Adaptive density estimation on bounded domains under mixing conditions

Karine Bertin[*],[†]

CIMFAV-INGEMAT, Universidad de Valparaíso, General Cruz 222, Valparaíso, Chile
e-mail: karine.bertin@uv.cl

Nicolas Klutchnikoff[†]

Univ Rennes, CNRS, IRMAR – UMR 6625, F-35000 Rennes, France
e-mail: nicolas.klutchnikoff@univ-rennes2.fr

Jose R. Léon[‡]

IMERL, Universidad de la República, Montevideo, Uruguay, Escuela de Matemática,
Facultad de Ciencias, Universidad Central de Venezuela, Caracas, Venezuela
e-mail: rlramos@fing.edu.uy

and

Clémentine Prieur[‡]

Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP*, LJK, F-38000 Grenoble, France,
*Institute of Engineering Univ. Grenoble Alpes
e-mail: clementine.prieur@univ-grenoble-alpes.fr

Abstract: In this article, we propose a new adaptive estimator for multivariate density functions defined on a bounded domain in the framework of multivariate mixing processes. Several procedures have been proposed in the literature to tackle the boundary bias issue encountered using classical kernel estimators. Most of them are designed to work in dimension $d = 1$ or on the unit d-dimensional hypercube. We extend such results to more general bounded domains such as simple polygons or regular domains that satisfy a rolling condition. We introduce a specific family of kernel-type estimators devoid of boundary bias. We then propose a data-driven Goldenshluger and Lepski type procedure to jointly select a kernel and a bandwidth. We prove the optimality of our procedure in the adaptive framework, stating an oracle-type inequality. We illustrate the good behavior of our new class of estimators on simulated data. Finally, we apply our procedure to a real dataset.

MSC 2010 subject classifications: Primary 62G07; secondary 60G10, 60K35.
Keywords and phrases: Compact supported density estimation, multivariate analysis, mixing stochastic processes.

Contents

1. Introduction

In this paper we study the classical problem of the estimation of a density function $f : \mathcal{D} \to \mathbb{R}$ where $\mathcal{D} \subset \mathbb{R}^d$ from the observation of n identically distributed random variables $X_1, \ldots, X_n$ not necessarily independent. In several modern applications, $\mathcal{D}$ is a known bounded domain whose shape can be quite complex (e.g. geographical region). Moreover these data often present a dependence structure that has to be taken into account. Our main objective is to propose a new kernel-type estimation procedure to address these two points and to establish results ensuring the optimality of this procedure from a theoretical point of view.

It is well known that classical kernel-type estimators present a severe bias when the density function f does not vanish near the boundary of $\mathcal{D}$. Several procedures have been proposed in the literature to tackle this issue in the univariate setting. Schuster (1985), Silverman (1986) and Cline and Hart (1991)

2200 K. Bertin et al.

studied the reflection of the data near the boundary. Marron and Ruppert (1994) proposed a previous transformation of the data. Müller (1991), Lejeune and Sarda (1992), Jones (1993), Chen (1999) and Botev et al. (2010) proposed to construct kernels which take into account the shape of the support of the density. These procedures can be easily adapted to the hypercube $\mathcal{D} = [0,1]^d$ using tensorization methods. Nevertheless, their generalization to more complex domains has not been studied and seems more tedious. Only few methods are designed to work with more general domains $\mathcal{D} \subset \mathbb{R}^d$. In this context, Müller and Stadtmüller (1999) proposed a generic method to construct kernels (of arbitrary orders) whose shape depends on both the estimation point $x \in \mathcal{D}$ and the bandwidth $h > 0$. These kernels are solutions, for each x and h, of tricky continuous least squares minimization problems. Marshall and Hazelton (2010) proposed an alternative approach dedicated to the case $d = 2$ that allows one to construct more tractable procedures. Few papers study in-depth the theoretical properties of these multivariate procedures. In the context of independent and identically distributed observations and twice differentiable density functions, we point out Bouezmarni and Rombouts (2010) who study the behavior of Beta kernels with a cross-validation selection procedure and Marshall and Hazelton (2010) who study the pointwise behavior of their estimators for a fixed bandwidth. The results stated in Müller and Stadtmüller (1999) could be used to prove pointwise minimax results over arbitrary isotropic Hölder classes. To our best knowledge only Bertin et al. (2018) proved adaptive results for integrated risks over $\mathcal{D} = [0,1]^d$ (in the sense that a single estimation procedure achieves the minimax rate of convergence over a large scale of regularity classes). They introduced a new family of kernel density estimators that do not suffer from the boundary bias problem and they proposed a data-driven procedure based on the Goldenshluger and Lepski (see Goldenshluger and Lepski, 2014) approach that jointly selects a kernel and a bandwidth.

Other classes of methods that consider density estimation on bounded domains are that based on selection models of histograms or piecewise polynomials (Castellan, 2000, 2003; Birgé and Rozenholc, 2006; Akakpo, 2012) using approach developed by Barron et al. (1999).

These papers studied above all estimation on the unit hypercube $[0,1]^d \subset \mathbb{R}^d$ in an independent framework. In the present paper, we aim at considering more general bounded domains of $\mathbb{R}^d$ such as the disk, simple polygons — that may be used to define geographical regions — or more regular domains satisfying the *rolling condition* (see Arias-Castro and Rodríguez-Casal, 2017). We assume that the observations are extracted from a stationary β-mixing process. In this framework, we introduce a new family of kernel density estimators, and we propose a data-driven selection procedure inspired by Goldenshluger and Lepski (2014) and Bertin et al. (2018) to jointly select a kernel and a bandwidth. Our main contribution consists in the statement of an oracle-type inequality in $\mathbb{L}_2-$ norm that allows us to prove the adaptivity (in the minimax sense) of our procedure over a large scale of Hölder classes without bound on the smoothness parameter. We also conduct simulation studies on the disk in the independent case and the dependent case where observations come from a diffusion process

with reflection. This type of process is of particular interest since it can be used to model population dynamics in bounded geographical areas (see Cholaquidis et al., 2016). Within this context, we consider a real data set corresponding to the trajectories of elephants in Hwange National Park in Zimbabwe.

The paper is organized as follows. We first introduce the statistical framework in Section 2. Section 3 is devoted to the description of our new estimation procedure and its theoretical properties are stated in Section 4. Simulation studies are shown in Section 5, as far as the application to the real data in Section 6. The proofs are postponed to Section 7.

2. Statistical framework

In what follows, we consider a strongly stationary and β-mixing process $\mathbb{X} = (X_i : i \in \mathbb{Z})$ that lies into a bounded domain $\mathcal{D} \subset \mathbb{R}^d$. We assume that the common marginal distribution of the random variables X_i is absolutely continuous with respect to the Lebesgue measure restricted to $\mathcal{D}$ and we denote by $f \colon \mathcal{D} \to \mathbb{R}$ the density of this distribution. We aim at finding an accurate estimation procedure for f based on the observations $X_1, \ldots, X_n$, where $n \in \mathbb{N}$.

In the rest of this section, we present the main assumptions we make on the law of the process $\mathbb{X}$ and on the geometry of the domain $\mathcal{D}$. We also present the adaptive minimax framework used to measure the statistical performances of the estimators.

2.1. Assumptions on the law of the process

The assumptions on the process $\mathbb{X}$ are divided into two parts: the assumptions on the marginal density f on the one hand and the assumptions on the dependence structure of the process on the other hand.

To state the assumptions on the marginal density, we recall the definition of a Hölder ball on the domain $\mathcal{D}$. Let γ and L be two positive numbers. A function $f \colon \mathcal{D} \to \mathbb{R}$ belongs to the Hölder class $\mathcal{H}_\mathcal{D}(\gamma, L)$ if the following conditions are fulfilled:

i) The partial derivatives $D^\alpha f = \partial^{|\alpha|} f / (\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d})$ exist for any $\alpha \in (\mathbb{N} \cup \{0\})^d$ such that $|\alpha| = \alpha_1 + \cdots + \alpha_d \leq \lfloor \gamma \rfloor$ where $\lfloor \gamma \rfloor = \max\{\ell \in \mathbb{N} \cup \{0\} \colon \ell < \gamma\}$.

ii) For any x, y in $\mathcal{D}$, $\sum_{|\alpha| = \lfloor \gamma \rfloor} |D^\alpha f(y) - D^\alpha f(x)| \leq L |y - x|_1^{\gamma - \lfloor \gamma \rfloor}$ where $|u|_p = (\sum_{i=1}^d |u_i|^p)^{1/p}$ if $1 \leq p < +\infty$, $|u|_\infty = \max\{|u_i| \colon i = 1, \ldots, d\}$.

Assumption 1. *Set $f_\infty > 0$. The sup-norm $\|f\|_\infty = \sup_{x \in \mathcal{D}} |f(x)|$, is less than or equal to f_∞.*

The absolute regularity (or β-mixing) condition of a process was introduced by Volkonskii and Rozanov (1959) and attributed there to Kolmogorov. For the convenience of the reader we recall the definition of the β-mixing coefficients of

the strictly stationary process $\mathbb{X}$. For each $k \geq 1$, we define:

$$\beta(k) = \sup \frac{1}{2} \sum_{i \in I} \sum_{j \in J} |\mathbf{P}(U_i \cap V_j) - \mathbf{P}(U_i)\mathbf{P}(V_j)|,$$

where the supremum is taken over all pairs of finite partitions $\{U_i : i \in I\}$ and $\{V_j : j \in J\}$ of the probability space Ω which are respectively measurable with respect to $\sigma(X_s : s \leq 0)$ and $\sigma(X_s : s \geq k)$.

Assumption 2. *Let c be a positive number and set $0 < \rho < 1$. We assume that the process $\mathbb{X}$ is strictly stationary and β-mixing at a geometric rate. More precisely, for any $k \geq 1$ we have $\beta(k) \leq c\rho^k$.*

Assumption 3. *Set $f_\infty > 0$. For any $k \geq 1$ the distribution of the random pair (X_1, X_{k+1}) admits a density $f_k \colon \mathcal{D}^2 \to \mathbb{R}$ (with respect to the Lebesgue measure restricted to $\mathcal{D}^2$) such that $\|f_k\|_\infty \leq f_\infty$.*

Without loss of generality we assume that the bounds f_∞ that appear in Assumptions 1 and 3 are the same if the two assumptions hold simultaneousely.

2.2. Geometric assumptions on the domain

In this section, we first state technical assumptions on the domain $\mathcal{D}$ and we offer some examples that satisfy these conditions.

Assumption 4. *Set $R > 0$. The domain $\mathcal{D} \neq \emptyset$ is a bounded open connected set such that, for any $x \in \mathcal{D}$, $|x|_\infty \leq R$.*

Remark that, since our goal is to consider the estimation on bounded domain, the existence of $R > 0$ is not a restrictive condition. Assuming that $\mathcal{D}$ is connected is also not restrictive since the same estimation procedure could be applied on each connected component. Finally $\mathcal{D}$ is assumed to be open to ensure that the ambiant dimension d is the correct one.

Assumption 5. *There exist $0 < r < 1$ and a finite family $\mathcal{A} = \{\mathbf{A}_1, \ldots, \mathbf{A}_\kappa\}$ of distinct elements in $GL_d(\mathbb{R})$ such that:*

i) *For any $j = 1, \ldots, \kappa$, $|\det \mathbf{A}_j| = 1$.*
ii) *For any $x \in \mathcal{D}$ there exists $A_x \in \mathcal{A}$ such that $x + A_x^{-1}([0,r]^d) \subset \mathcal{D}$.*

While Assumption 5 seems quite restrictive, it is satisfied by several domains as illustrated in the following examples. The first one was considered in Bertin et al. (2018) for independent data.

Example 1 (Hypercubes). *We consider the case where $\mathcal{D} = (0,1)^d$ and we define for $u \in \mathcal{D}$ and $x \in \mathcal{D}$: $A_x(u) = (\sigma(x_1)u_1, \ldots, \sigma(x_d)u_d)$ where $\sigma(x) = 1 - 2\mathbf{I}_{(1/2,1)}(x)$ for $x \in (0,1)$. In this case we have $r = 1/2$ and $\kappa = 2^d$.*

To state the second example, we denote by $\mathbb{D}_r = \{x \in \mathbb{R}^d : |x|_2 \leq r\}$ the Euclidean ball with radius $r > 0$. Note that this example is of particular interest since it is used throughout our simulation study in Section 5.

Example 2 (Disk). *For simplicity we only consider the case $d = 2$. We assume that $\mathcal{D} = \mathbb{D}_1$. For $k = 1, \ldots, 6$ we define:*

$$\mathbf{A}_k = \begin{pmatrix} \cos(\theta_k) & -\sin(\theta_k) \\ \sin(\theta_k) & \cos(\theta_k) \end{pmatrix}$$

where $\theta_k = -3\pi/4 - k\pi/3$. Now, for any $x \in \mathcal{D} \setminus \{0\}$ we identify $x/|x|_2 \in \mathbb{S}^1$ with a real number in $[0, 2\pi)$. Now, since

$$[0, 2\pi) = \bigcup_{k=1}^{6} I_k \quad where \quad I_k = \left[-\frac{\pi}{6} + k\frac{\pi}{3}, \frac{\pi}{6} + k\frac{\pi}{3} \right),$$

we define $A_x = \mathbf{A}_k$ if k is such that $x/|x|_2 \in I_k$ and $A_0 = \mathbf{A}_1$. Assumption 5 is satisfied with $\kappa = 6$ and $r = 1/4$.

However, generic classes of open subsets of $\mathbb{R}^2$ can be proven to satisfy Assumption 5 as in the two following examples (See Appendices A and B for the proofs).

