alpha(\delta_n) = h^{-1}\boldsymbol{c}_1(L)$ and $\beta(\delta_n) = 2h^2\boldsymbol{c}_2(L)^{1/2}$, for estimator $\tilde{f}_L$, and from (11) and (14) we have $\alpha(\delta_n) = g^{-1}\boldsymbol{d}_1(K)(1 + o(1))$ and $\beta(\delta_n) = 2g^2\boldsymbol{d}_2(K)^{1/2}(1+o(1))$, for estimator $\check{f}_K$. The stated expansions for $\mathrm{MISE}(f; \tilde{f}_L, h, n)$ and $\mathrm{MISE}(f; \check{f}_K, g, n)$ follow now straightforwardly from expansion (20) of Theorem 3. ■

Proof of Theorem 7: When $\hat{\theta}_{2,m}$ is the estimator $\hat{\hat{\theta}}_{2,\hat{m}}$ defined at (30), from (29) and (34) we deduce that

$$\frac{\hat{h}^*_{\hat{m}}}{h^*} - 1 = -\frac{1}{5}\tilde{\theta}_2^{-6/5}\theta_2(f)^{1/5}\big(\hat{\hat{\theta}}_{2,\hat{m}} - \theta_2(f)\big),$$

for some random value $\tilde{\theta}_2$ between $\hat{\hat{\theta}}_{2,\hat{m}}$ and $\theta_2(f)$. Therefore, the stated asymptotic behaviour of the relative error of $\hat{h}^*_{\hat{m}}$ follows straightforwardly from the asymptotic behaviour of $\hat{\hat{\theta}}_{2,\hat{m}}$ that is established in Lemma 1 of Tenreiro (2011, pp. 543–544). A similar reasoning also applies when $\hat{\theta}_{2,m}$ is the estimator $\hat{\hat{\theta}}_{2,\hat{m}}$ defined at (32) as $\hat{\hat{\theta}}_{2,m} = (1-n^{-1})\hat{\hat{\theta}}_{2,m} + (n\pi)^{-1}\sum_{k=1}^{m} k^4$. This completes the proof. ■

Supplementary information

The online version contains supplementary material available at
https://doi.org/10.1080/10485252.2022.2057974

Acknowledgments

The author would like to thank the anonymous reviewers and associate editor for their constructive comments and suggestions that greatly helped to improved this work.

Funding

This work was partially supported by the Centre for Mathematics of the University of Coimbra - UIDB/00324/2020, funded by the Portuguese Government through FCT/MCTES.

References

Bai, Z.D., Rao, C.R., and Zhao, L.C. (1988), 'Kernel Estimators of Density Function of Directional Data', *Journal of Multivariate Analysis*, 27, 24–39.

Batschelet, E. (1981), *Circular Statistics in Biology*, London: Academic Press.

Beran, R. (1979), 'Exponential Models for Directional Data', *The Annals of Statistics*, 7, 1162–1178.

Bosq, D., and Lecoutre, J.-P. (1987), *Théorie de l'Estimation Fonctionnelle*, Paris: Economica.

Butzer, P.L., and Nessel, R.J. (1971), *Fourier Analysis and Approximation. One-Dimensional Theory*, Basel: Birkhauser.

Chiu, S.-T. (1991), 'Bandwidth Selection for Kernel Density Estimation', *The Annals of Statistics*, 19, 1883–1905.

Deheuvels, P. (1977), 'Estimation Non Paramétrique de la Densité par Histogrammes Généralisés', *Revue de Statistique Appliquée*, 25 (3), 5–42.

Deheuvels, P., and Hominal, P. (1980), 'Estimation Automatique de la Densité', *Revue de Statistique Appliquée*, 28 (1), 25–55.

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

Di Marzio, M, Panzera, A. and Taylor, C.C. (2011), 'Kernel Density Estimation on the Torus', *Journal of Statistical Planning and Inference*, 141, 2156–2173.

Epanechnikov, V.A. (1969), 'Nonparametric Estimation of a Multivariate Probability Density', *Theory of Probability and Its Applications*, 14, 153–158.

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', *Electronic Journal of Statistics*, 7, 1655–1685.

