Pré-Publicações do Departamento de Matemática
Universidade de Coimbra
Preprint Number 21–10

KERNEL DENSITY ESTIMATION FOR CIRCULAR DATA: A FOURIER SERIES-BASED PLUG-IN APPROACH FOR BANDWIDTH SELECTION

CARLOS TENREIRO

ABSTRACT: Asymptotic expansions for the mean integrated squared errors of two classes of kernel density estimators for circular data are presented in this paper. Based on them, explicit expressions for the corresponding asymptotic optimal bandwidths are given, and a Fourier series-based plug-in approach for bandwidth selection is presented. The proposed bandwidth selectors have a $n^{-1/2}$ relative convergence rate whenever the underlying density is smooth enough and the simulation results testify that they present a very good finite sample performance against other bandwidth selection methods in the literature.

KEYWORDS: Circular density; Kernel estimator; Bandwidth selection; Plug-in rule; Fourier series-based projection estimators.
AMS SUBJECT CLASSIFICATION (2010): 62G07, 62G20.

1. Introduction

Kernel methods for estimating densities of q-dimensional unit spheres, for $q \geq 2$, have been studied in the seminal papers of Hall et al. (1987) and Bai et al. (1988) where several asymptotic properties of the considered estimators are stated. More recent developments on this topic include the works of Klemela (2000) and García-Portugués et al. (2013). Regarding the simplest univariate case of estimating densities in the unit circle ($q = 1$), some more recent works that address the important topic of the automatic selection of the kernel estimator smoothing parameter, usually called the bandwidth, comprise the papers of Taylor (2008), Di Marzio et al. (2009) and Oliveira et al. (2012). As for the well studied case of density estimation for linear data (see Wand and Jones, 1995, Chap. 3), the reference distribution scale rules or other more sophisticated plug-in methods for selecting the bandwidth are based on the theoretical expression for the bandwidth that minimises the main terms of the asymptotic expansion for the mean integrated squared error of the estimator. Unfortunately, as shown in the present work, the asymptotic expansions for the estimator mean integrated squared error considered

Received April 26, 2021.

in the previous papers are not correct leading to improper implementations of the proposed data-dependent bandwidth selectors. In this paper we propose to fill this gap in the nonparametric density estimation literature for circular data. Other than permitting the correction of the above mentioned plug-in selectors, the new theoretical expression for the asymptotic optimal bandwidth presented in this work lead us to propose a Fourier series-based plug-in approach for bandwidth selection in kernel estimation for circular data. These are the two main goals of this paper.

The mean integrated squared error asymptotic expansions derived in this work are established for a general class of δ-sequence density estimators. 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, a δ-sequence estimator of f 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 δ-function sequence, which satisfies a set of conditions described in Section 2.1. The reader is referred to Watson and Leadbetter (1964) for the concept of δ-function sequence in the context of linear data. This general class of estimators includes the class of kernel estimators considered in Hall et al. (1987) and Bai et al. (1988), defined, for $\theta \in [0, 2\pi[$, by

$$\hat{f}_{\mathrm{HB}}(\theta; h) = \frac{c_h(L)}{n} \sum_{i=1}^{n} L\left(\frac{1 - \cos(\theta - X_i)}{h^2}\right), \tag{2}$$

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 $\hat{f}_{\mathrm{HB}}(\cdot; h)$ integrates to unity. If $L(t) = e^{-t}$ then (2) is the density estimator considered in Taylor (2008), Di Marzio et al. (2009) and Oliveira et al. (2012). In this case the estimator is a combination of circular normal or von Mises distributions with mean directions X_i and concentration parameters equal to $\nu = h^{-2}$ as it takes the form

$$\hat{f}_{\text{HB}}(\theta; h) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{(2\pi) I_0(\nu)} \exp\big(\nu \cos(\theta - X_i)\big), \tag{3}$$

where $I_r(\nu)$ is, for $\nu \geq 0$ and $r \geq 0$, the modified Bessel function of order r defined by

$$I_r(\nu) = \frac{1}{2\pi} \int_0^{2\pi} \cos(r\theta) \exp(\nu \cos\theta) d\theta. \tag{4}$$

The considered class of δ-sequence estimators also comprises an estimator suggested in Silverman (1986, pp. 31–32), 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

$$\hat{f}_{\text{PR}}(\theta; g) = \frac{d_g(K)}{n} \sum_{i=1}^{n} K_g(\theta - X_i), \tag{5}$$

where K_g is a real-valued periodic function on $\mathbb{R}$, with period 2π, such that $K_g(\theta) = K(\theta/g)$, for $\theta \in [-\pi, \pi[$, with $K : \mathbb{R} \to \mathbb{R}$ a bounded and symmetric function, $g = g_n > 0$ 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 $\hat{f}_{\text{PR}}(\cdot; g)$ integrates to unity. Of course, for $K(u) = L(u^2)$ and $g = \sqrt{2}\, h$ the estimators $\hat{f}_{\text{HB}}$ and $\hat{f}_{\text{PR}}$ 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 $\hat{f}_{\text{PR}}$ is close to a kernel density estimator for linear data based on the normal or Gaussian kernel.

The rest of this paper is organised as follows. In Section 2 we establish sufficient conditions on the δ-function sequence which ensure the consistency of $\hat{f}_n$ as estimator of f and we present asymptotic expansions for its bias and variance. The particular cases of estimators (2), (3) and (5) are discussed in detail. In Section 3 we present asymptotic expansions for the mean integrated squared errors of estimators (2) and (5). As in kernel estimation for linear data, these asymptotic expansions enable us to derive explicit expressions for the asymptotic optimal bandwidths, to compare estimators (2) and (5), 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 be also quantified. As in optimal kernel theory for linear data, we conclude that one loses very little when suboptimal kernels are used. In Section 4 we follow the strategy of Tenreiro (2011) in order to propose Fourier series-based direct plug-in bandwidth selectors for estimators (2) and (5) (see also Tenreiro, 2020). They achieve the relative convergence rate $n^{-1/2}$ whenever the underlying density is smooth enough. In Section 5 the finite-sample behaviour of the proposed bandwidth selection method 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, while in Section 8 we gather some of the proofs.

The simulations and plots in this paper were performed using programs written in the R language (R Development Core Team, 2019) and the R package 'circular' (Lund and Agostinelli, 2017).

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. Throughout this section we assume that the δ-function sequence (δ_n), where the functions $\delta_n : \mathbb{R} \to [0, \infty[$ are periodic with period 2π, satisfies the following conditions that will ensure the consistency of $\hat{f}_n$:

(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 $\lambda > 0$;

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

On similar conditions for linear data see Watson and Leadbetter (1964, p. 102). See also Walter and Blum (1979).

In the case of estimator $\hat{f}_{\mathrm{HB}}$ given by (2), these conditions 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) $\int_0^\infty t^{-1/2} L(t) dt < \infty$.

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

$$\alpha(\delta_n) = h^{-1} 2^{-1/2} \int_0^\infty t^{-1/2} L(t)^2 dt \left(\int_0^\infty t^{-1/2} L(t) dt \right)^{-2} (1 + o(1)). \quad (6)$$

For $L(t) = e^{-t}$ and $\nu = h^{-2}$ we get

$$\alpha(\delta_n) = \frac{I_0(2\nu)}{2\pi I_0(\nu)^2} = \sqrt{\frac{\nu}{4\pi}} (1 + o(1)). \quad (7)$$

Concerning estimator $\hat{f}_{\mathrm{PR}}$ given by (5), the previous conditions 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) $\lim_{u \to +\infty} u K(u) = 0$;

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

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

$$\alpha(\delta_n) = g^{-1} \int_{-\infty}^\infty K(u)^2 du \left(\int_{-\infty}^\infty K(u) du \right)^{-2} (1 + o(1)). \quad (8)$$

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 asympotic expansions (6) and (8) coincide whenever $g = \sqrt{2}\, h$.