Example 3 (Rolling conditions). *Set $r_0 > 0$. The domain $\mathcal{D}$ is called r_0-regular if, for any $0 < r \leq r_0$, the ball $\mathbb{D}_r$ rolls freely in both $\mathcal{D}$ and $\mathcal{D}^c$. That is, for each $a \in \partial\mathcal{D}$, there exists x_a^r and y_a^r in $\mathbb{R}^d$ such that:*

$$a \in \left(x_a^r + \mathbb{D}_r \right) \subseteq \overline{\mathcal{D}} \quad and \quad a \in \left(y_a^r + \mathbb{D}_r \right) \subseteq \mathcal{D}^c.$$

Such regularity condition on $\mathcal{D}$ is well-known and widely used in statistics (see Arias-Castro and Rodríguez-Casal, 2017, and references therein). The following figure illustrates this condition:
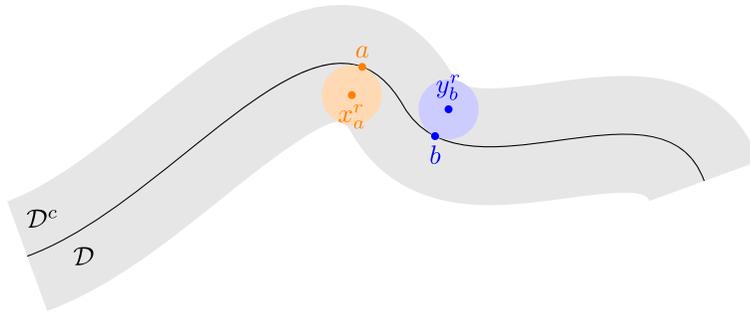
Finally, since in practical situations the boundary of a domain can be approximated by a simple polygonal path (as for geographical areas), the following example (see Appendix B for more details) seems to be of prime interest.

Example 4 (Simple polygons). *The interior of any simple polygon satisfies Assumption 5.*

2.3. Framework

Under Assumptions 1 and 4, the marginal density f belongs to the space of squared integrable functions that map $\mathcal{D}$ into $\mathbb{R}$. This set, denoted by $\mathbb{L}_2(\mathcal{D})$ is endowed with its natural Hilbertian norm:

$$\|g\|_2 = \left(\int_{\mathcal{D}} g^2(u) du \right)^{1/2}.$$

To measure the performance of an estimator, we consider its risk defined by:

$$R_n(\tilde{f}, f) = \left(\mathbf{E} \|\tilde{f} - f\|_2^2 \right)^{1/2}.$$

Let $\mathbb{F}$ be a subset of $\mathbb{L}_2(\mathcal{D})$. The maximal risk of $\tilde{f}$ over $\mathbb{F}$ is defined by:

$$R_n(\tilde{f}, \mathbb{F}) = \sup_{f \in \mathbb{F}} R_n(\tilde{f}, f),$$

whereas the minimax risk over $\mathbb{F}$ (see Tsybakov, 2009) is:

$$\phi_n(\mathbb{F}) = \inf_{\tilde{f}} R_n(\tilde{f}, \mathbb{F})$$

where the infimum is taken over all the estimators. For the problem of multivariate density estimation, the minimax rate of convergence over $\mathcal{H}_{\mathcal{D}}(\gamma, L)$ is given by $\phi_n(\gamma, L) = n^{-\gamma/(2\gamma+d)}$ (see Ibragimov and Hasminskii, 1980; Delyon and Juditsky, 1996). An estimator whose maximal risk is asymptotically bounded, up to a multiplicative factor, by $\phi_n(\mathbb{F})$ is called minimax over $\mathbb{F}$. Such an estimator is well-adapted to the estimation over $\mathbb{F}$ but it can perform poorly over another functional space. The problem of adaptive estimation consists in finding a single estimation procedure that is simultaneously minimax over a scale of functional classes. More precisely, given a family $\{\mathbb{F}_\lambda : \lambda \in \Lambda\}$ of subsets of $\mathbb{L}_2(\mathcal{D})$, the goal is to construct f^* such that $R_n(f^*, \mathbb{F}_\lambda)$ is asymptotically bounded, up to a multiplicative constant, by $\phi_n(\mathbb{F}_\lambda)$ for any $\lambda \in \Lambda$. One of the main tools to prove that an estimation procedure is adaptive over a scale of functional classes is to prove an oracle-type inequality that guarantees that this procedure performs almost as well as the best estimator in a rich family of estimators. Ideally, we would like to have an inequality of the following form:

$$R_n(f^*, f) \le \inf_{\eta \in H} R_n(\hat{f}_\eta, f), \tag{1}$$

where $\{\hat{f}_\eta : \eta \in H\}$ is a family of estimators well-adapted to our problem: for any $\lambda \in \Lambda$, there exists $\eta(\lambda)$ such that $\hat{f}_{\eta(\lambda)}$ is minimax over $\mathbb{F}_\lambda$. However, in many situations, (1) is relaxed and we prove a weaker inequality of the type:

$$R_n(f^*, f) \le \Upsilon_1 \inf_{\eta \in H} R_n^*(f, \eta) + \Upsilon_2 n^{-1/2}, \tag{2}$$

where Υ_1 and Υ_2 are two positive constants and $R_n^*(f, \eta)$ is an appropriate quantity to be determined that can be viewed as a tight upper bound on $R_n(\hat{f}_\eta, f)$. Inequalities of the form (2) are called oracle-type inequalities.

3. Statistical procedure

We propose in Section 3.1 a specific family of kernel-type estimators $(\hat{f}_\ell)$ which can tackle with the boundary bias problem encountered using classical kernel estimators (see Bertin et al., 2018, and references therein for more details) and that is well-adapted to our problem. The construction of this family is linked with geometrical assumptions on the domain $\mathcal{D}$. A selection procedure associated to this family is given in Section 3.2.

3.1. Family of estimators

A function $K : \mathbb{R} \to \mathbb{R}$ is called a univariate kernel if the support of K is included into $[0,1]$, $\|K\|_\infty < +\infty$ and $\int_0^1 K(u)du = 1$. Moreover, we say that K is a kernel of order $m \in \mathbb{N} \cup \{0\}$ if

$$\int_0^1 K(u)u^p du = \delta_{0,p}, \quad \text{for any } 0 \le p \le m.$$

For any bandwidth $h > 0$ and univariate kernel K, the normalized multivariate kernel $\mathbf{K}_h$ is defined by:

$$\mathbf{K}_h(s) = h^{-d} \prod_{i=1}^{d} K(h^{-1}s_i), \qquad s \in \mathbb{R}^d.$$

Equipped with these notations we note that, under Assumption 5 and for h small enough, for any $x \in \mathcal{D}$, $\mathbf{K}_h \circ A_x(u - x) = 0$ as soon as $u \notin \mathcal{D}$. This property implies that there is no loss of mass near the boundary (as in the classical convolution kernel). This leads us to consider the following estimators:

$$\hat{f}_{K,h}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{K}_h \circ A_x(X_i - x), \qquad x \in \mathcal{D}. \tag{3}$$

In the following proposition, we study the bias term and the variance term of the estimators $\hat{f}_{K,h}$ and show that they reach minimax rate of convergence over Hölder classes $\mathcal{H}_\mathcal{D}(\gamma, L)$.

Proposition 1. *Assume that Assumptions 1, 2, 4 and 5 are fulfilled. Set $\gamma > 0$ and $L > 0$. Assume that $f \in \mathcal{H}_\mathcal{D}(\gamma, L)$. Let K be a kernel of order greater than or equal to $\lfloor \gamma \rfloor$. Then, there exist two absolute constants $C(R, \gamma)$ and $C(\rho, c, f_\infty)$ such that for any $h > 0$ we have:*

$$\|\mathbf{E}\hat{f}_{K,h} - f\|_2 \le C(R, \gamma)L\|K\|_\infty^d h^\gamma \tag{4}$$

and

$$\mathbf{E}\|\hat{f}_{K,h} - \mathbf{E}\hat{f}_{K,h}\|_2^2 \le \frac{4\kappa\|K\|_2^{2d}}{nh^d} \frac{c}{1-\rho}. \tag{5}$$

With the additional Assumption 3, we get

$$\mathbf{E}\|\hat{f}_{K,h} - \mathbf{E}\hat{f}_{K,h}\|_2^2 \leq \frac{\kappa\|K\|_2^{2d}}{nh^d}\left(1 + C(c, \rho, f_\infty)h^{d/2}\right) \tag{6}$$

Moreover, in both cases, taking $h = n^{-1/(2\gamma+d)}$ the estimator $\hat{f}_{K,h}$ reaches the minimax rate $\phi_n(\gamma, L) \asymp n^{-\gamma/(2\gamma+d)}$ over $\mathcal{H}_\mathcal{D}(\gamma, L)$.

Even if the estimators $\hat{f}_{K,h}$ are well-adapted to our framework, selecting in a data driven way a kernel K and a bandwidth h is difficult from both theoretical and practical points of view. As a consequence, we construct a one-parameter subfamily which consists of predefined well-chosen pairs of kernels and bandwidths. To do so, we define, for any $m \in \mathbb{N} \cup \{0\}$, the kernel

$$K_m(u) = \sum_{r=0}^{m} \varphi_r(0)\varphi_r(u), \quad u \in [0, 1], \tag{7}$$

where $\varphi_k(u) = \sqrt{2k+1}L_k(2u-1)$ and L_k is the Legendre Polynomial of degree k on $[-1, 1]$. We also consider, for any $\ell \in \mathbb{N}$:

$$h(\ell) = e^{-\ell} \quad \text{and} \quad m_n(\ell) = \left[\frac{\log(n)}{2\ell} + \frac{1}{2}\right]$$

where $[\cdot]$ denotes the integer part. Using the notation introduced in (3) we define, for any $\ell \in \mathbb{N}$ the estimator:

$$\hat{f}_\ell = \hat{f}_{K_{m_n(\ell)}, h(\ell)}.$$

To obtain a finite collection of estimators we impose an additional restriction on $h(\ell)$ by considering only $\ell \in \mathcal{L}_n$ where

$$\mathcal{L}_n = \{\ell \in \mathbb{N} \colon \underline{h}_n < h(\ell) < \overline{h}_n\}.$$

Here, the bandwidths $\underline{h}_n$ and $\overline{h}_n$ are defined, for given $c_1 > 0$ and $c_2 > 0$, by

$$\overline{h}_n = (\log n)^{-2(1+c_1)/d} \quad \text{and} \quad n\underline{h}_n^d = \left(\log(n)\right)^{c_2}.$$

We have the following proposition about the family $(\hat{f}_\ell)$.

Proposition 2. *Assume that Assumptions 1, 2, 4 and 5 are fulfilled. Set $\gamma > 0$ and define $\ell_\gamma = [(2\gamma + d)^{-1}\log n]$. The estimator $\hat{f}_{\ell_\gamma}$ reaches the minimax rate of convergence $\phi_n(\gamma, L)$ over $\mathcal{H}_\mathcal{D}(\gamma, L)$ for any $L > 0$.*

Proposition 2 is a direct consequence of Proposition 1 and of the fact that the order of the kernel of $\hat{f}_{\ell_\gamma}$ is larger than γ. Proposition 2 shows an interesting property of the estimators $\hat{f}_\ell$. When we choose adequately ℓ in order to obtain a bandwidth of order $n^{-1/(2\gamma+d)}$, we automatically use a kernel of order larger than γ, which allows one to obtain the optimal rate of convergence on $\mathcal{H}_\mathcal{D}(\gamma, L)$.

Remark 1. *Note that the kernel K_m is of order m and satisfies*

$$\|K_m\|_\infty \le A(m+1)^B \tag{8}$$

with $A = 1$ and $B = 2$. More precisely it can be proven that $\|K_m\|_2 = m+1$, $\|K_m\|_\infty = (m+1)^2$ and that K_m is the kernel of order m with smaller L_2 norm. See Lemma 2 in Bertin et al. (2018) for more details. More generally, any sequence of kernels K_m of order m satisfying (8) with some constants A and B can be used instead of the family defined by (7).

3.2. Selection rule

Let $\tau > 0$. For any $\ell, \ell' \in \mathcal{L}_n$ we define the following majorants:

$$\widehat{M}(\ell) = \sqrt{2}\,\frac{\hat{\Gamma}\big(K_{m_n(\ell)}, h(\ell)\big) + \tau \|K_{m_n(\ell)}\|_2^d}{\sqrt{nh^d(\ell)}}$$

and

$$\widehat{M}(\ell, \ell') = \widehat{M}(\ell') + \widehat{M}(\ell' \wedge \ell),$$

where $\ell \wedge \ell'$ denotes the minimum between ℓ and ℓ' and $\hat{\Gamma}(K, h)$ is defined by

$$\hat{\Gamma}(K, h) = h^{d/2} \left(\int_{\mathcal{D}} \frac{1}{n} \sum_{i=1}^{n} \left(\mathbf{K}_h \circ A_x(X_i - x) \right)^2 dx \right)^{1/2}.$$

Now, we define:

$$\widehat{B}(\ell) = \max_{\ell' \in \mathcal{L}_n} \left\{ \|\hat{f}_{\ell \wedge \ell'} - \hat{f}_{\ell'}\|_2 - \widehat{M}(\ell, \ell') \right\}_+$$

with $x_+ = \max(x, 0)$ denotes the positive part of x. The final estimator, $\widehat{f}$ is then defined by

$$\widehat{f} = \hat{f}_{\widehat{\ell}} \qquad \text{with} \qquad \widehat{\ell} = \underset{\ell \in \mathcal{L}_n}{\arg\min} \left(\widehat{B}(\ell) + \widehat{M}(\ell) \right). \tag{9}$$

Remark 2. *This selection rule is an adaptation of the so-called Goldenshluger-Lepski (GL) method which consists in selecting, in a data-driven way, an estimator that realizes the trade-off (9) beetwen $\widehat{B}$ and $\widehat{M}$, estimators of respectively the bias term and the stochastic term. Finding tight majorants is the key-point of this procedure. Let us briefly comment on the form of the majorant $\widehat{M}(\ell)$. The ideal majorant would be*

$$\overline{M}(\ell) = \frac{\Gamma\big(K_{m_n(\ell)}, h(\ell)\big)}{\sqrt{nh^d(\ell)}},$$

where

$$\Gamma(K, h) = h^{d/2} \left(\int_{\mathcal{D}} \mathbf{E} \left(\mathbf{K}_h \circ A_x \right)^2 (X_1 - x) dx \right)^{1/2}. \tag{10}$$

However the term $\Gamma\big(K_{m_n(\ell)}, h(\ell)\big)$ depends on the unknown density f and is bounded, see (14), by the deterministic constant $\sqrt{\kappa}\|K\|_2^d$ which can be rough in

some situations (in Example 1, $\kappa = 2^d$). Recall that this is due to the specific form of our boundary kernels. To circumvent this drawback we replace this quantity by a simple estimator $\hat{\Gamma}\big(K_{m_n(\ell)}, h(\ell)\big)$. For technical reasons that appear in the proof of Lemma *7* we add the small corrective term $\tau\|K\|_2^d$. Finally, the extra $\sqrt{2}$ factor allows us to take into account the dependence structure of the observations using classical Berbee coupling techniques (see Berbee, *1979*; Comte et al., *2017*) in the proofs of Theorem *3*. Note also that the final procedure depends on the parameters τ, c_1 and c_2 that can be chosen from a theoretical point of view as small as desired.