García-Portugués, E. (2020), 'DirStats: Nonparametric Methods for Directional Data', R package version 0.1.6. http://CRAN.R-project.org/package=DirStats.

García-Portugués, E., Crujeiras, R.M., and González-Manteiga, W. (2013), 'Kernel Density Estimation for Directional-Linear Data', *Journal of Multivariate Analysis*, 121, 152–175.

Hall, P., Watson, G.S., and Cabrera, J. (1987), 'Kernel density Estimation with Spherical Data', *Biometrika*, 74, 751–762.

Hart, J.D. (1985), 'On the Choice of a Truncation Point in Fourier Series Density Estimation', *Journal of Statistical Computation and Simulation*, 21, 95–116.

Hornik, K., and Grün, B. (2014), 'movMF: An R Package for Fitting Mixtures of von Mises-Fisher Distributions', *Journal of Statistical Software*, 58 (10), 1–31.

Klemelä, J. (2000), 'Estimation of Densities and Derivatives of Densities with Directional Data', *Journal of Multivariate Analysis*, 73, 18–40.

Kronmal, R., and Tarter, M. (1968), 'The Estimation of Probability Densities and Cumulatives by Fourier Series Methods', *Journal of the American Statistical Association*, 63, 925–952.

Laurent, B. (1997), 'Estimation of Integral Functionals of a Density and its Derivatives', *Bernoulli*, 3, 181–211.

Lund, U., and Agostinelli, C. (2017), 'circular: Circular Statistics', R package version 0.4-93. http://CRAN.R-project.org/package=circular.

Mardia, K.V., and Jupp, P.E. (2000), *Directional Statistics*, Chischester: John Wiley.

Nadaraya, E.A. (1974), 'On the Integral Mean Square Error of Some Nonparametric Estimates for the Density Function', *Theory of Probability and Its Applications*, 19, 133–141.

Oliveira, M., Crujeiras, R.M., and Rodríguez-Casal, A. (2012), 'A Plug-in Rule for Bandwidth Selection fn Circular Density Estimation', *Computational Statistics and Data Analysis*, 56, 3898–3908.

Oliveira, M., Crujeiras, R.M., and Rodríguez-Casal, A. (2015), 'NPCirc: Nonparametric Circular Methods', R package version 2.0.1. http://CRAN.R-project.org/package=NPCirc.

Parzen, E. (1962), 'On Estimation of a Probability Density Function and Mode', *The Annals of Mathematical Statistics*, 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', *The Annals of Mathematical Statistics*, 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', *Computational Statistics and Data Analysis*, 52, 3493–3500.

Tenreiro, C. (2011), 'Fourier Series Based Direct Plug-in Bandwidth Selectors for Kernel Density Estimation', *Journal of Nonparametric Statistics*, 23, 533–545.

Tenreiro, C. (2020), 'Bandwidth Selection for Kernel Density Estimation: A Hermite Series-Based Direct Plug-in Approach', *Journal of Statistical Computation and Simulation*, 90, 3433–3453.

Tsuruta, Y., and Sagae, M. (2017a), 'Asymptotic Property of Wrapped Cauchy Kernel Density Estimation on the Circle', *Bulletin of Informatics and Cybernetics*, 49, 1–10.

Tsuruta, Y., and Sagae, M. (2017b), 'Higher Order Kernel Density Estimation on the Circle', *Statistics & Probability Letters*, 131, 46–50.

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

Watson, G.S., and Leadbetter, M.R. (1964), 'Hazard Analysis II', *Sankhyā: The Indian Journal of Statistics*, Ser. A, 26, 101–116.

Woodroofe, M. (1970), 'On Choosing a Delta-Sequence', *The Annals of Mathematical Statistics*, 41, 1665–1671.

Wu, T.-J. (1995), 'Adaptive Root n Estimates of Integrated Squared Density Derivatives', *The Annals of Statistics*, 23, 1474–1495.

Supplement to "Kernel density estimation for circular data: a Fourier series-based plug-in approach for bandwidth selection"*

Carlos Tenreiro[†]
CMUC, Department of Mathematics, University of Coimbra, Coimbra, Portugal

Abstract

This file contains the online supplementary material mentioned in Section 5 of the paper.