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

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

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

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

From Theorem 1 and (6) we conclude that under conditions (L.1) and (L.2) on L, $\hat{f}_{\mathrm{HB}}$ is a consistent estimators of f, for all density f continuous on $[0, 2\pi]$, whenever the smoothing parameter satisfy the classical conditions $h \to 0$, $nh \to +\infty$, as n tends to infinity. Taking into account (8) a similar result holds for estimator $\hat{f}_{\mathrm{PR}}$ whenever K satisfies conditions (K.1) and (K.2) and g is such that $g \to 0$, $ng \to +\infty$, as $n \to +\infty$.

2.2. Asymptotic expansions. From Theorem 1, an uniform asymptotic expansion for the variance of the estimator $\hat{f}_n$ is given by

$$\sup_{\theta\in[0,2\pi[} \big|\mathrm{Var}\hat{f}_n(\theta) - n^{-1}\alpha(\delta_n)f(\theta)\big| = o\big(n^{-1}\alpha(\delta_n)\big). \tag{9}$$

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

(Δ.4) $\displaystyle\int_{-\pi}^{\pi} y\delta_n(y)dy = 0$, 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

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

Taking into account their symmetry, the δ-function sequences associated to estimators $\hat{f}_{\mathrm{HB}}$ and $\hat{f}_{\mathrm{PR}}$ trivially fulfill the first of the previous assumptions. The second condition holds for estimator $\hat{f}_{\mathrm{HB}}$ if the kernel L is such that

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

In this case we have

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

For $L(t) = e^{-t}$ and $\nu = h^{-2}$ we get

$$\beta(\delta_n) = \frac{J_2(\nu)}{I_0(\nu)} = \nu^{-1} (1 + o(1)) \qquad (11)$$

with

$$J_2(\nu) = \frac{1}{2\pi} \int_{-\pi}^{\pi} y^2 \exp(\nu \cos y) dy. \qquad (12)$$

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

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

Moreover, we have

$$\beta(\delta_n) = g^2 \int_{-\infty}^{\infty} u^2 K(u) du \left(\int_{-\infty}^{\infty} K(u) du \right)^{-1} (1 + o(1)). \qquad (13)$$

Kernel L fulfills condition (L.3) iff the kernel K defined by $K(u) = L(u^2)$ satisfies condition (K.3), and the sequences (10) and (13) 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*

$$\left| f''(x) - f''(y) \right| \leq C|x - y|^\alpha, \; x, y \in [0, 2\pi], \qquad (14)$$

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

$$\sup_{\theta \in [0, 2\pi[} \left| \mathrm{E}\hat{f}_n(\theta) - f(\theta) - \frac{1}{2}\beta(\delta_n) f''(\theta) \right| = o\big(\beta(\delta_n)\big), \qquad (15)$$

where

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

It is possible to improve the asymptotic rate of convergence of $\hat{f}_n$ by allowing the δ-functions sequence (δ_n) to take negative values (cf. Wand and Jones, 1995, pp. 32–35). Nevertheless, this issue is not pursued here.

3. MISE expansions and asymptotic optimal bandwidths

The mean integrated squared error defined by

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

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 $n\to+\infty$, the variance and bias expansions (9) and (15) lead to the following asymptotic expansion for mean integrated squared error of the δ-sequence estimator (1):

$$\mathrm{MISE}(f;\hat{f}_n,n)=n^{-1}\alpha(\delta_n)+\frac{1}{4}\beta(\delta_n)^2\theta_2(f)+o\big(n^{-1}\alpha(\delta_n)+\beta(\delta_n)^2\big),\quad (16)$$

where $\theta_2(f)$ denotes the quadratic funcional

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

In this section we discuss some consequences of this general expansion in the case of kernel estimators $\hat{f}_{\mathrm{HB}}$ and $\hat{f}_{\mathrm{PR}}$ defined by (2) and (5), respectively.

Theorem 3. *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*

$$\mathrm{MISE}(f;\hat{f}_{\mathrm{HB}},h,n)=\frac{1}{nh}\boldsymbol{c}_1(L)+h^4\boldsymbol{c}_2(L)\,\theta_2(f)+o\bigg(\frac{1}{nh}+h^4\bigg),$$

where

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

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, 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

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

where

$$\boldsymbol{c}(L) = 2^{-1/2} \left(\int_0^\infty t^{-1/2} L(t)^2 dt \right)^{1/5} \left(\int_0^\infty t^{1/2} L(t) dt \right)^{-2/5}. \tag{17}$$

The previous asymptotic expansion for the mean integrated squared error corrects the corresponding expansions presented in Taylor (2008, p. 3495), Di Marzio et al. (2009, Theorem 1, p. 2068) and Oliveira et al. (2012, expression (2), p. 3899). In order to better compare the previous asymptotic expansion with those appearing in the literature, let us take $L(t) = e^{-t}$ and denote $\nu = h^{-2}$. In this case, from Theorem 3 the asymptotic mean integrated squared error of the estimator is given by

$$\text{AMISE}(f; \hat{f}_{\text{HB}}, h, n) = \frac{\nu^{1/2}}{2n\pi^{1/2}} + \frac{1}{4\nu^2}\theta_2(f). \tag{18}$$

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 $f_{vM(\mu,k)}$, we get

$$\theta_2(f_{vM(\mu,k)}) = \frac{3\kappa^2 I_0(2\kappa) - \kappa I_1(2\kappa)}{8\pi I_0(\kappa)^2}, \tag{19}$$

where the modified Bessel functions I_0 and I_1 are given by (4). Replacing $\theta_2(f)$ by (19) in (18), the previous asymptotic mean integrated squared error expression can be compared with that one given in Taylor (2008, p. 3495):

$$\text{AMISE}(f; \hat{f}_{\text{HB}}, h, n) = \frac{\nu^{1/2}}{2n\pi^{1/2}} + \frac{3\kappa^2 I_2(2\kappa)}{32\pi\nu^2 I_0(\kappa)^2}.$$

We see that integrated variance terms agree in both expressions, but the same does not happen with respect to the integrated squared bias terms.

From (7), (11) and (16) an alternative expression for the asymptotic mean integrated squared error is

$$\text{AMISE}(f; \hat{f}_{\text{HB}}, h, n) = \frac{I_0(2\nu)}{2n\pi I_0(\nu)^2} + \frac{J_2(\nu)^2}{4I_0(\nu)^2}\theta_2(f),$$

where J_2 is given by (12). This expression can be easily compared with expression (2) in Oliveira et al. (2012, p. 3899):

$$\text{AMISE}(f; \hat{f}_{\text{HB}}, h, n) = \frac{I_0(2\nu)}{2n\pi I_0(\nu)^2} + \frac{1}{16}\left[1 - \frac{I_2(\nu)}{I_0(\nu)}\right]^2 \theta_2(f).$$

As before, the integrated variance terms agree in both expressions but the integrated squared bias term of this last expression, taken from Di Marzio et al. (2009), is not correct.