4. Main results

In the following, we state an oracle inequality and adaptive properties of the selection procedure $\hat{f}$.

Theorem 3 (Oracle-type inequality). *Assume Assumptions* *1, 2, 3, 4* *and* *5* *are fulfilled. We have*

$$R_n(\hat{f}, f) \leq \Upsilon_1 \inf_{\ell \in \mathcal{L}_n} \left(\max_{\ell' \geq \ell} \|\mathbf{E}\hat{f}_{\ell'} - f\|_2 + \frac{\|K_{m_n(\ell)}\|_2^d}{\sqrt{nh^d(\ell)}} \right) + \Upsilon_2 n^{-1/2}$$

with Υ_1, Υ_2 two positive constants depending on τ, c, ρ, f_∞, κ, c_1 and c_2.

Note that this oracle-type inequality is of the form (2) with:

$$R_n^*(f, \ell) = \max_{\ell' \geq \ell} \|\mathbf{E}\hat{f}_{\ell'} - f\|_2 + \frac{\|K_{m_n(\ell)}\|_2^d}{\sqrt{nh^d(\ell)}}$$

which is, up to a multiplicative constant, an upper bound of the risk of $\hat{f}_\ell$ since, using (5):

$$R_n(\hat{f}_\ell, f) = \left(\|\mathbf{E}\hat{f}_\ell - f\|_2^2 + \mathbf{E}\|\hat{f}_\ell - \mathbf{E}\hat{f}_\ell\|_2^2 \right)^{1/2}$$

$$\leq \|\mathbf{E}\hat{f}_\ell - f\|_2 + \left(\frac{4c\kappa}{1 - \rho} \right)^{1/2} \frac{\|K_{m_n(\ell)}\|_2^d}{\sqrt{nh^d(\ell)}}.$$

This allows us to prove that the procedure is adaptive over a large scale of Hölder spaces.

Theorem 4 (Adaptive estimation). *Assume Assumptions* *1, 2, 3, 4* *and* *5* *are fulfilled. For any $\gamma > 0$ and $L > 0$, we have:*

$$\limsup_{n \to +\infty} n^{\gamma/(2\gamma+d)} \sup_{f \in \mathcal{H}_\mathcal{D}(\gamma, L)} R_n(\hat{f}, f) < \Upsilon. \tag{11}$$

with Υ a positive constant depending on τ, c, ρ, f_∞, κ, c_1, c_2, L and γ.

The proof of this theorem relies on Proposition 2 as well as the fact that, if $f \in \mathcal{H}_\mathcal{D}(\gamma, L)$ for some $\gamma > 0$, then $\max_{\ell' \geq \ell_\gamma} \|\mathbf{E}\hat{f}_{\ell'} - f\|_2$ is a tight upper bound of $\|\mathbf{E}\hat{f}_{\ell_\gamma} - f\|_2$ (see the proof for more details).

Remark 3. *This result is obtained without any restriction on the smoothness parameter $\gamma > 0$. This follows from the simultaneous choice of a bandwidth and a kernel and differs from the usual bandwidth selection procedure where the kernel remains fixed. In fact, in the usual approach, the same selection procedure proposed in Section 3.2 with K fixed instead of $K_{m_n(\ell)}$ can be applied to choose a bandwidth $\hat{h}$ and to obtain an estimator $\hat{f}_{K,\hat{h}}$. It can be proved similarly that this estimator satisfies an oracle-type inequality and the same inequality* (11) *but with Υ depending additionally on the order M of the kernel K. In this usual approach, adaptation can be obtained on Hölder classes $\mathcal{H}_{\mathcal{D}}(\gamma, L)$ of smoothness γ with $0 < \gamma \leq M + 1$. In fact the order of the kernel imposes a maximal value on the smoothness parameter for adaptation. Note also that in this case to get rid of this condition on the smoothness parameter, one may take a kernel with order $\varphi(n)$ that increases with n. Such a kernel satisfies $\|K\|_2 \geq (\varphi(n) + 1)$ (see Remark* 1), *and this will affect the rate of convergence by an extra $\varphi(n)$-term.*

5. Simulation study

In this section, we study the performance of our procedure using simulated data in Sections 5.1 (in the independent framework) and 5.2 (for β-mixing data). More precisely, in Section 5.1, we aim at estimating several densities defined on the disk that exhibit various behaviors near the boundary of their support. In Section 5.2, we study the stationary density of a two dimensional reflected Langevin diffusion on the disk. In each situation, we study the accuracy of our procedure as well as usual kernel estimators, calculating empirical risks using $M = 500$ Monte-Carlo replications. In the following, we detail our simulation scheme and comment on the obtained results.

5.1. Densities on the disk

Simulation scheme We consider a family of densities $\{f_{a,b,c}\}_{(a,b,c) \in (0,+\infty)^3}$ such that $f_{a,b,c} : \mathbb{D}_1 \to [0,+\infty)$ is defined for any x and y by

$$f_{a,b,c}(x,y) = g_{a,b}(x^2 + y^2)g_{c,c}\left(\frac{\text{atan2}(y,x)}{2\pi} \mod 1\right)$$

with $g_{a,b}$ the usual density of the beta distribution with parameters a and b. Equivalently, (X,Y) is distributed as $(\sqrt{R}\cos\Theta, \sqrt{R}\sin\Theta)$ with R and Θ independent with respective distributions $\text{Beta}(a,b)$ and $\text{Beta}(c,c)$. Four densities, plotted in Figure 1, are studied:

Case 1 $a = 1, b = 1, c = 1$. The density $f_1 := f_{1,1,1}$ is in fact the uniform density on the disk.

Case 2 $a = 1.5, b = 1, c = 1$. The density $f_2 := f_{1.5,1,1}$ takes small values in the centre of the disk and its values increase slowly as one gets close to the boundary.

Case 3 $a = 2, b = 1, c = 1$. The density $f_3 := f_{2,1,1}$ takes very small values in most of the disk and only on a small strip of the boundary takes larger values.

Case 4 $a = 1.5, b = 1, c = 3$. The density $f_4 := f_{1.5,1,3}$, contrary to the others, is not invariant by rotations. The mass is more important near the point $(-1, 0)$.
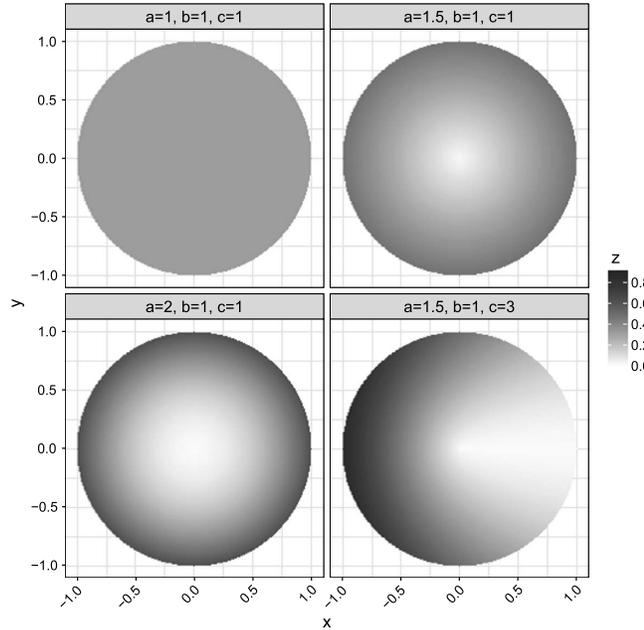

FIG 1. *Representation of the four densities.*

Quality criteria For each density function $f \in \{f_1, f_2, f_3, f_4\}$, we simulate $M = 500$ sequences of i.i.d. observations $(X_1, \ldots, X_n)$ where n is either 500, 1000, 2000 or 5000. Given an estimation procedure $\hat{f}$, we calculate M estimators $\hat{f}^{(1)}, \ldots, \hat{f}^{(M)}$. We consider the integrated squared error (ISE):

$$\mathrm{ISE}(\hat{f}^{(j)}) = \int_{\mathbb{D}_1} \left(f(x) - \hat{f}^{(j)}(x) \right)^2 dx \,.$$

Comparison of estimation procedures We consider a set $\mathcal{H}$ of 39 equally spaced bandwidths between 0.05 and 1 with step 0.025. For each $h \in \mathcal{H}$ we define the *usual* kernel estimator $\widetilde{f}_h$ and $\hat{f}_h$ which is a modified version of our estimator defined in Section 3.1 that we call *boundary* estimator. These estimators are defined as follows:

$$\widetilde{f}_h(x) = \frac{1}{nh^2} \sum_{i=1}^{n} \widetilde{K}(h^{-1}(X_i - x)) \text{ with } \widetilde{K}(u_1, \ldots, u_d) = \prod_{j=1}^{d} \mathbf{I}_{\left[-\frac{1}{2}, \frac{1}{2}\right]}(u_j) \quad (12)$$

and

$$\hat{f}_h(x) = \begin{cases} \widetilde{f}_h(x) & \text{if } x \in \mathbb{D}_{1-h/\sqrt{2}} \\ \hat{f}_{K,h}(x) & \text{otherwise} \end{cases}.$$

Here $K = \mathbf{I}_{[0,1]}$ and the transformation A_x is the one in Example 2. We consider only uniform kernels to allow an easier comparison of the performances of the different estimators. We choose to implement our Goldenshluger-Lepski procedure for the estimator $\hat{f}_h$ since it has better practical performances than the estimator $\hat{f}_\ell$. In fact it enjoys good properties of $\hat{f}_{K,h}$ or $\hat{f}_\ell$ in the boundary and good performances of the usual kernel far from the boundary.

In Figure 2 and Figure 3 we compare, plotting boxplots of the ISE for $n \in \{500, 1000, 2000, 5000\}$ observations and $M = 500$ replications, the behavior of

- the oracle of the boundary estimators (called *boundary*) defined by

$$\hat{f}_{or} = \arg\min_{h \in \mathcal{H}} \text{ISE}(\hat{f}_h),$$

- the Goldenshluger Lepski procedure based on the *boundary* estimators (called *boundary_GL*), More precisely the *boundary_GL* estimator is given by $\hat{f}_{\hat{h}}$ with

$$\hat{h} = \arg\min_{h \in \mathcal{H}} \left(\widehat{B}(h) + M(h) \right) \tag{13}$$

$$M(h) = \frac{\tau \|K\|_2^d}{\sqrt{nh^d}}, \quad M(h,h') = M(h') + M(h' \vee h)$$

and

$$\widehat{B}(h) = \max_{h' \in \mathcal{H}} \left\{ \|\hat{f}_{h' \vee h} - \hat{f}_{h'}\|_2 - M(h,h') \right\}_+$$

- the oracle of *usual* Kernel estimators (called *usual*) defined by

$$\tilde{f}_{or} = \arg\min_{h \in \mathcal{H}} \text{ISE}(\tilde{f}_h),$$

- the Goldenshluger Lepki procedure based on the *usual* Kernel estimators (called *usual_GL*) given by $\tilde{f}_{\hat{h}}$ with $\tilde{h} = \arg\min_{h \in \mathcal{H}}(\widetilde{B}(h) + M(h))$ where

$$\widetilde{B}(h) = \max_{h' \in \mathcal{H}} \left\{ \|\tilde{f}_{h' \vee h} - \tilde{f}_{h'}\|_2 - M(h,h') \right\}_+$$

For the *boundary_GL* and *usual_GL*, we use the slope heuristic to calibrate the constant τ (see Baudry et al., 2012, and Section 6 in which we explicit the use of the slope heuristic.)