*Supplementary material available at https://doi.org/10.1080/10485252.2022.2057974

[†]ORCID ID: 0000-0002-5495-6644. E-mail: tenreiro@mat.uc.pt URL: http://www.mat.uc.pt/~tenreiro/ Postal address: CMUC, Department of Mathematics, University of Coimbra, Apartado 3008, 3001–501 Coimbra, Portugal.

$\hat{h}$	M1 $n=50$	$n=100$	$n=200$	$n=400$	$n=800$
FO	0.8571	0.3753	0.2539	0.0892	0.0465
vM	0.4601	0.2080	0.1004	0.0496	0.0307
OLI	5.3599	1.4722	0.5376	0.1781	0.0747
AMI	1.1582	0.3711	0.1614	0.0434	0.0274
EMI	0.8425	0.3061	0.1466	0.0489	0.0306
LCV	1.4039	0.7841	0.4307	0.2116	0.1015
$\hat{M}=2$	2.1836	0.6566	0.2774	0.1090	0.0584

$\hat{h}$	M2 $n=50$	$n=100$	$n=200$	$n=400$	$n=800$
FO	1.6693	0.9014	0.4815	0.2666	0.1477
vM	1.2620	0.7172	0.4136	0.2389	0.1395
OLI	8.2672	2.1555	0.5965	0.2845	0.1471
AMI	1.7170	0.8206	0.4305	0.2387	0.1394
EMI	1.4168	0.7677	0.4240	0.2387	0.1394
LCV	1.7146	0.9634	0.5684	0.3311	0.1713
$\hat{M}=2$	2.2802	1.1078	0.4495	0.2433	0.1414

$\hat{h}$	M3 $n=50$	$n=100$	$n=200$	$n=400$	$n=800$
FO	3.0550	1.9259	1.1365	0.6567	0.3518
vM	2.5923	1.5536	0.9192	0.5500	0.3280
OLI	16.4319	5.4680	1.4616	0.7283	0.3358
AMI	2.9567	1.5721	0.9184	0.5502	0.3282
EMI	2.7802	1.5472	0.9131	0.5484	0.3277
LCV	3.4402	2.1343	1.1236	0.6332	0.3612
$\hat{M}=2$	5.4840	2.3821	1.0479	0.5881	0.3301

$\hat{h}$	M4 $n=50$	$n=100$	$n=200$	$n=400$	$n=800$
FO	1.2593	0.6841	0.3998	0.2241	0.1248
vM	1.1154	0.6115	0.3683	0.2191	0.1243
OLI	10.2377	2.2710	0.7423	0.3083	0.1605
AMI	1.5720	0.7545	0.4431	0.2466	0.1391
EMI	1.3151	0.6662	0.4118	0.2307	0.1303
LCV	1.7162	0.9995	0.5814	0.3049	0.1832
$\hat{M}=2$	1.9023	0.8911	0.4383	0.2372	0.1324

Table 1: *Empirical L_2–norm of* $\mathrm{ISE}(f;\tilde{f}_{\mathrm{vM}},\hat{h},n)\times 100$ *for the bandwidths* $\hat{h}_{\mathrm{FO}}$, $\hat{h}_{\mathrm{vM}}$, $\hat{h}_{\mathrm{OLI}}$, $\hat{h}_{\mathrm{AMI}}$, $\hat{h}_{\mathrm{EMI}}$, $\hat{h}_{\mathrm{LCV}}$ *and* $\hat{h}_{\mathrm{OLI}}$ *with* $\hat{M}=2$, *and circular density models* 1 *to* 4. *The number of replications is* 500.