Theorem 4. *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*

$$\mathrm{MISE}(f; \hat{f}_{\mathrm{PR}}, g, n) = \frac{1}{ng}\boldsymbol{d}_1(K) + \frac{g^4}{4}\boldsymbol{d}_2(K)\,\theta_2(f) + o\Big(\frac{1}{ng} + g^4\Big),$$

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 $\hat{f}_{\mathrm{PR}}$ is given by

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

where

$$\boldsymbol{d}(K) = 2^{-2/5}\Big(\int_{-\infty}^{\infty} K(u)^2 du\Big)^{1/5} \Big(\int_{-\infty}^{\infty} u^2 K(u)du\Big)^{-2/5}. \tag{20}$$

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 $\hat{f}_{\mathrm{HB}}$ with kernel L and bandwidth h and the estimator $\hat{f}_{\mathrm{PR}}$ with kernel $K(u) = L(u^2)$ and bandwidth $g = \sqrt{2}\,h$ share the same first-order asymptotics 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 δ-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 δ-function sequence satisfying conditions (Δ.1) and (Δ.3). In the particular case of estimator $\hat{f}_{\rm HB}$, and under very general conditions on the kernel L, we have $\mathrm{MISE}(f;\hat{f}_{\rm HB},h,n) = o(1)$ 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 $\mathrm{MISE}(f;\hat{f}_{\rm HB},h,n) = o(n^{-1})$. A similar result is valid for estimator $\hat{f}_{\rm PR}$.

The asymptotic comparison between estimators $\hat{f}_{\rm HB}$ and $\hat{f}_{\rm PR}$, 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 $\hat{f}_{\rm HB}$ and $\hat{f}_{\rm PR}$, from Theorems 3 and 4 we know that their mean integrated squared errors are such that

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

and

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

where

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

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

and

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

This last functional is well-known in the context of kernel estimation for linear data. We know that the parabolic kernel $K^*(u) = (1-u^2)I(|u|\le 1)$ 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 $L^*(t) = (1-t)I(t\le 1)$ 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

$L(t)$	$K(u)$	$\mathrm{eff}(L) = \mathrm{eff}(K)$
$I(t \le 1)$	$I(\lvert u\rvert \le 1)$	0.9295
$(1-t)I(0 \le t \le 1)$	$(1-u^2)I(\lvert u\rvert \le 1)$	1
$(1-t)^2 I(t \le 1)$	$(1-u^2)^2 I(\lvert u\rvert \le 1)$	0.9939
$(1-t)^3 I(t \le 1)$	$(1-u^2)^3 I(\lvert u\rvert \le 1)$	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^*.*

each one of the classes of estimators $\hat f_{\mathrm{HB}}$ and $\hat f_{\mathrm{PR}}$, 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 $\mathrm{eff}(L) := \big(\phi(L^*)/\phi(L)\big)^{5/4}$ and $\mathrm{eff}(K) := \big(\psi(K^*)/\psi(K)\big)^{5/4}$ (see Wand and Jones, 1995, p. 31). For some kernels, these efficiencies are reported in Table 1. As in kernel estimation for linear data framework, 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 (2), or when a nonnegative, symmetric and bounded kernel K satisfying conditions (K.1)–(K.3) is used in (5), under some smoothness assumptions on f we have seen that the asymptotic optimal bandwidths for estimators (2) and (5) are respectively given by

$$h^* = \boldsymbol{c}(L)\,\theta_2(f)^{-1/5} n^{-1/5} \text{ and } g^* = \boldsymbol{d}(K)\,\theta_2(f)^{-1/5} n^{-1/5}, \tag{21}$$

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 (17) and (20), respectively. As the only unknown quantity in (21) is $\theta_2(f)$, 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 idea of the direct plug-in approach to bandwidth selection. For classical references on the direct plug-in method the reader is referred to Woodroofe (1970), Nadaraya (1974) and Deheuvels and Hominal (1980).

In order to define plug-in selectors based on the previous theoretical bandwidths, we follow the approach of Tenreiro (2011) where the Fourier series-based estimators studied in Laurent (1997) are used to estimate the quadratic functional $\theta_2(f)$. In order to define these estimators, let us denote by $\{p_\ell, \ell \in \mathbb{N}_0\}$ the orthonormal Fourier basis of $L_2([0,2\pi])$ given by

$$p_0(x) = \frac{1}{\sqrt{2\pi}},\ \ p_{2\ell-1}(x) = \frac{1}{\sqrt{\pi}}\sin(\ell x),\ \ p_{2\ell}(x) = \frac{1}{\sqrt{\pi}}\cos(\ell x),$$

for $\ell = 1, 2, \ldots$. The Fourier series-based or projection estimator of $\theta_2(f)$ is motivated by the representation $\theta_2(f) = \sum_{\ell=1}^{\infty} \ell^4 c_\ell$, where $c_\ell = a_{2\ell-1}^2 + a_{2\ell}^2$ with $a_\ell = \int_0^{2\pi} f(x) p_\ell(x) dx$ the ℓ-th Fourier coefficients of f. It is defined by

$$\hat{\hat{\theta}}_{2,m} = \sum_{\ell=1}^{m} \ell^4\, \hat{\hat{c}}_\ell, \tag{22}$$

where $\hat{\hat{c}}_\ell$ is the unbiased estimator of c_ℓ given by

$$\hat{\hat{c}}_\ell = \frac{2}{n(n-1)} \sum_{1 \le j < k \le n} \{p_{2\ell-1}(X_j)p_{2\ell-1}(X_k) + p_{2\ell}(X_j)p_{2\ell}(X_k)\}, \tag{23}$$

and $m = m(n)$ is a sequence on integers converging to infinity. As shown in Laurent (1997), these estimators achieve the $n^{-1/2}$ rate of convergence, whenever f is smooth enough and they are 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

$$\hat{\hat{\theta}}_{2,m} = \theta_2(\tilde{f}_m) = \sum_{\ell=1}^{m} \ell^4\, \hat{\hat{c}}_\ell, \tag{24}$$

where $\tilde{f}_m(x) = \sum_{\ell=0}^{2m} \hat{a}_\ell p_\ell(x)$ is the Fourier series-based estimator of f studied in Kronmal and Tarter (1968), $\hat{a}_\ell$ is the unbiased estimator of a_ℓ given by $\hat{a}_\ell = \frac{1}{n}\sum_{i=1}^{n} p_\ell(X_i)$ and $\hat{\hat{c}}_\ell = \hat{a}_{2\ell-1}^2 + \hat{a}_{2\ell}^2$.

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 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 $\hat{\tilde{\theta}}_{2,\hat{m}}$ of $\theta_2(f)$, whose asymptotic behaviour is established in Tenreiro (2011, Lemma 1, pp. 543–544).