In almost all the cases (except for $n = 500$, $a = 2$, $b = 1$ and $c = 1$), the oracle based on *boundary* estimators as well as the GL procedure based on *boundary* estimators outperform the oracle and the GL procedures based on *usual* kernels. Note also that the ratio between the MISE of the GL procedure and the one of the oracle, both based on boundary estimator is around 1.6 (see Table 1) which means the GL procedure mimics quite well the oracle.
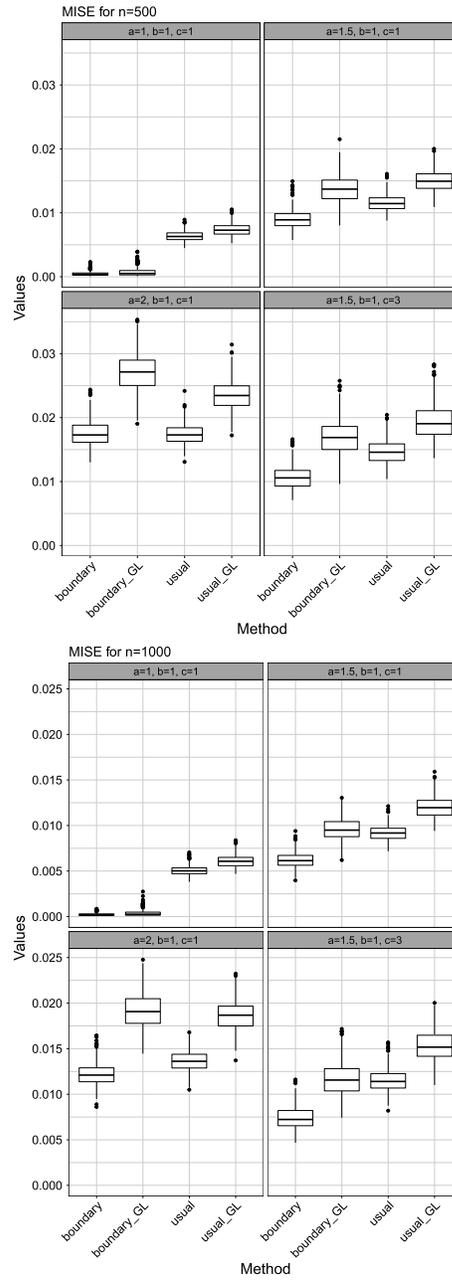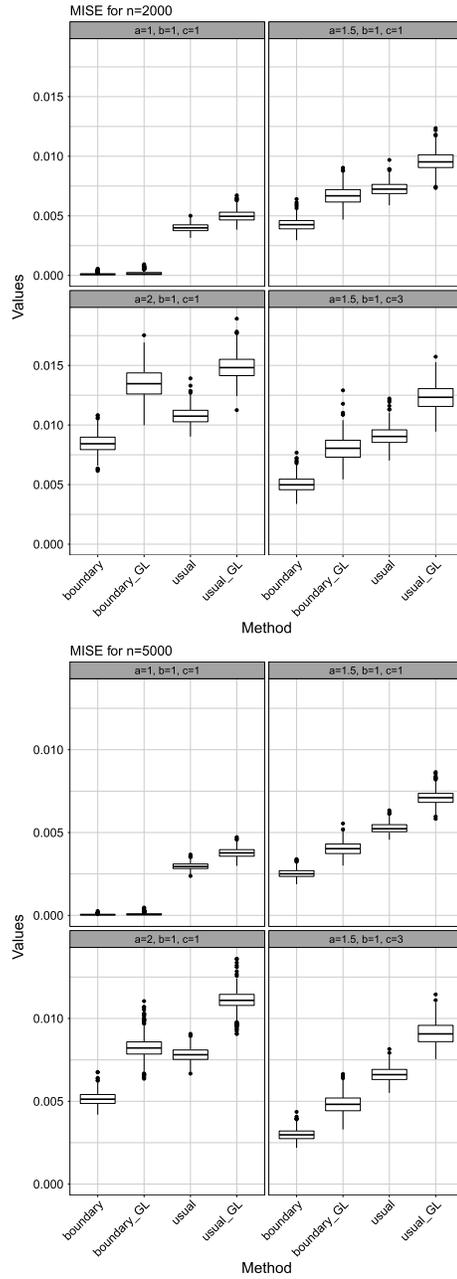
FIG 2. *Boxplots of the integrated squared error (ISE) on the disk for the models described by Cases 1 to 4, and sample sizes equal to 500 and 1000 for the four estimators boundary, boundary_GL, usual and usual_GL.*

FIG 3. *Boxplots of the integrated squared error (ISE) on the disk for the models described by Cases 1 to 4, and sample sizes equal to* 2000 *and* 5000 *for the four estimators boundary, boundary_GL, usual and usual_GL.*

TABLE 1
Mean and standard-deviation for the ratio between the ISE of the GL procedure and the one of the oracle, both for the boundary estimator, computed from $M = 500$ replications. The results are provided for different sample sizes: $n = 500$, 1000, 2000 and 5000.

Case	n	mean	sd	n	mean	sd
a = 1, b = 1, c = 1	500	1.63	1.12	2000	1.71	1.23
	1000	1.65	1.18	5000	1.55	1.03
a = 1.5, b = 1, c = 1	500	1.55	0.28	2000	1.57	0.21
	1000	1.57	0.23	5000	1.60	0.19
a = 2, b = 1, c = 1	500	1.56	0.21	2000	1.60	0.16
	1000	1.58	0.17	5000	1.60	0.13
a = 1.5, b = 1, c = 3	500	1.62	0.31	2000	1.62	0.23
	1000	1.59	0.25	5000	1.62	0.19

5.2. Diffusion

Let us consider in this section the following two dimensional reflected Langevin diffusion on the disk:

$$\begin{cases} dX_t &= dW_t^1 - \beta \frac{X_t}{(1+X_t^2+Y_t^2)^\beta} dt + n_1(X_t, Y_t) dL_t \\ dY_t &= dW_t^2 - \beta \frac{Y_t}{(1+X_t^2+Y_t^2)^\beta} dt + n_2(X_t, Y_t) dL_t \end{cases}$$

with $\beta > 1$, $(n_1(x,y), n_2(x,y))$, $(x,y) \in \partial\mathcal{D}$ defined the normal vector to the boundary of the domain $\mathcal{D} = \mathbb{D}_r$, W^1 and W^2 are two independent standard Brownian motions and L the local time on $\partial\mathcal{D}$. The process $Z_t = (X_t, Y_t)_{t \geq 0}$ is well known as Brownian motion with drift. This process is ergodic and exponential Φ-mixing in the sense of Ibragimov (1962). Recall that the ϕ-mixing coefficients of the strictly stationary process $Z_t = (X_t, Y_t)_{t \geq 0}$ are defined, for each $k \geq 1$, as:

$$\phi(k) = \sup \left| \frac{\mathbf{P}(U \cap V)}{\mathbf{P}(U)} - \mathbf{P}(V) \right|,$$

where the supremum is taken over all pairs U and V of the probability space Ω which are respectively measurable with respect to $\sigma(Z_s : s \leq 0)$ and $\sigma(Z_s : s \geq k)$. Remark that ϕ-mixing is stronger than β-mixing (see, e.g., Doukhan, 1994). The invariant measure is absolutely continuous with respect to the Lebesgue measure restricted to the disk $\mathbb{D}_r$. The invariant density writes as follows:

$$f(x, y) = \frac{1 - \beta}{\pi[(1 + r^2)^{1-\beta} - 1]} \frac{1}{(1 + x^2 + y^2)^\beta}.$$

Note that this density has most of its mass concentrated in the centre of the disk. In the simulations we fix $\beta = 2$ and we run as in Cattiaux et al. (2017) the Euler reflected scheme introduced in Bossy et al. (2004). As we are interested in the stationary regime, we throw away the first runs of the scheme. As in Section 5.1, we simulate $M = 500$ Monte-Carlo replications with sample size $n \in \{500, 1000, 2000, 5000\}$. We obtain that for all sample sizes, the GL procedure based on *boundary* estimators outperforms the GL procedure based on *usual* kernels (see Figure 4 for more details).
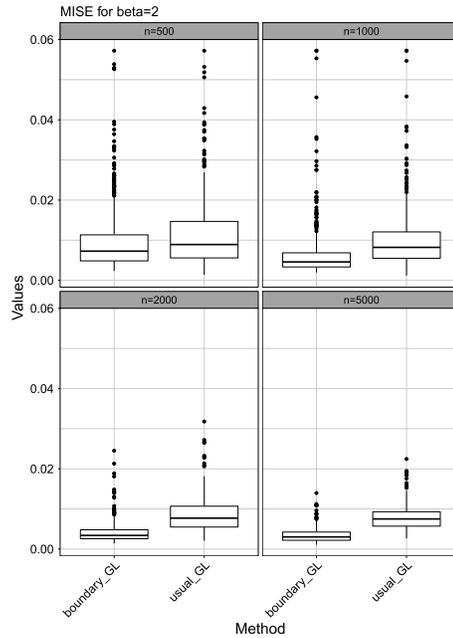
FIG 4. *Boxplots of the integrated squared error (ISE) of the GL procedure based on the boundary estimator for $\beta = 2$ and different sample sizes, $n = 500$, 1000, 2000 and 5000.*

Note that, similar types of processes (defined on a more complex domain) have been used to model population dynamics, in particular the home-range and the core-area of an animal based on tracking data (see Cholaquidis et al., 2016). Roughly speaking, the authors consider that the process under observation behaves in the interior of a geographical area $\mathcal{D} \subset \mathbb{R}^d$ like an ordinary Brownian motion with drift, and reflects (normally) at the boundary $\partial\mathcal{D}$ of $\mathcal{D}$. Under regularity assumptions on the drift and geometric constraints on the support $\mathcal{D}$, Cholaquidis et al. (2016) prove the existence of a unique stationary distribution, absolutely continuous with respect to the Lebesgue measure on $\mathcal{D}$. They also prove that the process is geometrically ergodic. Then the trajectories of the animals allow estimating the density of the invariant probability measure. In Section 6 we study a real data set of this type.

6. Application to a real data set

In this section our method is applied to a database obtained from the study "African Elephant (Migration) Chamaillé-Jammes Hwange NP" that resides in the repository Movebank (Wikelski and Kays, 2018, accessed on 2018/11/05).

The data were obtained from the GPS tracking of the migratory trajectories of 30 elephants, each from a different herd, evolving in the Hwange National Park in Zimbabwe (HNP in the following). The daily observations of their dis-

placements were taken by means of GPS devices located in collars installed in individuals of different herds (see Tshipa et al., 2017; Valls-Fox et al., 2018, and references therein). The date of installation of the collars was as follows: August 2009 (10 elephants), November 2012 (10 elephants), November 2014 (8 elephants) and February 2015 (2 elephants). Each elephant is observed during approximately 2 years.

We are interested in the spatial density of the whole set of elephants into the park. Our work represents a first statistical insight in the data, where we gather all the measures collected along time from the 30 elephants. At least the independence of all the trajectories does not seem unreasonable as only one elephant per herd was monitored. As in Cholaquidis et al. (2016) we then assume that the animal movements can be modeled by a reflected diffusion. This allows us to guarantee that the process that modeled the trajectories of the elephants satisfies the main assumptions of our model. Moreover, nine elephants were removed from the initial database as their behavior seems atypic.

6.1. Boundary of the park

Figure 5, obtained from © *OpenStreetMap contributors*, represents the boundary of HNP as well as the $n = 17501$ GPS positions of the elephants.
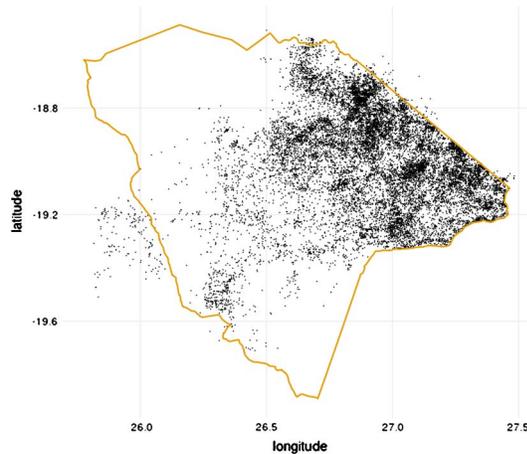

FIG 5. *Map of HNP with the $n = 17501$ GPS positions of the elephants.*

Note that the boundary is very simple on the east side (it consists mainly in straight lines) and more complex on the west side. However most of the observations lie in the east side. This is especially true for observations that are close to a boundary. Our methodology requires the knowledge of this boundary. Moreover, the more complex the boundary, the more difficult our method is to implement. With this in mind we propose to approximate the boundary of HNP by a simple polygon that adjusts quite well the real boundary in the east

part whereas a more rough approximation is used in the west part. Finally, a simple polygon with only 11 edges was chosen. Figure 6 represents both the real boundary (in yellow) and the approximating simple polygon (in blue):
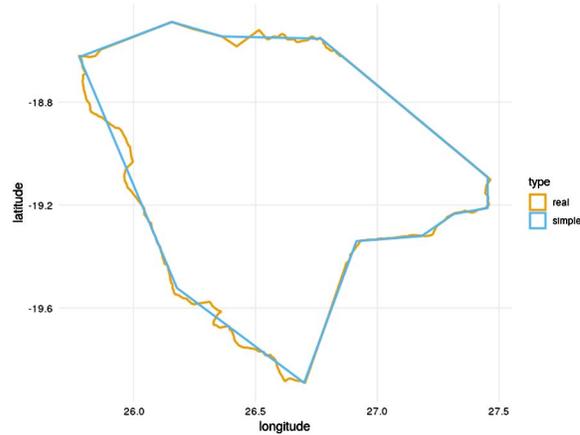

FIG 6. *Real and approximating boundary of HNP.*

6.2. *Estimation procedure*

In this section we follow the same strategy as in the simulation study: our boundary estimators are only used in a "neighborhood" of the boundary of the east side of HNP while classical kernel estimators are used otherwhere. More precisely, for each bandwidth h, we define two specific zones: the northeast zone C_1 and the southeast zone C_2. Examples of such zones are represented, for different values of h, in Figure 7.
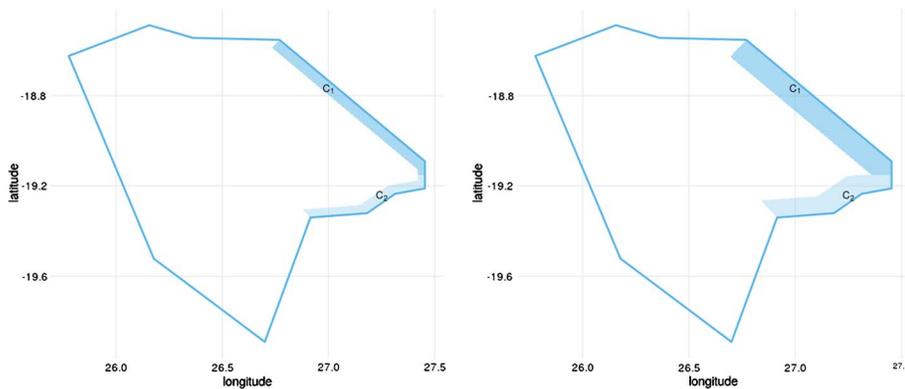

FIG 7. *Zones C_1 and C_2 for $h = 0.07$ (left), and $h = 0.15$ (right).*

Now, our procedure consists in selecting in a data-driven way a bandwidth among a family of 30 bandwidths equally spaced between 0.02 and 0.31. For each bandwidth h and any point x in HNP, the estimator $\hat{f}_h(x)$ is defined in two different ways depending on the position of x. If x belongs to $C_1 \cup C_2$ and if the distance of x to the boundary is less than h, then the boundary estimator is used and we define:

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{K}_h \circ A_x(X_i - x),$$

where $A_x = \begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}$ if $x \in C_1$ and $A_x = \begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}$ if $x \in C_2$. Otherwise the usual kernel estimator defined by (12) is used.