	M5				
$\hat{h}$	$n = 50$	$n = 100$	$n = 200$	$n = 400$	$n = 800$
FO	6.7272	3.9458	2.3400	1.2551	0.7410
vM	9.4133	6.9467	5.0972	3.5811	2.4998
OLI	12.8475	5.3785	2.3329	1.2411	0.7289
AMI	6.8758	3.6908	2.1811	1.2454	0.7404
EMI	6.8570	3.7029	2.2276	1.2704	0.7555
LCV	12.0817	7.2185	4.0636	2.1326	1.1698
$\hat{M}=2$	6.8204	4.0971	2.1765	1.2304	0.7381

	M6				
$\hat{h}$	$n = 50$	$n = 100$	$n = 200$	$n = 400$	$n = 800$
FO	4.9139	3.1217	1.9565	1.2189	0.6892
vM	4.3050	3.1154	2.2625	1.6861	1.1995
OLI	15.4331	6.2135	2.3851	1.2409	0.6847
AMI	5.0196	3.1224	1.8874	1.1863	0.6707
EMI	4.6088	2.9371	1.8057	1.1400	0.6482
LCV	4.9555	3.1464	1.9568	1.2448	0.7608
$\hat{M}=2$	5.0210	2.9589	1.8281	1.1342	0.7082

	M7				
$\hat{h}$	$n = 50$	$n = 100$	$n = 200$	$n = 400$	$n = 800$
FO	2.4285	1.4011	0.8029	0.4646	0.2578
vM	9.1469	9.0068	9.0070	8.9819	9.0158
OLI	13.9815	2.8960	0.9267	0.4587	0.2536
AMI	2.6414	1.3996	0.7518	0.4402	0.2514
EMI	2.3033	1.3346	0.7442	0.4376	0.2506
LCV	2.5369	1.4291	0.8020	0.4724	0.2696
$\hat{M}=2$	2.3146	1.3136	0.7438	0.4401	0.2567

	M8				
$\hat{h}$	$n = 50$	$n = 100$	$n = 200$	$n = 400$	$n = 800$
FO	2.8533	1.6173	0.8935	0.5090	0.3050
vM	3.8957	2.7796	1.8634	1.2325	0.7972
OLI	9.9637	4.0549	1.0260	0.5258	0.3013
AMI	2.7655	1.5876	0.8445	0.4946	0.2991
EMI	2.6106	1.5368	0.8365	0.4928	0.2985
LCV	2.8680	1.7182	0.9146	0.5415	0.3225
$\hat{M}=2$	2.6569	1.5114	0.8444	0.4957	0.3027

Table 2: *Empirical L_2–norm of* $\mathrm{ISE}(f; \tilde{f}_{\mathrm{vM}}, \hat{h}, n) \times 100$ *for the bandwidths* $\hat{h}_{\mathrm{FO}}$, $\hat{h}_{\mathrm{vM}}$, $\hat{h}_{\mathrm{OLI}}$, $\hat{h}_{\mathrm{AMI}}$, $\hat{h}_{\mathrm{EMI}}$, $\hat{h}_{\mathrm{LCV}}$ *and* $\hat{h}_{\mathrm{OLI}}$ *with* $\hat{M} = 2$, *and circular density models* 5 *to* 8. *The number of replications is* 500.

$\hat{h}$	M9 $n=50$	$n=100$	$n=200$	$n=400$	$n=800$
FO	1.7187	1.0985	0.6028	0.3329	0.1918
vM	1.3130	0.8613	0.5034	0.2985	0.1777
OLI	8.6427	2.4272	0.7294	0.3274	0.1771
AMI	1.9213	1.1825	0.5476	0.3087	0.1743
EMI	1.5845	1.0670	0.5270	0.3038	0.1740
LCV	1.7386	1.0603	0.6300	0.3442	0.1975
$\hat{M}=2$	3.4859	1.0851	0.5407	0.3005	0.1779

$\hat{h}$	M10 $n=50$	$n=100$	$n=200$	$n=400$	$n=800$
FO	4.8376	3.3165	2.0037	1.1973	0.7260
vM	3.9471	2.9478	2.1437	1.6246	1.2514
OLI	13.8096	5.5740	2.6244	1.2146	0.7051
AMI	4.5031	3.1760	2.0460	1.2332	0.7076
EMI	4.2247	3.0130	1.9835	1.2035	0.6993
LCV	4.8477	3.3501	2.3723	1.6579	1.1073
$\hat{M}=2$	5.8261	3.2672	2.2440	1.3428	0.7750