Next we describe the asymptotic behaviour of the relative errors associated to the plug-in bandwidth selector defined by

$$\hat{h}^*_{\hat{m}} = \boldsymbol{c}(L)\,\hat{\theta}_{2,\hat{m}}^{-1/5} n^{-1/5}, \tag{25}$$

where $\hat{\theta}_{2,m}$ denotes either $\hat{\bar{\theta}}_{2,m}$ or $\hat{\tilde{\theta}}_{2,m}$ defined by (22) and (24), respectively. Of course, if K is a nonnegative, bounded and symmetric kernel satisfying assumptions (K.1)–(K.3), the same asymptotic behaviour can be established for the relative error associated to the plug-in bandwidth selector defined by

$$\hat{g}^*_{\hat{m}} = \boldsymbol{d}(K)\,\hat{\theta}_{2,\hat{m}}^{-1/5} n^{-1/5}. \tag{26}$$

Theorem 5. *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*

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

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*

$$\mathrm{P}\big(C_1 n^{\xi_1} \leq \hat{m} \leq C_2 n^{\xi_2}\big) \to 1, \tag{27}$$

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

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

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 (27) with*

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

then

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

with

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

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 above mentioned Fourier series-based estimator $\tilde{f}_m$ (Kronmal and Tarter, 1968). 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

$$\mathrm{MISE}(\tilde{f}_m)=H(m)+\sum_{\ell=1}^{\infty}a_\ell^2,$$

where

$$H(m)=\frac{m}{n\pi}-\frac{n+1}{n}\sum_{\ell=1}^{m}\big(a_{2\ell-1}^2+a_{2\ell}^2\big),$$

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

$$\hat{m}_\gamma=\arg\min_{m\in\mathscr{M}_n}\hat{H}_\gamma(m), \tag{28}$$

where

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

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 nonnegative integers, $0<\gamma\le 1$ needs to be chosen by the user, and $\hat{\hat{c}}_\ell$ is given by (23). The value $\hat{m}_\gamma$ 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 $L_n=\lfloor C_1n^{\xi_1}\rfloor+1$ and $U_n=\lfloor C_2n^{\xi_2}\rfloor$, 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 5, 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 5, 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 the empirical distribution of the integrated squared error $\mathrm{ISE}(f; \hat{f}_{\mathrm{HB}}, h, n) = \int_0^{2\pi} \{\hat{f}_{\mathrm{HB}}(\theta) - f(\theta)\}^2 d\theta$, based on 500 samples from the circular densities M3 and M19 considered in Oliveira et al. (2012), where $\hat{f}_{\mathrm{HB}}$ is given by (2) with $L(t) = e^{-t}$ and $h = \hat{h}^*_m$ 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 $\mathrm{EISE}(m) := \mathrm{E}\big(\mathrm{ISE}(f; \hat{f}_{\mathrm{HB}}, \hat{h}^*_m, n)\big)$. 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 1500 points and the composite Simpson's rule. As for other densities that present simple distribution 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 distribution 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

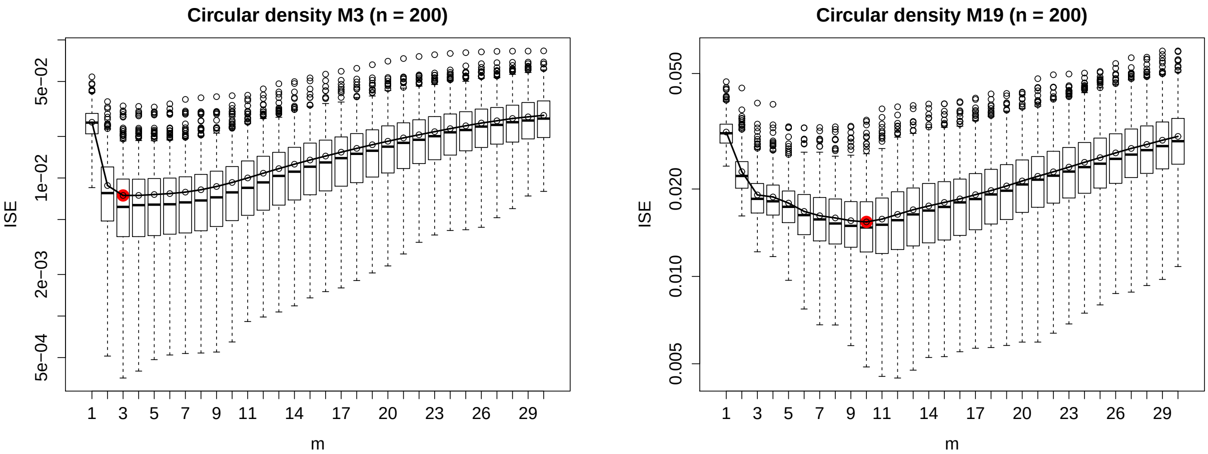

FIGURE 1. Empirical distribution of $\mathrm{ISE}(f; \hat{f}_{\mathrm{HB}}, \hat{h}^*_m, n)$ depending on m for models M3 and M19 ($n = 200$) from the Oliveira et al. (2012) set of circular density models. The number of replications is 500.

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

(25) and (26), with $\hat{m}$ given by (28), were very similar, we will restrict our attention to the kernel estimator (2) associated to the first of these bandwidths. Moreover, as the estimator $\hat{\bar{\theta}}_{2,\hat{m}}$ of θ_2 defined by (22) may occasionally produce poor, sometimes negative, estimates of $\theta_2(f)$ when the size of the sample is small, and it performs similarly to $\hat{\hat{\theta}}_{2,\hat{m}}$ defined by (24) when the sample size is moderate or large, the data-dependent bandwidth based on $\hat{\bar{\theta}}_{2,\hat{m}}$ is not considered hereafter. Finally, as the estimator (3) is used in the already mentioned papers of Taylor (2008), Di Marzio et al. (2009) and Oliveira et al. (2012), that address the automatic selection of the smoothing parameter, from now on we take the kernel $L(t) = e^{-t}, t \geq 0$. Therefore, as $c(L) = (4\pi)^{-1/10}$ for this kernel, the Fourier series-based plug-in bandwidth we consider henceforth, we denote by $\hat{h}_{\mathrm{FO}}$, is defined by

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

where $\hat{m} = \hat{m}_\gamma$ is given by (28) with L_n, U_n and γ chosen as explained in the previous section, and $\tilde{f}_m$ is the above mentioned Fourier series-based estimator of f (Kronmal and Tarter, 1968).

Two other plug-in bandwidths existing in the literature are included in this study. The simplest one is considered for the first time in Taylor (2008, p. 3496), who adapted 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 k, from (19) and (21) we can define the von Mises reference distribution bandwidth selector by

$$\begin{aligned}\hat{h}_{\mathrm{vM}} &= (4\pi)^{-1/10}\theta_2(f_{vM(\hat{\mu},\hat{\kappa})})^{-1/5}n^{-1/5}\\ &= (4\pi)^{-1/10}\left(\frac{3\hat{\kappa}^2 I_0(2\hat{\kappa}) - \hat{\kappa}I_1(2\hat{\kappa})}{8\pi I_0(\hat{\kappa})^2}\right)^{-1/5} n^{-1/5},\end{aligned}$$

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

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

The previous expression for $\hat{h}_{\text{vM}}$ corrects those considered in Taylor (2008, p. 3495) and Oliveira et al. (2012, p. 3899).

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 distributions, that is, f takes the form $f_{M,\alpha,\mu,\kappa} = \sum_{i=1}^{M} \alpha_i f_{vM(\mu_i,\kappa_i)}$, where the proportions α_i are such that $\sum_{i=1}^{M} \alpha_i = 1$, and μ_i and κ_i denote the mean directions and the concentration parameters of the different von Mises distributions. 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 and the selection of the number of mixture components is performed by using Akaike Information Criterion (AIC) (see Oliveira et al., 2012, p. 3900). Denoting by $f_{\hat{M},\hat{\alpha},\hat{\mu},\hat{\kappa}}$ the selected reference distribution density, the associated plug-in bandwidth is given by

$$\hat{h}_{\text{mvM}} = (4\pi)^{-1/10}\theta_2(f_{\hat{M},\hat{\alpha},\hat{\mu},\hat{\kappa}})^{-1/5}n^{-1/5}.$$

We adopt the implementation of the method as described in Oliveira et al. (2012, p. 3903) and given by the function bw.PI of the R package 'NPCirc' (Oliveira et al., 2015). 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 and/or from the numerical approximation of the integral $\theta_2(f_{\hat{M},\hat{\alpha},\hat{\mu},\hat{\kappa}})$, which may not be finite. In this situation one takes $\hat{M} = 1$ in which case the bandwidth selected by $\hat{h}_{\text{mvM}}$ is the von Mises reference distribution bandwidth.