We use the same selection procedure defined in Section 5 defined by (13). Since our selection procedure depends on a tuning parameter τ, we propose to use a slope heuristic to determine this constant. We refer the reader to Baudry et al. (2012) for more details. As a consequence, for each tuning parameter τ, our procedure selects the quantity $\widehat{h}(\tau)$ and the slope heuristic consists in defining τ_{min} as the largest slope of the function $\tau \mapsto M(\widehat{h}(\tau))$. As it can be observed in Figure 8, we obtain $\tau_{\min} = 0.8$ so we implement our GL procedure with $\tau = 2\tau_{\min} = 1.6$. The bandwidth which is selected then equals to $h = 0.07$.
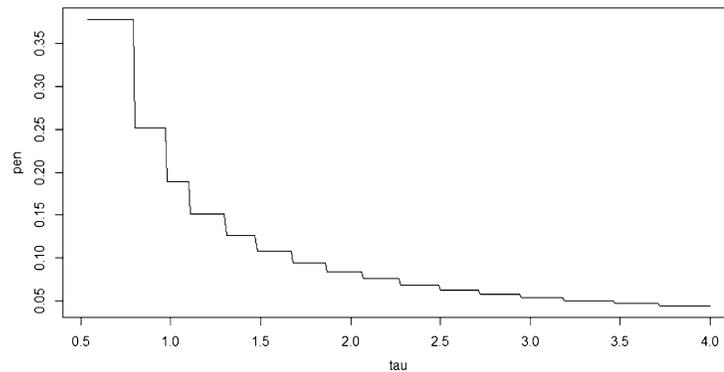

FIG 8. *Value of the penalization $\widehat{M}(\widehat{h}(\tau))$ in terms of τ.*

6.3. Results

Figure 9 represents the final density estimation plotted on a spatial regular grid (with step $\delta = 0.01$) within the real boundary.

The result we obtain seems to confirm that the density of elephants is related to the placement of artificial water pumps. It would be interesting to investigate further that issue with the owners of this database.
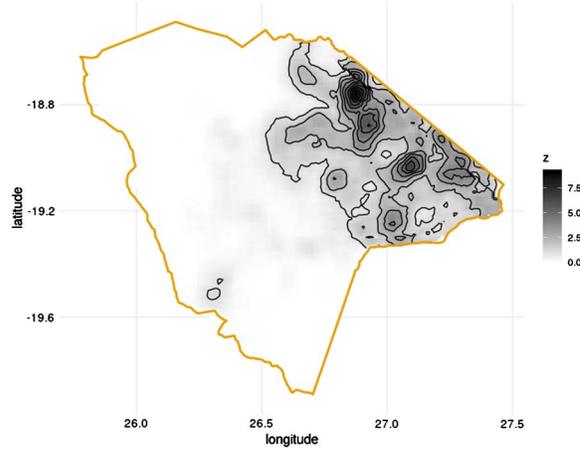
FIG 9. *The estimated repartition of the elephants within HNP.*

7. Proofs

7.1. Preliminary notations and technical lemmas

For any $\ell \in \mathcal{L}_n$, define

$$M(\ell) = \frac{\Gamma(\ell) + \tau \|K_{m_n(\ell)}\|_2^d}{\sqrt{nh^d(\ell)}} \text{ where } \Gamma(\ell) = \Gamma(K_{m_n(\ell)}, h(\ell))$$

with $\Gamma(K, h)$ defined by (10). For any kernel K and bandwidth $h > 0$, define:

$$\xi_{K,h}(x) = \frac{\sqrt{nh^d}}{n} \sum_{i=1}^n \left(\mathbf{K}_h \circ A_x(X_i - x) - \mathbf{E}\mathbf{K}_h \circ A_x(X_i - x) \right), \, x \in \mathcal{D}.$$

We denote $\mathcal{D}_i = \{x \in \mathcal{D} \colon A_x = A_i\}$. The following lemma, whose proof is postponed in Section 7.5, provides some properties of $\xi_{K,h}$.

Lemma 5. *We have*

$$\Gamma(K, h) \leq \sqrt{\kappa} \|K\|_2^d. \tag{14}$$

Moreover under Assumptions 1, 2, 3, 4 and 5, there exists an absolute constant $C(c, \rho, f_\infty)$ such that we have

$$\mathbf{E}\|\xi_{K,h}\|_2^2 \leq \Gamma^2(K, h) + \kappa \|K\|_2^{2d} C(c, \rho, f_\infty) h^{d/2} \tag{15}$$

$$\leq \kappa \|K\|_2^{2d} \left(1 + C(c, \rho, f_\infty) h^{d/2} \right) \tag{16}$$

and $\exists n_0(c, \rho, f_\infty, c_1, \kappa, \tau)$ such that for $n \geq n_0(c, \rho, f_\infty, c_1, \kappa, \tau)$ and $h \leq \overline{h}_n$ we have:

$$\mathbf{E}\|\xi_{K,h}\|_2 \leq \Gamma(K, h) + \|K\|_2^d \frac{\tau}{4}. \tag{17}$$

The proof uses the Bousquet inequality from Boucheron et al. (2004) (see also Theorem 12.5 in Boucheron et al., 2013). It also makes use of ingredients in Lemma 4.2 in Viennet (1997). Both Lemmas are recalled in Appendix C. Note also that we will use in several parts that

$$\|K_m\|_\infty \leq A(m+1)^B$$

where $A = 1$ and $B = 2$.

7.2. Proof of Proposition 1

Proof of Bound (4)   Set $f \in \mathcal{H}_\mathcal{D}(\gamma, L)$ and let K be a kernel of order $m = \lfloor \gamma \rfloor$. Using that $|\det(A_x)| = 1$, we have for $x \in \mathcal{D}$

$$\mathcal{B}_{K,h}(f,x) = \mathbf{E}\left(\hat{f}_{K,h}(x)\right) - f(x) = \int_\mathcal{D} \mathbf{K}_h \circ A_x(u-x)f(u)du - f(x)$$

$$= \int_{[0,1]^d} \prod_{i=1}^d K(s_i) \left[f(x + hA_x^{-1}(s)) - f(x)\right] ds.$$

Using a Taylor expansion we obtain

$$f(x + hA_x^{-1}(s)) - f(x) = \sum_{|\alpha| \leq m} \frac{D^\alpha f(x)}{\alpha!} h^{|\alpha|} (A_x^{-1}(s))^\alpha$$

$$+m \sum_{|\alpha|=m} \frac{h^m (A_x^{-1}(s))^\alpha}{\alpha!} \int_0^1 (1-t)^{m-1} \left[D^\alpha f(x + thA_x^{-1}(s)) - D^\alpha f(x)\right] dt,$$

where for $x \in \mathbb{R}^d$ and $\alpha \in (\mathbb{N} \cup \{0\})^d$ we define $x^\alpha = \prod_{i=1}^d x_i^{\alpha_i}$. Using that K is a kernel of order m and that $(A_x^{-1}(s))^\alpha$ is a polynomial in $(s_1, \ldots, s_d)$ of degree α, we have

$$|\mathcal{B}_{K,h}(f,x)| \leq Lh^\gamma \sum_{|\alpha|=m} \int_{[0,1]^d} \prod_{i=1}^d |K(s_i)||(A_x^{-1}(s))^\alpha||A_x^{-1}(s)|^{\gamma-m} ds$$

$$\leq Lh^\gamma \|K\|_\infty^d C(R,\gamma),$$

where $C(R, \gamma)$ is a positive constant that depends only on R and γ.

Proof of Bound (5)   We have

$$\mathbf{E}\|\hat{f}_{K,h} - \mathbf{E}\hat{f}_{K,h}\|_2^2 = \int_\mathcal{D} \mathrm{Var}\left(\frac{1}{n} \sum_{i=1}^n \mathbf{K}_h \circ A_x(X_i - x)\right) dx$$

$$\leq \frac{\kappa}{n^2} \sup_{j=1,\ldots,\kappa} \int_{\mathcal{D}_j} \mathrm{Var}\left(\sum_{i=1}^n \mathbf{K}_h \circ A_j(X_i - x)\right) dx.$$

Then using Lemma 9 in Appendix C, we deduce that there exists a sequence of random variables $(b_k(X_0))_{k \geq 0}$ such that $0 \leq b_k(X_0) \leq 1$ and $\mathbf{E}(b_k(X_0)) = \beta(k)$ and we have

$$\mathbf{E}\|\hat{f}_{K,h} - \mathbf{E}\hat{f}_{K,h}\|_2^2 \leq \frac{4n\kappa}{n^2} \sup_{j=1,\dots,\kappa} \mathbf{E}\left(\sum_{k=0}^n b_k(X_0) \int_{\mathcal{D}_j} \mathbf{K}_h^2 \circ A_j(X_0 - x)dx\right)$$

$$\leq \frac{4\kappa\|K\|_2^{2d}}{nh^d} \sum_{k=0}^n \beta(k) \leq \frac{4\kappa\|K\|_2^{2d}}{nh^d} \frac{c}{1-\rho}.$$

Final step Note that (6) is a direct consequence of (15) and (16) of Lemma 5. Finally choosing $h = n^{-1/(2\gamma+d)}$, the estimator $\hat{f}_{K,h}$ reaches the minimax rate $n^{-2\gamma/(2\gamma+d)}$.

7.3. Proof of Theorem 3

To prove Theorem 3, we use Berbee's coupling method as in Viennet (1997) (proof of Proposition 5.1) and proof of Theorem 1 of Comte et al. (2017). We assume $n = 2p_n q_n$ with $q_n = [\log n]^2$. Then there exist random variables X_i^*, $i = 1, \dots, n$ satisfying the following properties:

- For $r = 1, \dots, p_n$, the random vectors $\vec{U}_{r,1} = (X_{2(r-1)q_n+1}, \dots, X_{(2r-1)q_n})^T$ and $\vec{U}_{r,1}^* = (X_{2(r-1)q_n+1}^*, \dots, X_{(2r-1)q_n}^*)^T$ have the same distribution, and so have the vector $\vec{U}_{r,2} = (X_{(2r-1)q_n+1}, \dots, X_{2rq_n})^T$ and the vector $\vec{U}_{r,2}^* = (X_{(2r-1)q_n+1}^*, \dots, X_{2rq_n}^*)^T$.
- For $r = 1, \dots, p_n$, $\mathbb{P}(\vec{U}_{r,1} \neq \vec{U}_{r,1}^*) \leq \beta(q_n)$ and $\mathbb{P}(\vec{U}_{r,2} \neq \vec{U}_{r,2}^*) \leq \beta(q_n)$.
- For each $i \in \{1, 2\}$, the random vectors $\vec{U}_{1,i}^*, \dots, \vec{U}_{p_n,i}^*$ are independent.

We define $\Omega^* = \{X_i = X_i^*, i = 1, \dots, n\}$ and $\overline{\Omega}^*$ its complementary set in Ω. We have (see Comte et al., 2017)

$$\mathbb{P}(\overline{\Omega}^*) \leq 2p_n\beta(q_n) \leq n\beta(q_n). \tag{18}$$

Now denoting $\mathbf{K}_\ell(\cdot) = h(\ell)^{-d} \bigotimes_{i=1}^d K_{m_n(\ell)}\left((h(\ell))^{-1}\cdot\right)$, we define $\hat{f}_\ell^* = (\hat{f}_\ell^{*(1)} + \hat{f}_\ell^{*(2)})/2$ where

$$\hat{f}_\ell^{*(1)}(x) = \frac{2}{n} \sum_{r=1}^{p_n} \sum_{s=1}^{q_n} \mathbf{K}_\ell \circ A_x(X_{2(r-1)q_n+s}^* - x),$$

$$\hat{f}_\ell^{*(2)}(x) = \frac{2}{n} \sum_{r=1}^{p_n} \sum_{s=1}^{q_n} \mathbf{K}_\ell \circ A_x(X_{(2r-1)q_n+s}^* - x).$$

For any kernel K and any bandwith h, let $\xi_{K,h}^{*(1)}(x)$ (*resp.* $\xi_{K,h}^{*(2)}(x)$) defined as

$$2\frac{\sqrt{nh^d}}{n} \sum_{r=1}^{p_n} \sum_{s=1}^{q_n} \left(\mathbf{K}_h \circ A_x(X_{2(r-1)q_n+s}^* - x) - \mathbf{E}\mathbf{K}_h \circ A_x(X_{2(r-1)q_n+s}^* - x)\right),$$

$$2\frac{\sqrt{nh^d}}{n}\sum_{r=1}^{p_n}\sum_{s=1}^{q_n}\left(\mathbf{K}_h\circ A_x(X^*_{(2r-1)q_n+s}-x)-\mathbf{E}\mathbf{K}_h\circ A_x(X^*_{(2r-1)q_n+s}-x)\right),$$

and $\xi^*_{K,h}(x)=\frac{1}{2}(\xi^{*(1)}_{K,h}(x)+\xi^{*(2)}_{K,h}(x))$. The two following lemmas give some properties of $\xi^{*(1)}_{K,h}$ and $\xi^{*(2)}_{K,h}$. Lemma 6 follows immediately from Lemma 5.

Lemma 6. *We have*

$$\mathbf{E}\|\xi^{*(i)}_{K,h}\|_2\leq\left(2\Gamma^2(K,h)+2\kappa\|K\|_2^{2d}C(c,\rho,f_\infty)h^{d/2}\right)^{1/2}.$$

Moreover, $\exists\, n_0(c,\rho,f_\infty,c_1,\kappa,\tau)$ *such that for* $n\geq n_0(c,\rho,f_\infty,c_1,\kappa,\tau)$ *and* $h\leq\overline{h}_n$ *we also have:* $\mathbf{E}\|\xi^{*(i)}_{K,h}\|_2\leq\sqrt{2}\left(\Gamma(K,h)+\|K\|_2^d\frac{\tau}{4}\right)$.

Lemma 7. *For any* $\delta>0$, $x\geq 0$, *any* $i=1,2$ *and* $n^{-1}\leq h\leq\overline{h}_n$, $\mathbf{P}(\|\xi^{*(i)}_{K,h}\|_2-\mathbf{E}\|\xi^{*(i)}_{K,h}\|_2>\delta\|K\|_2^d+x)$ *is bounded by*

$$\widetilde{C}_0\exp\left(-\frac{\widetilde{C}_1 x^2\overline{h}_n^{-d/2}}{\|K\|_2^{2d}+x\|K\|_2^d}\right)\exp\left(-\widetilde{C}_2\overline{h}_n^{-d/2}\right)$$

where $\widetilde{C}_0$, $\widetilde{C}_1$, $\widetilde{C}_2$ *are positive constants that depend only on* c_1, δ, ρ, c, κ *and* f_∞.