$\hat{h}$	M11 $n=50$	$n=100$	$n=200$	$n=400$	$n=800$
FO	2.8191	1.5545	0.8953	0.5278	0.3032
vM	6.4178	6.3177	6.2753	6.2498	6.2353
OLI	10.8292	2.8382	1.0283	0.5457	0.3025
AMI	4.3670	1.7991	0.9024	0.5244	0.3011
EMI	4.0373	1.7286	0.8835	0.5210	0.3002
LCV	2.8350	1.6607	0.9516	0.5512	0.3135
$\hat{M}=2$	3.6061	2.5096	2.4624	2.5638	3.1216

$\hat{h}$	M12 $n=50$	$n=100$	$n=200$	$n=400$	$n=800$
FO	2.2935	1.3014	0.7581	0.4271	0.2482
vM	3.4075	2.8887	2.3247	1.7232	1.2425
OLI	7.5424	2.4560	0.8745	0.4317	0.2477
AMI	3.0110	1.4042	0.7777	0.4236	0.2469
EMI	2.6032	1.2841	0.7200	0.4129	0.2447
LCV	2.3336	1.3394	0.7589	0.4475	0.2596
$\hat{M}=2$	2.4594	1.2889	0.6890	0.4129	0.2534

Table 3: *Empirical L_2–norm of* $\mathrm{ISE}(f;\tilde{f}_{\mathrm{vM}},\hat{h},n)\times 100$ *for the bandwidths* $\hat{h}_{\mathrm{FO}}$, $\hat{h}_{\mathrm{vM}}$, $\hat{h}_{\mathrm{OLI}}$, $\hat{h}_{\mathrm{AMI}}$, $\hat{h}_{\mathrm{EMI}}$, $\hat{h}_{\mathrm{LCV}}$ *and* $\hat{h}_{\mathrm{OLI}}$ *with* $\hat{M}=2$, *and circular density models* 9 *to* 12. *The number of replications is* 500.

$\hat{h}$	M13				
	$n=50$	$n=100$	$n=200$	$n=400$	$n=800$
FO	3.2671	2.0340	1.2106	0.6901	0.4024
vM	10.3324	10.1878	10.0906	9.9161	9.6580
OLI	8.3516	3.7104	1.4454	0.6889	0.4008
AMI	4.1362	2.1725	1.2166	0.6743	0.3986
EMI	3.5210	1.9845	1.1492	0.6590	0.3927
LCV	3.3956	2.0397	1.1844	0.6867	0.4045
$\hat{M}=2$	3.4464	2.0698	1.2045	0.7394	0.4439

$\hat{h}$	M14				
	$n=50$	$n=100$	$n=200$	$n=400$	$n=800$
FO	3.6230	1.9958	1.1898	0.7004	0.4098
vM	8.3183	8.2128	8.1636	8.1378	8.1253
OLI	9.8771	3.0111	1.3360	0.7177	0.4099
AMI	6.1660	2.4824	1.2250	0.7066	0.4087
EMI	5.7398	2.3636	1.1939	0.6993	0.4064
LCV	3.7064	2.1106	1.2384	0.7240	0.4234
$\hat{M}=2$	6.8369	6.9169	7.1482	7.2239	7.1870

$\hat{h}$	M15				
	$n=50$	$n=100$	$n=200$	$n=400$	$n=800$
FO	1.2110	0.8531	0.6888	0.4971	0.2650
vM	0.9041	0.7319	0.6517	0.5897	0.5276
OLI	6.9877	1.9514	0.8750	0.4293	0.2355
AMI	1.4372	0.8776	0.6678	0.5911	0.5127
EMI	1.1675	0.7966	0.6585	0.5793	0.4953
LCV	1.6780	1.0622	0.7052	0.4346	0.2523
$\hat{M}=2$	2.3539	0.9864	0.6264	0.4894	0.4140

$\hat{h}$	M16				
	$n=50$	$n=100$	$n=200$	$n=400$	$n=800$
FO	4.9118	2.4273	1.3928	0.8214	0.4762
vM	8.0286	7.9228	7.8650	7.8360	7.8251
OLI	7.8557	2.7624	1.4522	0.8349	0.4778
AMI	7.6447	4.2652	1.4530	0.8351	0.4780
EMI	7.3014	4.1098	1.4066	0.8235	0.4748
LCV	4.2739	2.5145	1.4474	0.8518	0.4881
$\hat{M}=2$	7.5458	7.5261	7.5333	7.5203	7.5019