Two other data-driven procedures for selection the bandwidth, already 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 $\hat{f}_{\text{HB},-i}$ the kernel density estimator (3) by leaving out the i-th observation, the least-square cross-validation bandwidth $\hat{h}_{\text{LSCV}}$ is obtained by minimising the classic least-square cross-validation criterion function given by $\text{LSCV}(h) = \int_0^{2\pi} \hat{f}_{\text{HB}}(\theta; h)d\theta - 2n^{-1}\sum_{i=1}^{n} \hat{f}_{\text{HB},-i}(X_i, h)$, whereas the likelihood cross-validation bandwidth $\hat{h}_{\text{LCV}}$ is obtained by maximising the likelihood cross-validation criterion function defined by $\text{LCV}(h) = \prod_{i=1}^{n} \hat{f}_{\text{HB},-i}(X_i; h)$. These methods are implemented by the function bw.CV from the package 'NPCirc' in R (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 compared it with the data-driven bandwidths $\hat{h}_{\mathrm{vM}}$, $\hat{h}_{\mathrm{mvM}}$, $\hat{h}_{\mathrm{LSCV}}$ 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 is analysed through the measure of stochastic performance defined by

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

This performance measure takes into account not only the mean of the $\mathrm{ISE}(f;\hat{f}_{\mathrm{HB}},\hat{h},n)$ 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 $\hat{h}_{\mathrm{FO}}$, $\hat{h}_{\mathrm{vM}}$, $\hat{h}_{\mathrm{mvM}}$ and $\hat{h}_{\mathrm{LCV}}$ are reported in Figures 2, 3, 4 and 5. In these figures the empirical L_2–norm of $\mathrm{ISE}(f;\hat{f}_{\mathrm{HB}},\hat{h},n)$, based on 500 replications, is shown for sample sizes $n = 25 \cdot 2^k$, $k = 1,\ldots,5$.

As we can see from the graphics, the bandwidth $\hat{h}_{\mathrm{vM}}$ is suitable when the underlying density has a distribution structure that is close to a von Mises distribution. This situation occurs with models 1, 2, 3, 4, 9, 15. 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, 20. In all these cases the sampling distribution of the considered concentration parameter estimator $\hat{\kappa}$ is distributed near zero leading to large bandwidths that provide uniform estimates for the underlying circular density (on this situation, see Oliveira et al., 2012, pp. 3906). With respect to the bandwidth $\hat{h}_{\mathrm{mvM}}$, we can see

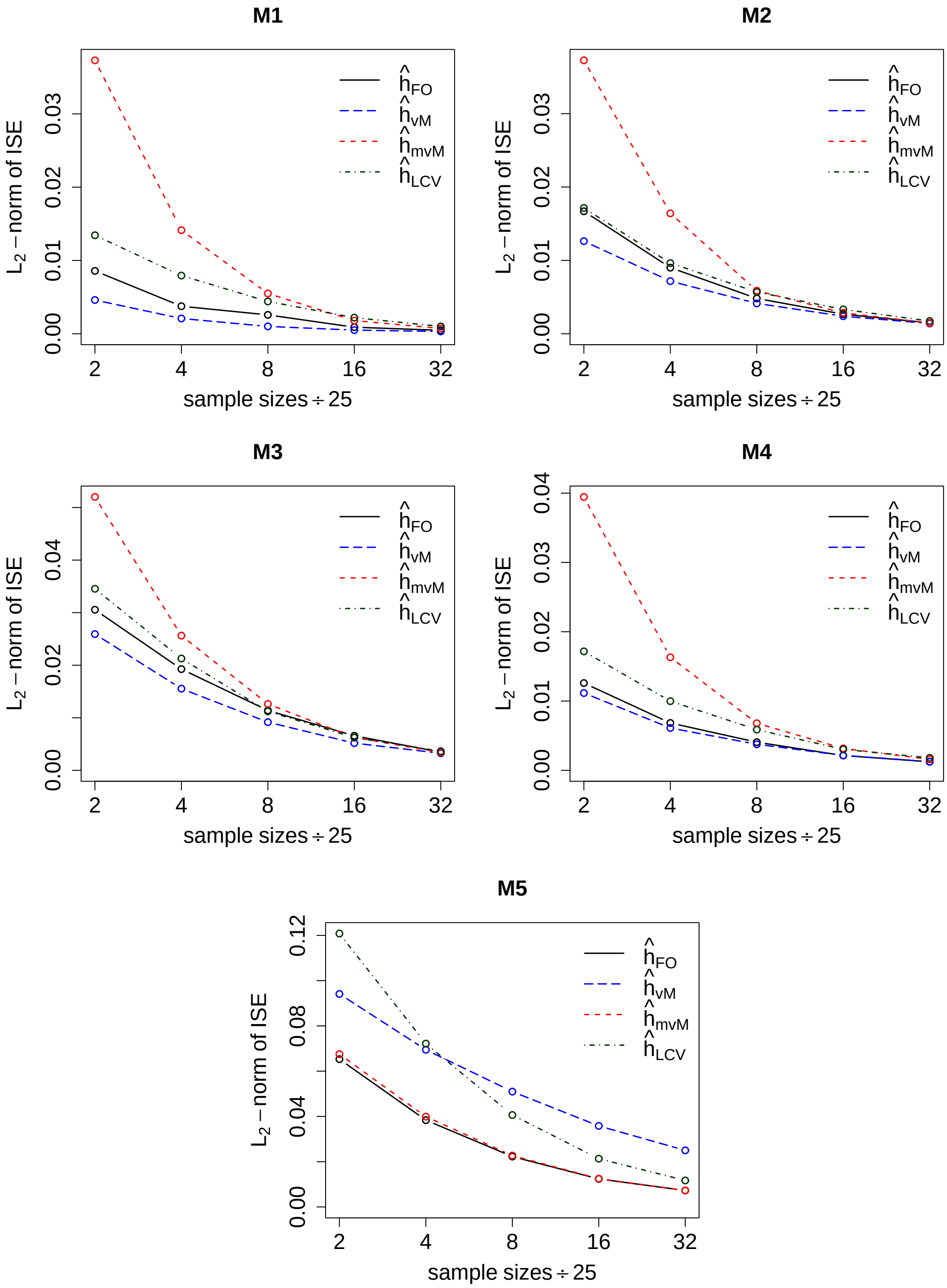

FIGURE 2. *Empirical* L_2*–norm of* $\mathrm{ISE}(f;\hat{f}_{\mathrm{HB}},\hat{h},n)$ *associated to the bandwidths* $\hat{h}_{\mathrm{FO}}$*,* $\hat{h}_{\mathrm{vM}}$*,* $\hat{h}_{\mathrm{mvM}}$ *and* $\hat{h}_{\mathrm{LCV}}$*, for circular density models* 1 *to* 5*. The number of replications is* 500*.*