We are now able to prove the oracle inequality. Set $\ell\in\mathcal{L}_n$. Using the triangular inequality we get:

$$\|f-\widehat{f}\|_2\leq\|f-\hat{f}_\ell\|_2+\|\hat{f}_{\widehat{\ell}\wedge\ell}-\hat{f}_\ell\|_2+\|\hat{f}_{\widehat{\ell}\wedge\ell}-\hat{f}_{\widehat{\ell}}\|_2.$$

Note that if $\ell\geq\widehat{\ell}$, using the definitions of $\widehat{B}(\ell)$ and $\widehat{M}(\ell)$, we easily obtain:

$$\begin{aligned}\|f-\widehat{f}\|_2&\leq\|f-\hat{f}_\ell\|_2+\|\hat{f}_{\widehat{\ell}\wedge\ell}-\hat{f}_\ell\|_2\leq\|f-\hat{f}_\ell\|_2+\widehat{B}(\widehat{\ell})+\widehat{M}(\widehat{\ell},\ell)\\&\leq\|f-\hat{f}_\ell\|_2+\widehat{B}(\widehat{\ell})+\widehat{M}(\ell)+\widehat{M}(\widehat{\ell})\\&\leq\|f-\hat{f}_\ell\|_2+2\left(\widehat{B}(\ell)+\widehat{M}(\ell)\right).\end{aligned}$$

Last inequality comes from the definition of $\widehat{\ell}$. The same bound remains valid if $\ell\leq\widehat{\ell}$. This implies:

$$\left(\mathbf{E}\|\widehat{f}-f\|_2^2\right)^{1/2}\leq\left(\mathbf{E}\|\hat{f}_\ell-f\|_2^2\right)^{1/2}+2\left(\mathbf{E}\widehat{B}^2(\ell)\right)^{1/2}+2\left(\mathbf{E}\widehat{M}^2(\ell)\right)^{1/2}.\quad(19)$$

It remains to bound each term of the right hand side of this inequality.

1 We have:

$$\begin{aligned}\left(\mathbf{E}\|\hat{f}_\ell-f\|_2^2\right)^{1/2}&\leq\|\mathbf{E}\hat{f}_\ell-f\|_2+\left(\mathbf{E}\|\hat{f}_\ell-\mathbf{E}\hat{f}_\ell\|_2^2\right)^{1/2}\\&\leq\|\mathbf{E}\hat{f}_\ell-f\|_2+\sqrt{\kappa}\|K_{m_n(\ell)}\|_2^d\sqrt{\frac{1}{nh^d(\ell)}+\frac{C(\rho,c,f_\infty)}{nh^{d/2}(\ell)}}.\quad(20)\end{aligned}$$

Last line follows from Bound (6) of Proposition 1.

2 Using triangular inequality and (14), last term can be bounded by:

$$\left(\mathbf{E}\widehat{M}^2(\ell)\right)^{1/2} \leq \sqrt{2}\,M(\ell) \leq \frac{\sqrt{2}(\sqrt{\kappa}+\tau)\|K_{m_n(\ell)}\|_2^d}{\sqrt{nh^d(\ell)}}. \tag{21}$$

3 Remark that, using the triangular inequality, we have:

$$\begin{aligned}
\widehat{B}(\ell) &\leq 2\max_{\ell'\in\mathcal{L}_n}\left\{\|\hat{f}_{\ell'}-\mathbf{E}\hat{f}_{\ell'}\|_2 - \widehat{M}(\ell')\right\}_+ + \max_{\ell'\in\mathcal{L}_n}\|\mathbf{E}\hat{f}_{\ell'}-\mathbf{E}\hat{f}_{\ell\wedge\ell'}\|_2\\
&\leq 2\max_{\ell'\in\mathcal{L}_n}\left\{\|\hat{f}_{\ell'}-\mathbf{E}\hat{f}_{\ell'}\|_2 - \widehat{M}(\ell')\right\}_+ + 2\max_{\ell'\geq\ell}\|\mathbf{E}\hat{f}_{\ell'}-f\|_2\\
&\leq \max_{\ell'\in\mathcal{L}_n}\left\{\|\hat{f}_{\ell'}^{*(1)}-\mathbf{E}\hat{f}_{\ell'}^{*(1)}\|_2 - \widehat{M}(\ell')\right\}_+\\
&\quad + \max_{\ell'\in\mathcal{L}_n}\left\{\|\hat{f}_{\ell'}^{*(2)}-\mathbf{E}\hat{f}_{\ell'}^{*(2)}\|_2 - \widehat{M}(\ell')\right\}_+\\
&\quad + 2\max_{\ell'\geq\ell}\|\mathbf{E}\hat{f}_{\ell'}-f\|_2 + 2\max_{\ell'\in\mathcal{L}_n}\|\hat{f}_{\ell'}^*-\hat{f}_{\ell'}\|_2.
\end{aligned}$$

This implies that $(\mathbf{E}\widehat{B}^2(\ell))^{1/2}$ is bounded by:

$$2(\mathbf{E}(\max_{\ell'\in\mathcal{L}_n}\|\hat{f}_{\ell'}^*-\hat{f}_{\ell'}\|_2)^2)^{1/2} + 2\max_{\ell'\geq\ell}\|\mathbf{E}\hat{f}_{\ell'}-f\|_2 + (\sqrt{\Delta_n^{(1)}}+\sqrt{\Delta_n^{(2)}}), \tag{22}$$

where for $i=1,2$ $\Delta_n^{(i)} = \mathbf{E}(\max_{\ell'\in\mathcal{L}_n}\{\|\hat{f}_{\ell'}^{*(i)}-E\hat{f}_{\ell'}^{*(i)}\|_2 - \widehat{M}(\ell')\}_+)^2$. It remains to study the terms of the right hand side of (22).

4. Study of $\Delta_n^{(i)}$ First define for $m\in\mathbb{N}$, the kernel $K_m^* = K_m^2/\|K_m\|_2^2$ and consider, for $\ell'\in\mathcal{L}_n$ the event

$$\mathcal{D}_{\ell'} = \left\{\|\xi_{K_{m_n(\ell')}^*,h(\ell')}\|_1^{1/2} \leq \frac{\tau}{2}\left(nh^d(\ell')\right)^{1/4}\right\}.$$

Now, remark that, on the event $\mathcal{D}_{\ell'}$ we have:

$$\begin{aligned}
\frac{|\widehat{M}(\ell')-\sqrt{2}\,M(\ell')|}{\sqrt{2}} &= \left|\frac{\hat{\Gamma}(\ell')-\Gamma(\ell')}{\sqrt{nh^d(\ell')}}\right|\\
&= \frac{\|K_{m_n(\ell')}\|_2^d}{\sqrt{nh^d(\ell')}}\left|\|\hat{f}_{K_{m_n(\ell')}^*,h(\ell')}\|_1^{1/2} - (\mathbf{E}\|\hat{f}_{K_{m_n(\ell')}^*,h(\ell')}\|_1)^{1/2}\right|\\
&\leq \frac{\|K_{m_n(\ell')}\|_2^d}{\sqrt{nh^d(\ell')}}\left((nh^d(\ell'))^{-1/2}\|\xi_{K_{m_n(\ell')}^*,h(\ell')}\|_1\right)^{1/2}\\
&\leq \frac{\|K_{m_n(\ell')}\|_2^d}{(nh^d(\ell'))^{3/4}}\|\xi_{K_{m_n(\ell')}^*,h(\ell')}\|_1^{1/2} \leq \frac{\tau\|K_{m_n(\ell')}\|_2^d}{2(nh^d(\ell'))^{1/2}}.
\end{aligned}$$

This implies that, on $\mathcal{D}_{\ell'}$

$$\widehat{M}(\ell') \geq \sqrt{2}\frac{\Gamma(\ell')+\frac{\tau}{2}\|K_{m_n(\ell')}\|_2^d}{(nh^d(\ell'))^{1/2}}. \tag{23}$$

Using (23), we obtain for $i = 1, 2$

$$\left\{ \|\hat{f}_{\ell'}^{*(i)} - \mathbf{E}\hat{f}_{\ell'}^{*(i)}\|_2 - \widehat{M}(\ell') \right\}_+$$

$$\leq \|\hat{f}_{\ell'}^{*(i)} - \mathbf{E}\hat{f}_{\ell'}^{*(i)}\|_\infty \mathbf{I}_{\bar{\mathcal{D}}_{\ell'}} + \left\{ \|\hat{f}_{\ell'}^{*(i)} - \mathbf{E}\hat{f}_{\ell'}^{*(i)}\|_2 - \widehat{M}(\ell') \right\}_+ \mathbf{I}_{\mathcal{D}_{\ell'}},$$

$$\leq \frac{2\|K_{m_n(\ell')}\|_\infty^d}{\underline{h}_n^d} \mathbf{I}_{\bar{\mathcal{D}}_{\ell'}}$$

$$+ \left\{ \|\hat{f}_{\ell'}^{*(i)} - \mathbf{E}\hat{f}_{\ell'}^{*(i)}\|_2 - \sqrt{2}\frac{\Gamma(\ell') + \frac{\tau}{2}\|K_{m_n(\ell')}\|_2^d}{(nh(\ell')^d)^{1/2}} \right\}_+.$$

Now, using triangular inequality

$$\left(\Delta_n^{(i)}\right)^{1/2} \leq \sum_{\ell' \in \mathcal{L}_n} \left(\mathbf{E}\left\{ \|\hat{f}_{\ell'}^{*(i)} - \mathbf{E}\hat{f}_{\ell'}^{*(i)}\|_2 - \widehat{M}(\ell') \right\}_+^2 \right)^{1/2}$$

$$\leq \sum_{\ell' \in \mathcal{L}_n} \frac{2\|K_{m_n(\ell')}\|_\infty^d}{\underline{h}_n^d} \sqrt{\mathbf{P}(\bar{\mathcal{D}}_{\ell'})}$$

$$+ \sqrt{ \mathbf{E}\left\{ \|\hat{f}_{\ell'}^{*(i)} - \mathbf{E}\hat{f}_{\ell'}^{*(i)}\|_2 - \sqrt{2}\frac{\Gamma(\ell') + \frac{\tau}{2}\|K_{m_n(\ell')}\|_2^d}{(nh(\ell')^d)^{1/2}} \right\}_+^2 } \quad (24)$$

Assume that here and after we have $n \geq n_0(c, \rho, f_\infty, c_1, \kappa, \tau)$. We consider:

$$[nh^d(\ell')]\mathbf{E}\left\{ \|\hat{f}_{\ell'}^{*(i)} - \mathbf{E}\hat{f}_{\ell'}^{*(i)}\|_2 - \sqrt{2}\frac{\Gamma(\ell') + \frac{\tau}{2}\|K_{m_n(\ell')}\|_2^d}{(nh^d(\ell'))^{1/2}} \right\}_+^2$$

$$= \mathbf{E}\left\{ \|\xi_{K_{m_n(\ell')},h(\ell')}^{*(i)}\|_2 - \sqrt{2}\Gamma(\ell') - \sqrt{2}\frac{\tau}{2}\|K_{m_n(\ell')}\|_2^d \right\}_+^2$$

$$= \int_0^{+\infty} \mathbf{P}\left(\|\xi_{K_{m_n(\ell')},h(\ell')}^{*(i)}\|_2 - \sqrt{2}\left(\Gamma(\ell') + \frac{\tau}{2}\|K_{m_n(\ell')}\|_2^d\right) > x^{1/2} \right) dx$$

$$\leq 2\int_0^{+\infty} xP(x)dx.$$

where

$$P(x) = \mathbf{P}\left(\|\xi_{K_{m_n(\ell')},h(\ell')}^{*(i)}\|_2 - \mathbf{E}\|\xi_{K_{m_n(\ell')},h(\ell')}^{*(i)}\|_2 > \frac{\tau\|K_{m_n(\ell')}\|_2^d}{\sqrt{2}} + x - \mathcal{H}(\ell') \right)$$

and $\mathcal{H}(\ell') = \frac{\tau}{2\sqrt{2}}\|K_{m_n(\ell')}\|_2^d$. Last line follows from Lemma 6.

Using Lemma 7 with $\delta = \tau/(2\sqrt{2})$ and $x = 0$ we obtain:

$$\mathbf{E}\left\{ \|\hat{f}_{\ell'}^{*(i)} - \mathbf{E}\hat{f}_{\ell'}^{*(i)}\|_2 - \sqrt{2}\frac{\Gamma(\ell') + \frac{\tau}{2}\|K_{m_n(\ell')}\|_2^d}{(nh(\ell')^d)^{1/2}} \right\}_+^2$$

$$\leq \frac{2\widetilde{C}_0 \exp\left(-\widetilde{C}_2 \overline{h}_n^{-d/2}\right)}{nh^d(\ell')} \int_0^{+\infty} x \exp\left(-\widetilde{C}_1 \frac{x^2 \overline{h}_n^{-d/2}}{\|K_{m_n(\ell')}\|_2^{2d} + x\|K_{m_n(\ell')}\|_2^d}\right) dx$$

$$\leq \frac{2\widetilde{C}_0 \exp\left(-\widetilde{C}_2 \overline{h}_n^{-d/2}\right)}{nh^d(\ell')} \int_0^{+\infty} x \exp\left(-\widetilde{C}_1 \frac{x^2 \overline{h}_n^{-d/2}}{\alpha_n^2 + x\alpha_n}\right) dx$$

where $\alpha_n = A^d (\log n + 3/2)^{Bd}$. This follows from the fact that the kernel $K_{m_n(\ell')}$ satisfies (8) combined with the expression of $m_n(\ell')$. Now, splitting the integral into two terms, depending on the position of x with respect to α_n, we obtain:

$$\mathbf{E}\left\{\|\hat{f}_{\ell'}^{*(i)} - \mathbf{E}\hat{f}_{\ell'}^{*(i)}\|_2 - \sqrt{2}\frac{\Gamma(\ell') + \frac{\tau}{2}\|K_{m_n(\ell')}\|_2^d}{(nh(\ell')^d)^{1/2}}\right\}_+^2$$

$$\leq \frac{\widetilde{C}_3}{2} \frac{\exp\left(-\widetilde{C}_2 \overline{h}_n^{-d/2}\right)}{n\underline{h}_n^d} \alpha_n^2 \left(\overline{h}_n^{d/2} + \overline{h}_n^d\right)$$

$$\leq \widetilde{C}_3 \frac{\exp\left(-\widetilde{C}_2 \overline{h}_n^{-d/2}\right)}{n\underline{h}_n^d} \alpha_n^2 \overline{h}_n^{d/2}$$

where $\widetilde{C}_3$ is a positive constant that depends on τ, ρ, c, κ, A, B, c_1 and f_∞. This implies that there exists $\widetilde{C}_4$ a positive constant that depends on τ, ρ, c, κ, A, B, c_1, and f_∞ such that we have:

$$\sum_{\ell' \in \mathcal{L}_n} \left(\mathbf{E}\left\{\|\hat{f}_{\ell'}^{*(i)} - \mathbf{E}\hat{f}_{\ell'}^{*(i)}\|_2 - \sqrt{2}\frac{\Gamma(\ell') + \frac{\tau}{2}\|K_{m_n(\ell')}\|_2^d}{(nh(\ell')^d)^{1/2}}\right\}_+^2\right)^{1/2} \leq \widetilde{C}_4 n^{-1/2}.$$