Table 4: *Empirical* L_2*–norm of* $\mathrm{ISE}(f;\tilde{f}_{\mathrm{vM}},\hat{h},n)\times 100$ *for the bandwidths* $\hat{h}_{\mathrm{FO}}$, $\hat{h}_{\mathrm{vM}}$, $\hat{h}_{\mathrm{OLI}}$, $\hat{h}_{\mathrm{AMI}}$, $\hat{h}_{\mathrm{EMI}}$, $\hat{h}_{\mathrm{LCV}}$ *and* $\hat{h}_{\mathrm{OLI}}$ *with* $\hat{M}=2$*, and circular density models* 13 *to* 16*. The number of replications is* 500*.*

	M17				
$\hat{h}$	$n=50$	$n=100$	$n=200$	$n=400$	$n=800$
FO	7.9383	5.1597	2.8398	1.5884	0.9327
vM	8.5780	7.5080	6.3768	5.4802	4.6068
OLI	11.2766	5.0852	2.8716	1.6259	0.9690
AMI	8.6342	6.6524	4.1481	2.0114	0.9581
EMI	8.5868	6.7004	4.1978	2.0217	0.9870
LCV	8.0645	6.0661	4.0445	2.5430	1.5454
$\hat{M}=2$	8.3586	6.2830	5.1367	4.4947	3.8189

	M18				
$\hat{h}$	$n=50$	$n=100$	$n=200$	$n=400$	$n=800$
FO	4.7903	3.3195	1.9234	0.8934	0.5195
vM	4.1513	3.4189	2.9064	2.4034	1.9160
OLI	13.0529	3.6827	1.9193	0.9363	0.5210
AMI	4.5964	3.4338	2.7348	1.9140	0.6537
EMI	4.4064	3.4217	2.7582	1.9506	0.6580
LCV	4.2462	3.0265	1.8874	1.0421	0.5786
$\hat{M}=2$	4.5001	3.2736	2.6187	2.2033	1.7430

	M15				
$\hat{h}$	$n=50$	$n=100$	$n=200$	$n=400$	$n=800$
FO	4.1508	2.7296	1.9320	1.2602	0.6418
vM	4.1312	3.2961	2.6970	2.2342	1.8989
OLI	9.9848	3.8756	1.9673	1.1166	0.6647
AMI	4.2250	2.7070	1.8736	1.2955	0.8254
EMI	3.8657	2.6095	1.8802	1.3187	0.8606
LCV	3.9235	2.6252	1.8401	1.2215	0.7229
$\hat{M}=2$	3.6402	2.5023	1.8617	1.4142	1.0788

	M20				
$\hat{h}$	$n=50$	$n=100$	$n=200$	$n=400$	$n=800$
FO	6.7279	3.6745	2.3416	1.4669	0.8479
vM	10.6886	10.6007	10.5651	10.5246	10.5163
OLI	10.7344	4.4795	2.4327	1.4316	0.8729
AMI	8.1015	4.4374	2.4644	1.4396	0.8672
EMI	7.6957	4.3605	2.4590	1.4721	0.8908
LCV	5.7748	3.6836	2.3009	1.4248	0.8634
$\hat{M}=2$	7.4523	6.4313	5.5767	4.6722	3.8507

Table 5: *Empirical L_2–norm of* $\mathrm{ISE}(f;\tilde{f}_{\mathrm{vM}},\hat{h},n)\times 100$ *for the bandwidths* $\hat{h}_{\mathrm{FO}}$, $\hat{h}_{\mathrm{vM}}$, $\hat{h}_{\mathrm{OLI}}$, $\hat{h}_{\mathrm{AMI}}$, $\hat{h}_{\mathrm{EMI}}$, $\hat{h}_{\mathrm{LCV}}$ *and* $\hat{h}_{\mathrm{OLI}}$ *with* $\hat{M}=2$*, and circular density models* 17 *to* 20*. The number of replications is* 500.