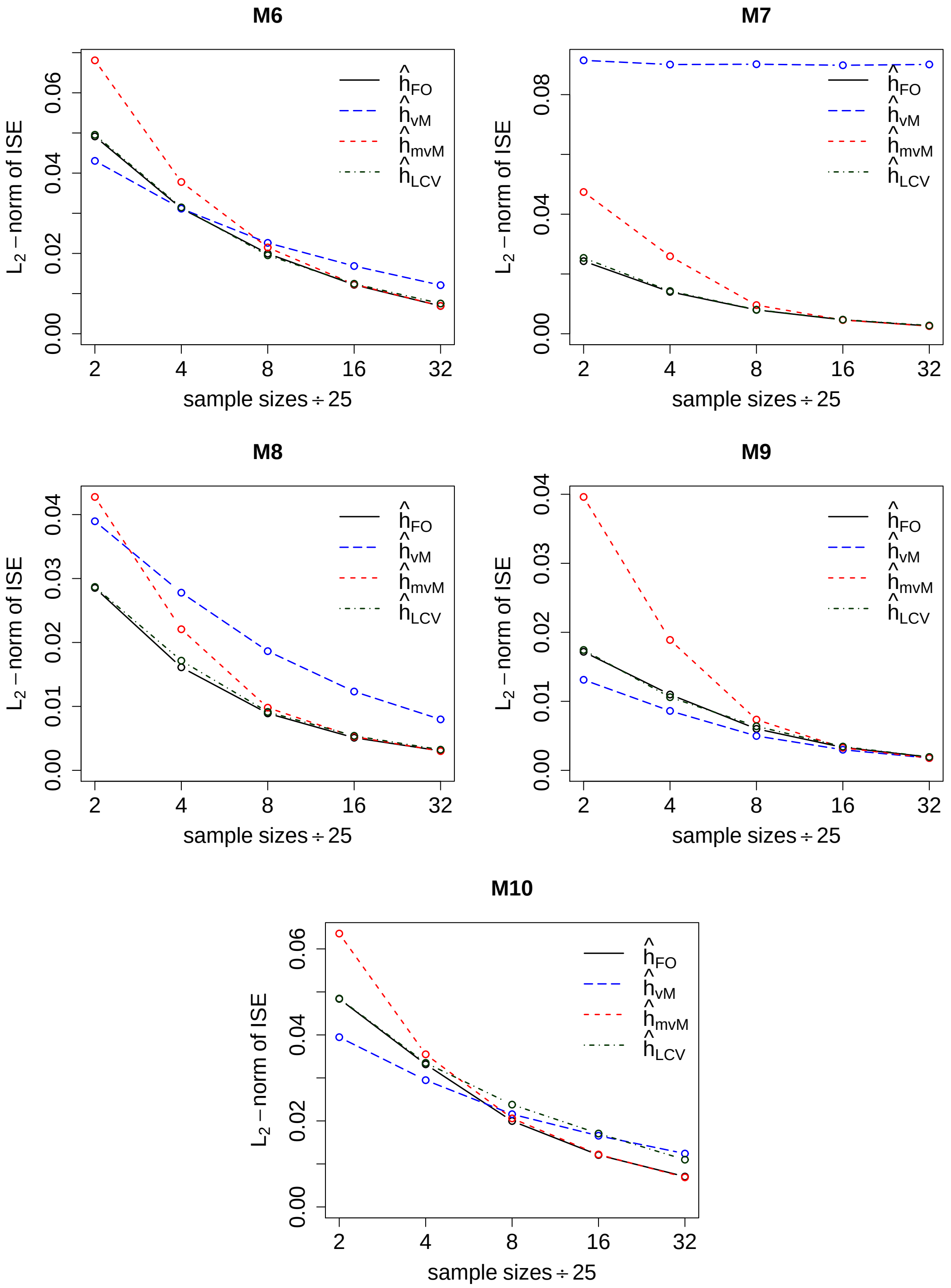

FIGURE 3. *Empirical* L_2*–norm of* $\mathrm{ISE}(f; \hat{f}_{\mathrm{HB}}, \hat{h}, n)$ *associated to the bandwidths* $\hat{h}_{\mathrm{FO}}$, $\hat{h}_{\mathrm{vM}}$, $\hat{h}_{\mathrm{mvM}}$ *and* $\hat{h}_{\mathrm{LCV}}$*, for circular density models* 6 *to* 10*. The number of replications is* 500*.*