(25)

It remains to bound $\sum_{\ell' \in \mathcal{L}_n} \frac{2\|K_{m_n(\ell')}\|_\infty^d}{\underline{h}_n^d} \sqrt{\mathbf{P}(\overline{\mathcal{D}}_{\ell'})}$. As $\|\cdot\|_1 \leq \sqrt{\mathrm{Vol}(\mathcal{D})}\|\cdot\|_2$:

$$\mathbf{P}(\overline{\mathcal{D}}_{\ell'}) \leq \mathbf{P}\left(\|\xi_{K_{m_n(\ell')}^*,h(\ell')}\|_2 \geq \frac{\tau^2 (nh^d(\ell'))^{1/2}}{4\sqrt{\mathrm{Vol}(\mathcal{D})}}\right)$$

$$\leq 2\mathbf{P}\left(\|\xi_{K_{m_n(\ell')}^*,h(\ell')}^{*(1)}\|_2 \geq \frac{2}{3}\mathcal{U}(\ell')\right)$$

$$+ 3\mathbf{E}\|\xi_{K_{m_n(\ell')}^*,h(\ell')} - \xi_{K_{m_n(\ell')}^*,h(\ell')}^*\|_2 \frac{1}{\mathcal{U}(\ell')}$$

with $\mathcal{U}(\ell') = \frac{\tau^2 (nh^d(\ell'))^{1/2}}{4\sqrt{\mathrm{Vol}(\mathcal{D})}}$. Then, using Lemma 5 and Lemma 6, we have:

$$\mathbf{E}\|\xi_{K_{m_n(\ell')}^*,h(\ell')}^{*(1)}\|_2 + \|K_{m_n(\ell')}^*\|_2^d \leq \left(\frac{\tau}{2\sqrt{2}} + \sqrt{2}\sqrt{\kappa} + 1\right)\|K_{m_n(\ell')}^*\|_2^d.$$

Now using that $\|K_{m_n(\ell')}\|_2 \geq 1$ and (8), we deduce that

$$\|K_{m_n(\ell')}^*\|_2^d = \frac{\|K_{m_n(\ell')}\|_4^{2d}}{\|K_{m_n(\ell')}\|_2^{2d}} \leq \|K_{m_n(\ell')}\|_\infty^{2d}$$

$$\leq A^{2d}(m_n(\ell') + 1)^{2Bd} \leq A^{2d}\left(\frac{\log n}{2\ell'} + \frac{3}{2}\right)^{2Bd}.$$

Now remark that:

$$\left(\frac{\log n}{2\ell'} + \frac{3}{2}\right)^{2Bd} \leq (nh^d(\ell'))^{1/2}\left(\frac{n^{-1/(4Bd)}\log n}{2\ell'e^{-\ell'/(4B)}} + \frac{3(\log n)^{-c_2/(4Bd)}}{2}\right)^{2Bd}.$$

Moreover it is easily seen that, for n large enough:

$$\frac{n^{-1/(4Bd)}\log n}{2\ell'e^{-\ell'/(4B)}} \leq \max\left(\frac{n^{-1/(4Bd)}\log n}{2e^{-1/(4B)}}, \frac{n^{-1/(4Bd)}\log n}{2\ell^*e^{-\ell^*/(4B)}}\right)$$

$$\leq \max\left(\frac{n^{-1/(4Bd)}\log n}{2e^{-1/(4B)}}, d(\log n)^{-c_2/(4Bd)}\right)$$

$$\leq d(\log n)^{-c_2/(4Bd)},$$

where, $\ell^* = [\frac{1}{d}(\log n - c_2 \log \log n)]$ is such that, for n large enough:

$$\frac{\log n}{2d} \leq \ell^* \leq \frac{1}{d}(\log n - c_2 \log \log n).$$

Then we obtain for n large enough:

$$\mathbf{E}\|\xi^{*(1)}_{K^*_{m_n(\ell')},h(\ell')}\|_2 + \|K^*_{m_n(\ell')}\|_2^d$$

$$\leq \sqrt{nh^d(\ell')}\left(\frac{\tau}{2\sqrt{2}} + \sqrt{2}\sqrt{\kappa} + 1\right)A^{2d}\left(d + \frac{3}{2}\right)^{2Bd}(\log n)^{-c_2/2}$$

$$\leq \frac{\tau^2(nh^d(\ell'))^{1/2}}{6\sqrt{\mathrm{Vol}(\mathcal{D})}}.$$

This implies, using Lemma 7 with $\delta = 1$ and $x = 0$:

$$\exists \widetilde{C}_5 > 0 \text{ s.t. } \mathbf{P}\left(\|\xi^{*(1)}_{K^*_{m_n(\ell')},h(\ell')}\|_2 \geq \frac{2}{3}\mathcal{U}(\ell')\right) \leq \widetilde{C}_5 \exp(-\widetilde{C}_2 \overline{h}_n^{-d/2}).$$

We also have:

$$\mathbf{E}\|\xi_{K^*_{m_n(\ell')},h(\ell')} - \xi^*_{K^*_{m_n(\ell')},h(\ell')}\|_2$$

$$\leq \sqrt{nh(\ell')^d}\sqrt{\frac{4}{h(\ell')^d}}\|K^*_{m_n(\ell')}\|_2^d\,\mathbf{P}(\overline{\Omega}^*) \leq 2n\sqrt{n}\beta(q_n)\|K^*_{m_n(\ell')}\|_\infty^{2d}. \qquad (26)$$

Thus we have:

$$\mathbf{P}(\bar{\mathcal{D}}_{\ell'}) \leq 2\widetilde{C}_5 \exp(-\widetilde{C}_2(\overline{h}_n)^{-d/2}) + 2n\beta(q_n)\sqrt{n}\|K^*_{m_n(\ell')}\|_\infty^{2d}\frac{3 \times 4\sqrt{\mathrm{Vol}(\mathcal{D})}}{\tau^2\sqrt{nh^d}}.$$

This implies that we have that $\sum_{\ell' \in \mathcal{L}_n} \frac{2\|K_{m_n(\ell')}\|_\infty^d}{\underline{h}_n^d} \sqrt{\mathbf{P}(\bar{\mathcal{D}}_{\ell'})}$ is bounded by:

$$\sum_{\ell' \in \mathcal{L}_n} \frac{2\|K_{m_n(\ell')}\|_\infty^d}{\underline{h}_n^d} \sqrt{2\widetilde{C}_5 \exp(-\widetilde{C}_2 \bar{h}_n^{-d/2})}$$
$$+ \sum_{\ell' \in \mathcal{L}_n} \frac{2\|K_{m_n(\ell')}\|_\infty^{2d}}{\underline{h}_n^d} \sqrt{2n\sqrt{n}\beta(q_n)} \frac{\sqrt{3} \times 2(\mathrm{Vol}(\mathcal{D}))^{1/4}}{\tau(nh^d)^{1/4}}.$$

Thus there exists a positive constant $\widetilde{C}_6$ such that

$$\sum_{\ell' \in \mathcal{L}_n} \frac{2\|K_{m_n(\ell')}\|_\infty^d}{\underline{h}_n^d} \sqrt{\mathbf{P}(\bar{\mathcal{D}}_{\ell'})} \leq \widetilde{C}_6 n^{-1/2} \tag{27}$$

as $q_n = [\log n]^2$. Using (24), (25) and (27), we can conclude for $i = 1, 2$ that

$$\left(\Delta_n^{(i)} \right)^{1/2} \leq (\widetilde{C}_4 + \widetilde{C}_6) n^{-1/2} \tag{28}$$

5. Final step Using (18) and proceeding as in (26) we obtain that

$$\sqrt{\mathbf{E} \left(\max_{\ell' \in \mathcal{L}_n} \|\hat{f}_{\ell'}^* - \hat{f}_{\ell'}\|_2 \right)^2} \leq \widetilde{C}_7 n^{-1/2} \tag{29}$$

and taking (22), (28) and (29) together we obtain:

$$\left(\mathbf{E} \widehat{B}^2(\ell) \right)^{1/2} \leq \widetilde{C}_8 n^{-1/2} + 2 \max_{\ell' \geq \ell} \|\mathbf{E} \hat{f}_{\ell'} - f\|_2 \tag{30}$$

with $\widetilde{C}_7$ and $\widetilde{C}_8$ are positive constants depending on c_1, c_2, τ, c, ρ, f_∞, κ, A, B.

6. Conclusion Combining (19), (20), (21), (30) we finally obtain

$$\left(\mathbf{E} \|\hat{f} - f\|_2^2 \right)^{1/2} \leq 5 \max_{\ell' \geq \ell} \|\mathbf{E} \hat{f}_{\ell'} - f\|_2 + \widetilde{C}_9 \left(\frac{\|K_{m_n(\ell)}\|_2^d}{\sqrt{nh^d(\ell)}} + n^{-1/2} \right)$$

with $\widetilde{C}_9$ a positive constant depending on c_1, c_2, τ, c, ρ, f_∞, κ, A, B. We thus get the result with $\widetilde{\Upsilon}_1 = 5 + \widetilde{C}_9$ and $\widetilde{\Upsilon}_2 = \widetilde{C}_9$.

7.4. Proof of Theorem 4

Set $\gamma > 0$ and $L > 0$ and let $f \in \mathcal{H}_{\mathcal{D}}(\gamma, L)$. Define $\ell_0 = \lceil (2\gamma + d)^{-1} \log n \rceil$ where for $x \in \mathbb{R}$ $\lceil x \rceil$ is the smallest integer greater or equal to x. Since ℓ_0 belongs to $\mathcal{L}_n$ for n large enough, Theorem 3 implies that we only have to bound the two following quantities:

$$\max_{\ell' \geq \ell_0} \|\mathbf{E} \hat{f}_{\ell'} - f\|_2 \quad \text{and} \quad \frac{\|K_{m_n(\ell_0)}\|_2^d}{\sqrt{nh^d(\ell_0)}}.$$

Defining $h_0 = n^{-\frac{1}{2\gamma+d}}$ we have:

$$e^{-1}h_0 < h(\ell_0) \le h_0 \quad \text{and} \quad m_n(\ell_0) \le \gamma + d/2 + \frac{1}{2} \tag{31}$$

This, implies:

$$\|K_{m_n(\ell_0)}\|_2 \le A\left(\gamma + d/2 + \frac{3}{2}\right)^B. \tag{32}$$

Using (31) and (32), we obtain that $\frac{\|K_{m_n(\ell_0)}\|_2^d}{\sqrt{nh^d(\ell_0)}} \le Cn^{-\gamma/(2\gamma+d)}$, where C depends on A, B and γ. It remains to bound the bias term. Set $\ell' \ge \ell_0$. Remark that: $m_n(\ell') \le m_n(\ell_0)$, $h(\ell') \le h(\ell_0)$ and $\|K_{m_n(\ell')}\|_\infty \le A\left(\gamma + d/2 + \frac{3}{2}\right)^B$. We consider two cases.

Case 1   Assume that $m_n(\ell') \ge \lfloor\gamma\rfloor$. Using Proposition 1 we obtain that

$$\|\mathbf{E}\hat{f}_{\ell'} - f\|_2 \le C(\gamma, R)\|K_{m_n(\ell')}\|_\infty^d L(h(\ell'))^\gamma \le Cn^{-\gamma/(2\gamma+d)}$$

where C depends on L, γ, R, τ, κ.

Case 2   Assume now that $m_n(\ell') < \lfloor\gamma\rfloor$. Define $\gamma' = m_n(\ell') + 1 \le \gamma$ and note that there exists L' that depends on α, L and $\mathcal{D}$ such that $f \in \mathcal{H}_\mathcal{D}(\gamma', L')$. Using Proposition 1 we have:

$$\|\mathbf{E}\hat{f}_{\ell'} - f\|_2 \le C(R, \gamma')\|K_{m_n(\ell')}\|_\infty^d L(h(\ell'))^{\gamma'} \le C\exp\left(-\ell'(m_n(\ell')+1)\right)$$

$$\le C\exp\left(-\ell'\left(\frac{\log n}{2\ell'} + \frac{1}{2}\right)\right) \le Cn^{-1/2}$$

where C depends on L, γ, R, τ and κ. Theorem follows.