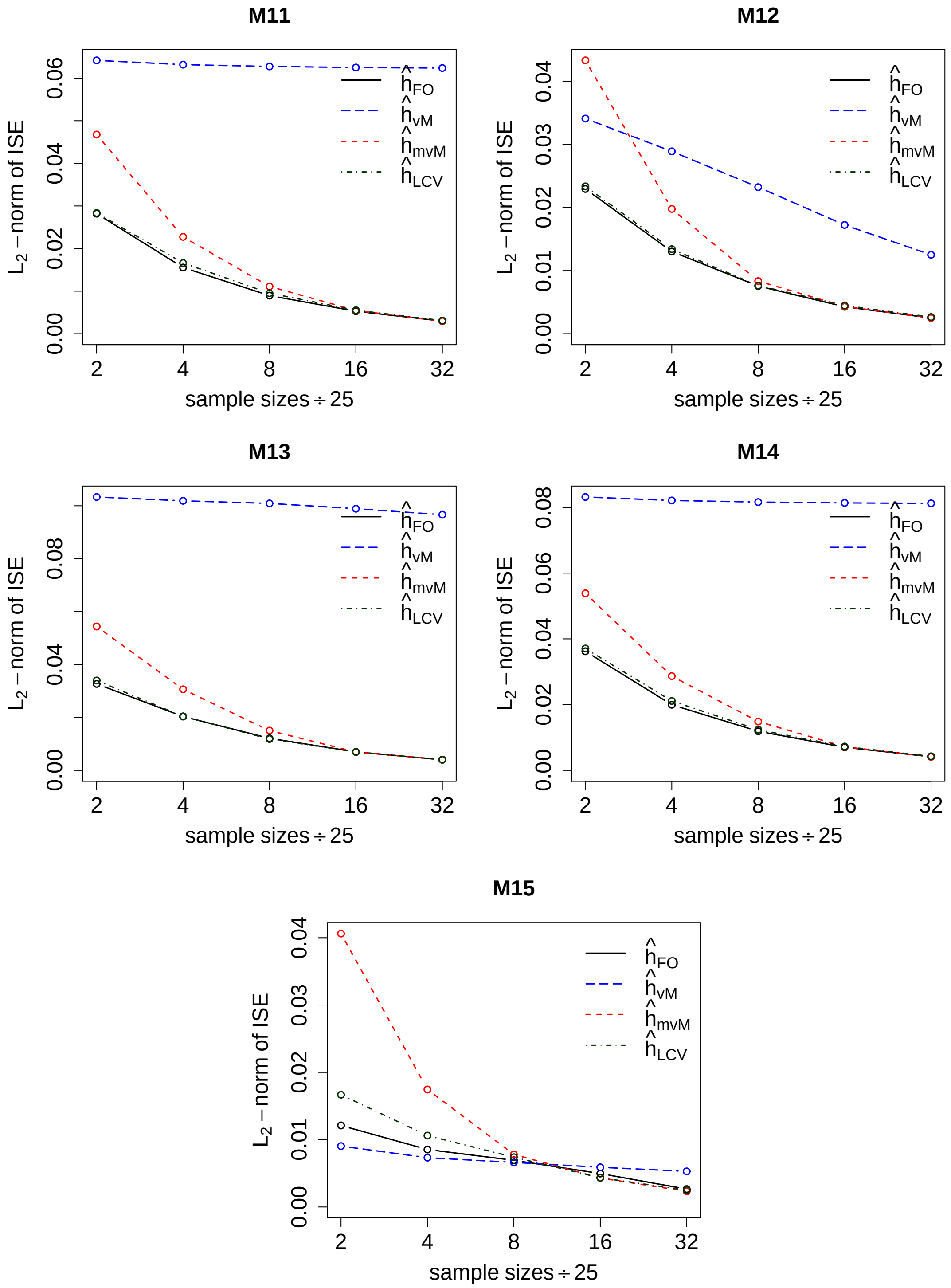

FIGURE 4. *Empirical L_2–norm of* $\mathrm{ISE}(f;\hat{f}_{\mathrm{HB}},\hat{h},n)$ *associated to the bandwidths* $\hat{h}_{\mathrm{FO}}$, $\hat{h}_{\mathrm{vM}}$, $\hat{h}_{\mathrm{mvM}}$ *and* $\hat{h}_{\mathrm{LCV}}$, *for circular density models* 11 *to* 15. *The number of replications is* 500.

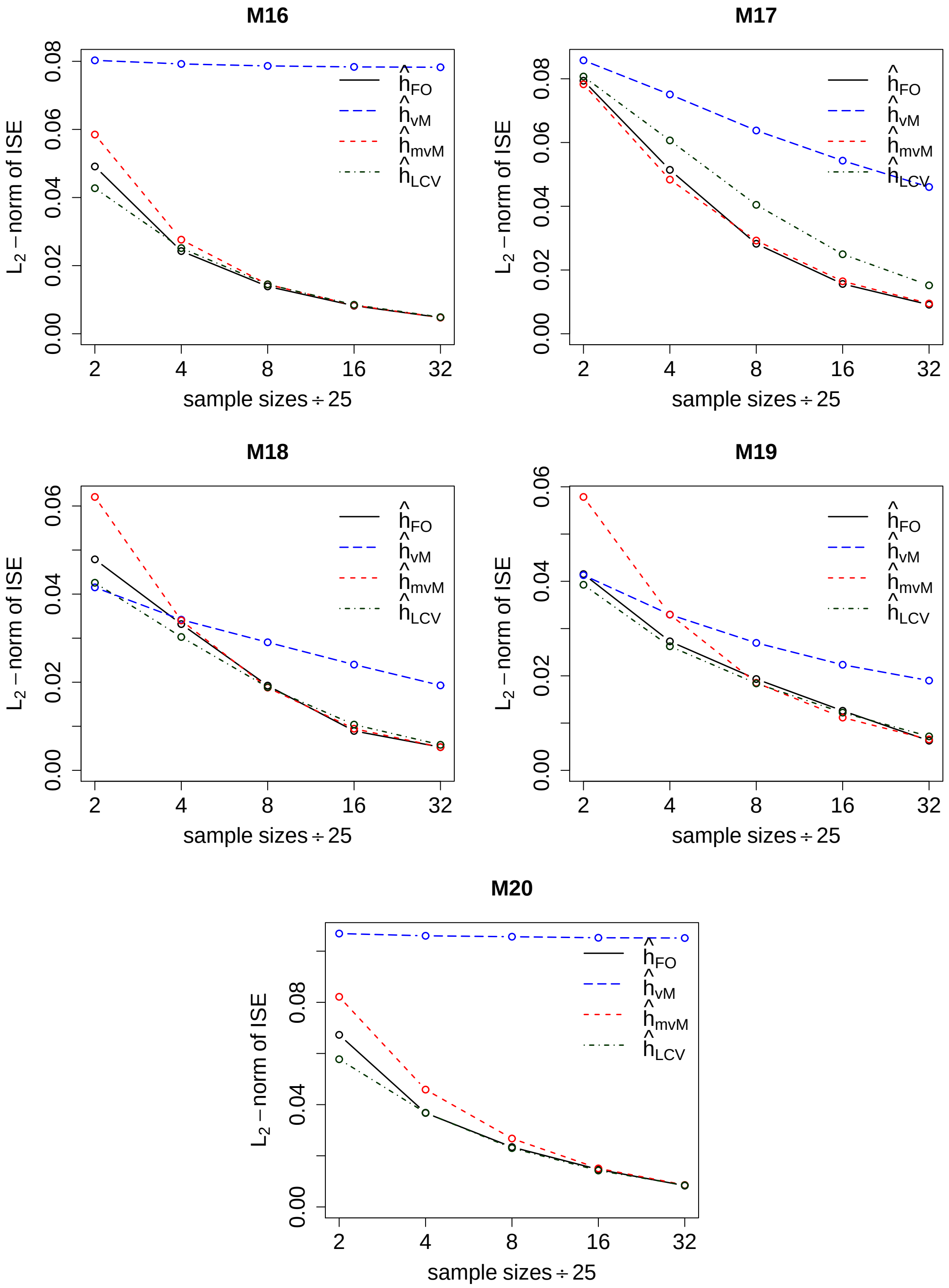

FIGURE 5. *Empirical L_2–norm of* $\mathrm{ISE}(f; \hat{f}_{\mathrm{HB}}, \hat{h}, n)$ *associated to the bandwidths* $\hat{h}_{\mathrm{FO}}$, $\hat{h}_{\mathrm{vM}}$, $\hat{h}_{\mathrm{mvM}}$ *and* $\hat{h}_{\mathrm{LCV}}$, *for circular density models* 16 *to* 20. *The number of replications is* 500.

that it shows a poor behaviour for almost all the considered test models when the sample size is small. Its performance improves significantly for moderate or large sample sizes, and only in this case this bandwidth selector should be used in practice. With the exceptions of models M5 and M17 whose densities present a single strong peak, 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, 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 $\hat{h}_{\mathrm{FO}}$, $\hat{h}_{\mathrm{vM}}$, $\hat{h}_{\mathrm{mvM}}$ and $\hat{h}_{\mathrm{LSC}}$, are used. For comparison purposes, after the values obtained for the bandwidths $\hat{h}_{\mathrm{vM}}$ and $\hat{h}_{\mathrm{mvM}}$ we also indicate, between square brackets, the bandwidths generated by the functions $(\mathrm{bw.rt})^{-1/2}$ and $(\mathrm{bw.pi})^{-1/2}$ from the 'NPCirc' package in R.

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 $\hat{h}_{\mathrm{mvM}} = 0.359$ [0.380] ($\hat{M} = 2$) yield identical bandwidths, while larger bandwidths are produced by $\hat{h}_{\mathrm{vM}} = 0.442$ [0.594] and $\hat{h}_{\mathrm{LSC}} = 0.508$. As we can see from Figure 6 the different smoothing parameters provide similar density estimates. 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, p. 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

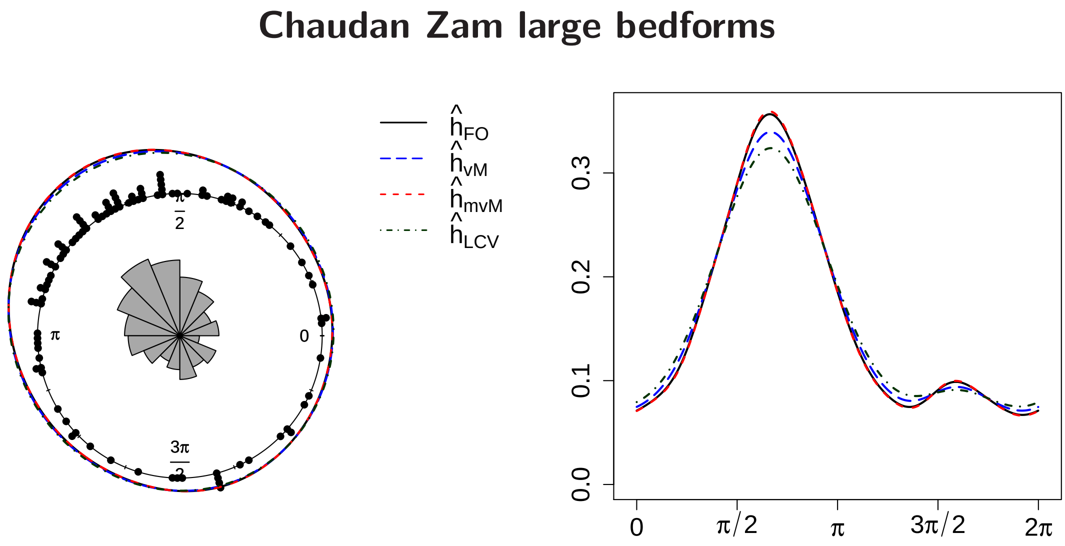

FIGURE 6. *Kernel density estimates for the cross-bed density by using the bandwidth selectors* $\hat{h}_{\mathrm{FO}}$, $\hat{h}_{\mathrm{vM}}$, $\hat{h}_{\mathrm{mvM}}$ *and* $\hat{h}_{\mathrm{LSC}}$.

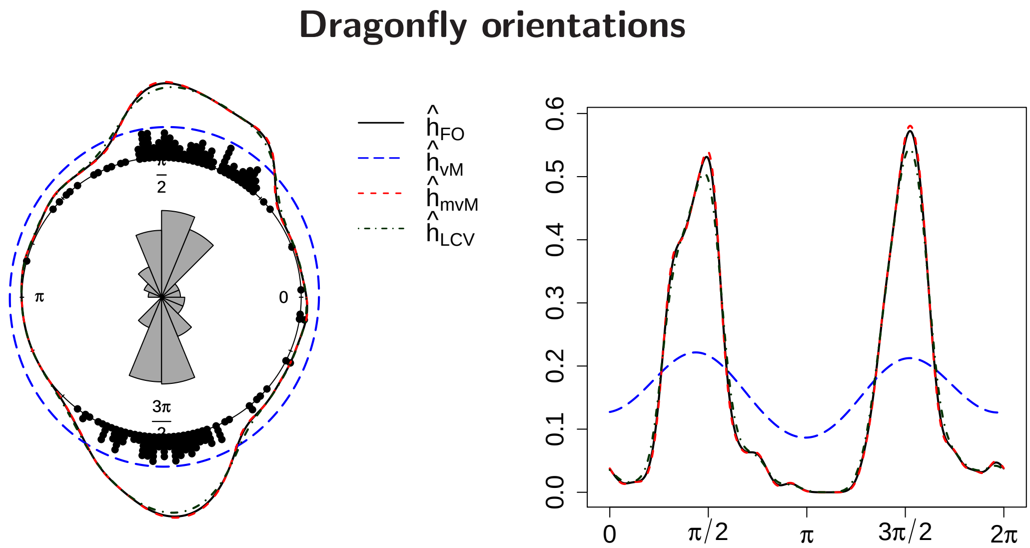

FIGURE 7. *Kernel density estimates for the dragonflies orientation density by using the bandwidth selectors* $\hat{h}_{\mathrm{FO}}$, $\hat{h}_{\mathrm{vM}}$, $\hat{h}_{\mathrm{mvM}}$ *and* $\hat{h}_{\mathrm{LSC}}$.

bimodal. As for the previous data set the bandwidth selectors $\hat{h}_{\mathrm{FO}} = 0.136$ and $\hat{h}_{\mathrm{mvM}} = 0.126$ [0.127] ($\hat{M} = 4$) yield identical bandwidths. Although a slightly larger bandwidth is produced by the likelihood cross-validation selector $\hat{h}_{\mathrm{LSC}} = 0.168$, the different kernel density estimates corresponding to these three bandwidths 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 $\hat{h}_{\text{vM}} = 0.778$ [1.479]. However, based on the simulation results for the circular density with opposite modes M7 obtained in the previous section (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 kernel density estimators for circular data presented in this paper enabled us to derive explicit expressions for the kernel density estimator asymptotic optimal bandwidth, and to propose a Fourier series-based plug-in approach for kernel density bandwidth selection. The theoretical properties established for the new bandwidth selector method, not shared by other existing 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 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

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

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

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

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 (14), from classic arguments we get

$$\sup_{\theta \in [0, 2\pi[} \left| \mathrm{E} \hat{f}_n(\theta) - f(\theta) - \frac{1}{2} \beta(\delta_n) f''(\theta) \right| \leq \frac{C}{6} \int_{-\pi}^{\pi} |y|^{2+\alpha} \delta_n(y) dy.$$

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

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

Batschelet, E. (1981). *Circular statistics in biology.* London: Academic Press.

Bosq, D., Lecoutre, J.-P. (1987). *Théorie de l'estimation fonctionnelle.* Paris: Economica.

Deheuvels, P. (1977). Estimation non paramétrique de la densité par histogrammes généralisés *Rev. Statist. Appl.* 25, 3, 5–42.

Deheuvels, P., Hominal, P. (1980). Estimation automatique de la densité. *Rev. Statist. Appl.* 28, 25–55.

Di Marzio, M, Panzera, A., Taylor, C.C. (2009). Local polynomial regression for circular predictors. *Statis. Probab. Lett.* 79, 2066–2075.

Epanechnikov, V.A. (1969). Nonparametric estimation of a multivariate probability density. *Theory Probab. Appl.* 14, 153–158.

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

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.

Hart, J.D. (1985). On the choice of a truncation point in Fourier series density estimation. *J. Stat. Comput. Simul.* 21, 95–116.

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

Kronmal, R., Tarter, M. (1968). The estimation of probability densities and cumulatives by Fourier series methods. *J. Amer. Statist. Assoc.* 63, 925–952.

Laurent, B. (1997). Estimation of integral functionals of a density and its derivatives. *Bernoulli* 3, 181–211.

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

Nadaraya, E.A. (1974). On the integral mean square error of some nonparametric estimates for the density function. *Theory Probab. Appl.* 19, 133–141.

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.

Oliveira, M., Crujeiras, R.M., Rodríguez-Casal, A. (2015). 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. *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. (2011). Fourier series based direct plug-in bandwidth selectors for kernel density estimation. *J. Nonparametr. Stat.* 23, 533–545.

Tenreiro, C. (2020). Bandwidth selection for kernel density estimation: a Hermite series-based direct plug-in approach. *J. Stat. Comput. Simul.* 90, 3433–3453.

Walter, G., Blum, J. (1979). Probability density estimation using delta sequences. *Ann. Statist.* 7, 328–340.

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

Watson, G.S., Leadbetter, M.R. (1964). Hazard analysis II *Sankhya Ser. A* 26, 101–116.

Woodroofe, M. (1970). On choosing a delta-sequence. *Ann. Math. Statist.* 41, 1665–1671.

Carlos Tenreiro
CMUC, Department of Mathematics, University of Coimbra, 3001–501 Coimbra, Portugal
E-mail address: tenreiro@mat.uc.pt