7.5. *Proof of lemma 5*

We have

$$\Gamma^2(K,h) = h^d \sum_{i=1}^{\kappa} \int_{\mathcal{D}_i} \int_{\mathcal{D}} |\mathbf{K}_h \circ A_i(u-x)|^2 f(u)dudx$$

$$\le \sum_{i=1}^{\kappa} \int_{[0,1]^d} \int_{\mathcal{D}} \prod_{j=1}^{d} K^2(v_j)f(u)dudv \le \kappa\|K\|_2^{2d}. \tag{33}$$

We have

$$\mathbf{E}\|\xi_{K,h}\|_2^2 = h^d\left\{\int_{\mathcal{D}} \operatorname{Var}\mathbf{K}_h \circ A_x(X_1-x)dx + 2\sum_{k=1}^{n-1} \frac{(n-k)}{n}\mathfrak{c}(k)\right\}$$

where

$$\mathfrak{c}(k) = \int_{\mathcal{D}} \operatorname{Cov}\left(\mathbf{K}_h \circ A_x(X_1-x), \mathbf{K}_h \circ A_x(X_{k+1}-x)\right)dx.$$

The variance term can be easily bounded since

$$\int_{\mathcal{D}} \text{Var} \, \mathbf{K}_h \circ A_x(X_1 - x) dx \leq \int_{\mathcal{D}} \mathbf{E} |\mathbf{K}_h \circ A_x(X_1 - x)|^2 dx = \frac{\Gamma^2(K, h)}{h^d}. \quad (34)$$

Concerning the covariance terms, on one hand we bound, using Assumption 3 the term $\mathfrak{c}(k)$ by

$$\int_{\mathcal{D}} \mathbf{E} |\mathbf{K}_h \circ A_x(X_1 - x) \mathbf{K}_h \circ A_x(X_{k+1} - x)| dx + \int_{\mathcal{D}} \left(\mathbf{E} |\mathbf{K}_h \circ A_x(X_1 - x)| \right)^2 dx$$

$$\leq \sum_{i=1}^{\kappa} \int_{\mathcal{D}_i} \int_{\mathcal{D}} |\mathbf{K}_h \circ A_x(u - x)| |\mathbf{K}_h \circ A_x(v - x)| f_k(u, v) du dv dx$$

$$+ \int_{\mathcal{D}} \left(\mathbf{E} |\mathbf{K}_h \circ A_x(X_1 - x)| \right)^2 dx$$

$$\leq f_\infty \kappa \|K\|_1^{2d} + \sum_{i=1}^{\kappa} \int_{\mathcal{D}_i} \mathbf{E} |\mathbf{K}_h \circ \mathbf{A}_i(X_1 - x)| \int_{\mathcal{D}} |\mathbf{K}_h \circ \mathbf{A}_i(u - x)| f(u) du dx$$

$$\leq 2\kappa f_\infty \|K\|_1^{2d} \leq 2\kappa f_\infty \|K\|_2^{2d}. \quad (35)$$

On the other hand, using Lemma 9 in Appendix C, we get:

$$\mathfrak{c}(k) \leq \frac{2\kappa \|K\|_2^{2d} c \rho^k}{h^d}. \quad (36)$$

Combining (35) with (36), we obtain that for any $a \in (0, 1)$

$$\sum_{k=1}^{n-1} \mathfrak{c}(k) \leq \frac{2\kappa \|K\|_2^{2d} f_\infty^{1-a} (c\rho)^a}{h^{da}(1 - \rho^a)}. \quad (37)$$

Thus, using (33), (34) and (37) with $a = 1/2$, we get

$$\mathbf{E} \|\xi_{K,h}\|_2^2 \leq \Gamma^2(K, h) + \frac{4\kappa \|K\|_2^{2d} \sqrt{f_\infty} \sqrt{c\rho}}{(1 - \sqrt{\rho})} h^{d/2} \leq \kappa \|K\|_2^{2d} \left(1 + \frac{4\sqrt{f_\infty} \sqrt{c\rho}}{1 - \sqrt{\rho}} h^{d/2} \right)$$

which concludes the proof of (15) and (16). Finally (17) follows from previous inequality and from the fact that $h \leq \overline{h}_n$.

7.6. Proof of lemma 7

Define $Y^{*1} = \|\xi_{K,h}^{*(1)}\|_2$. Using duality arguments and Banach-Alaoglu theorem, there exists a countable set $\Lambda = (\lambda_k)_{k \in \mathbb{N}}$ of functions such that $\|\lambda_k\|_2 \leq 1$ and:

$$Y^{*1} = \sup_{k \in \mathbb{N}} \int_{\mathcal{D}} \lambda_k(t) \xi_{K,h}^{*(1)}(t) dt$$

$$= \sup_{k \in \mathbb{N}} \frac{1}{\sqrt{n/2}} \sum_{r=1}^{p_n} \sum_{s=1}^{q_n} \left(g_{\lambda_k}(X_{2(r-1)q_n+s}^*) - \mathbf{E} g_{\lambda_k}(X_{2(r-1)q_n+s}^*) \right)$$

$$= \sup_{k \in \mathbb{N}} \frac{1}{\sqrt{p_n}} \sum_{r=1}^{p_n} \left(g_{\lambda_k, q_n}(\vec{U}_{r,1}^*) - \mathbf{E} g_{\lambda_k, q_n}(\vec{U}_{r,1}^*) \right),$$

where for any $\lambda \in \Lambda$, we define

$$g_{\lambda, q_n}(x_1, \ldots, x_{q_n}) = \frac{\sqrt{2} h^{d/2}}{\sqrt{q_n}} \int_{\mathcal{D}} \lambda(t) \sum_{s=1}^{q_n} \mathbf{K}_h \circ A_t(x_s - t) dt$$

and

$$\overline{g}_{\lambda, q_n}(x_1, \ldots, x_{q_n}) = g_{\lambda, q_n}(x_1, \ldots, x_{q_n}) - \mathbf{E} g_{\lambda, q_n}(X_1, \ldots, X_{q_n})$$

Fix $\lambda \in \Lambda$. We then have $\|\overline{g}_{\lambda, q_n}\|_\infty \leq 2\|g_{\lambda, q_n}\|_\infty$ and

$$|g_{\lambda, q_n}(x_1, \ldots, x_{q_n})| \leq \sqrt{2} h^{d/2} \frac{1}{\sqrt{q_n}} \sum_{j=1}^{\kappa} \left| \int_{\mathcal{D}_j} \lambda(t) \sum_{s=1}^{q_n} \mathbf{K}_h \circ A_j(x_s - t) dt \right|$$

$$\leq \sqrt{2} h^{d/2} \kappa \sqrt{q_n} \sup_{j=1,\ldots,\kappa} \|\lambda \star (\mathbf{K}_h \circ A_j)\|_\infty$$

$$\leq \sqrt{2} h^{d/2} \kappa \sqrt{q_n} h^{-d/2} \|K\|_2^d \quad \leq \sqrt{2} \kappa \sqrt{q_n} \|K\|_2^d.$$

This implies that $\|\overline{g}_{\lambda, q_n}\|_\infty \leq \mathfrak{b}$ where $\mathfrak{b} = 2\sqrt{2} \kappa \sqrt{q_n} \|K\|_2^d$. Moreover:

$$\mathbf{E}\left(\overline{g}_{\lambda, q_n}^2(\vec{U}_{r,1}^*) \right) = \frac{2h^d}{q_n} \mathrm{Var}\left(\sum_{s=1}^{q_n} \int_{\mathcal{D}} \lambda(t)(\mathbf{K}_h \circ A_t)(X_{2(r-1)q_n + s}^* - t) dt \right)$$

$$\leq \frac{2\kappa^2 h^d}{q_n} \sup_{j=1,\ldots,\kappa} \mathrm{Var}\left(\sum_{s=1}^{q_n} \int_{\mathcal{D}_j} \lambda(t)(\mathbf{K}_h \circ \mathbf{A}_j)(X_s - t) dt \right).$$

Fix $j \in \{1, \ldots, \kappa\}$. Denote $\Psi = \lambda \star (\mathbf{K}_h \circ \mathbf{A}_j)$. Using these notations we have:

$$\mathrm{Var}\left(\sum_{s=1}^{q_n} \int_{\mathcal{D}_j} \lambda(t)(\mathbf{K}_h \circ \mathbf{A}_j)(X_s - t) dt \right) = \mathrm{Var}\left(\sum_{s=1}^{q_n} \Psi(X_s) \right)$$

$$\leq 4q_n \left(\mathbf{E} \Psi^4(X_1) \right)^{1/2} \left(2 \sum_{k=1}^{+\infty} (k+1)\beta(k) \right)^{1/2}.$$

Last line is deduced from the third inequation in Lemma 9 in Appendix C with $p = p' = 2$. Using Young's inequality for convolution products with $r = 4$, $p = 4/3$ and $q = 2$, we obtain:

$$\mathbf{E} \Psi^4(X_1) \leq f_\infty \|\Psi\|_4^4 \leq f_\infty \left(\|\lambda\|_2 \cdot \|\mathbf{K}_h \circ \mathbf{A}_j\|_{4/3} \right)^4.$$

Since $\|\lambda\|_2 = 1$, this leads to $\mathbf{E} \Psi^4(X_1) \leq f_\infty h^{-d} \|K\|_{4/3}^{4d} \leq f_\infty h^{-d} \|K\|_2^{4d}$.

Finally we obtain the following bound:

$$\mathbf{E}\left(\overline{g}_{\lambda, q_n}^2(\vec{U}_{r,1}^*) \right) \leq \mathfrak{a} \quad \text{where} \quad \mathfrak{a} = 8\kappa^2 \left(2 \sum_{k=1}^{+\infty} (k+1)\beta(k) \right)^{1/2} f_\infty^{1/2} \|K\|_2^{2d} h^{d/2}.$$

Using Lemma 8 in Appendix C, with $X_{r,\lambda} = \overline{g}_{\lambda,q_n}(\vec{U}_{r,1}^*)\frac{\sqrt{p_n}}{\mathfrak{b}}$ we get:

$$\mathbf{P}(Y^{*1} - \mathbf{E}Y^{*1} > \delta\|K\|_2^d + x) \leq \exp\left(-\frac{(x + \delta\|K\|_2^d)^2}{2\mathfrak{a} + \frac{4\mathfrak{b}\mathbf{E}Y^{*1}}{\sqrt{p_n}} + \frac{2\mathfrak{b}}{3\sqrt{p_n}}(x + \delta\|K\|_2^d)}\right).$$

Lemma 6 with Lemma 5 implies that

$$\mathbf{E}Y^{*1} \leq \sqrt{2}\sqrt{\kappa}\|K\|_2^d\sqrt{1 + C(c,\rho,f_\infty)h^{d/2}}.$$

This, combined with basic calculations, implies that there exist positive constants $\widetilde{C}$, $\widetilde{C}_0$, $\widetilde{C}_1$, $\widetilde{C}_2$ that depend only on c_1, δ, κ, c, ρ and f_∞ such that:

$$\mathbf{P}(Y^{*1} - \mathbf{E}Y^{*1} > \delta\|K\|_2^d + x)$$

$$\leq \exp\left(-\frac{\widetilde{C}x^2}{(\|K\|_2^{2d} + x\|K\|_2^d)(h^{d/2} + q_n n^{-1/2})}\right)\exp\left(-\frac{\widetilde{C}}{h^{d/2} + q_n n^{-1/2}}\right)$$

$$\leq \widetilde{C}_0\exp\left(-\frac{\widetilde{C}_1 x^2 \bar{h}_n^{-d/2}}{\|K\|_2^{2d} + x\|K\|_2^d}\right)\exp\left(-\widetilde{C}_2\bar{h}_n^{-d/2}\right).$$

Last line follows from the fact that, for n large enough we have $q_n n^{-1/2} \leq \bar{h}_n^{d/2}$. Similar arguments can be applied to the study of Y^{*2}.

Appendix A: Regular domains satisfy Assumption 5

Step 1 Walther (1999) proved that the domain $\mathcal{D}$ is r_0-regular if, and only if, the following assumption is satisfied: $\partial\mathcal{D}$ is a 1-dimensional C^1 submanifold in $\mathbb{R}^2$ with the outward-pointing unit normal vector $n(a)$ at $a \in \partial\mathcal{D}$ satisfying the Lipschitz condition:

$$|n(a) - n(b)|_2 \leq \frac{1}{r_0}|a - b|_2, \qquad \forall a, b \in \partial\mathcal{D}.$$

Moreover, if x belongs to $\mathcal{D}$ is such that $\inf\{|x - a| : a \in \partial\mathcal{D}\} \leq 2r_0$ then x projects uniquely onto $\partial\mathcal{D}$. We then denote by $a(x)$ this projection which satisfies:

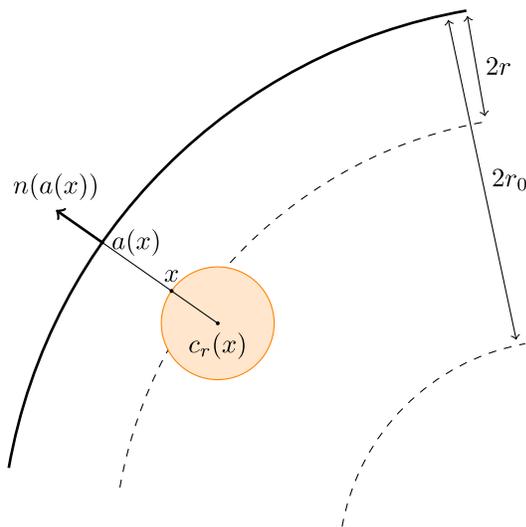$$a(x) - x = |a(x) - x| \cdot n(a(x)).$$

Now we define, for any $0 < r \leq r_0$:

$$\mathcal{C}(r) = \{x \in \mathcal{D} : \inf\{|x - a| : a \in \partial\mathcal{D}\} \leq 2r\} = \bigcup_{a \in \partial\mathcal{D}}(x_a^r + \mathbb{D}_r).$$

Using these notations and the triangle inequality we remark that, for any $0 < r \leq r_0/2$ and any $x \in \mathcal{C}(r)$, we have:

$$x \in \left(c_r(x) + \mathbb{D}_r\right) \subseteq \mathcal{C}(r_0) \subset \overline{\mathcal{D}}$$
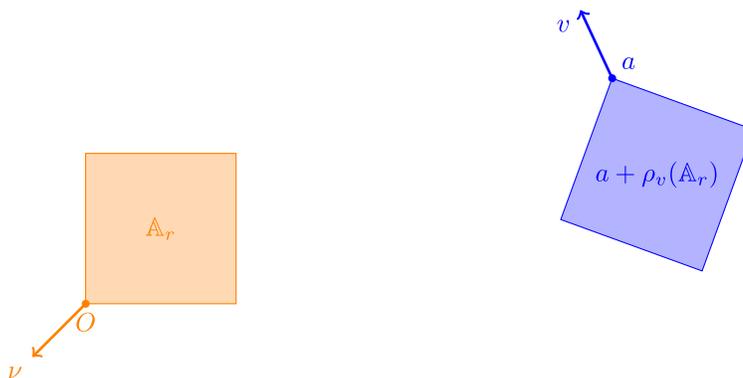
where $c_r(x) = x - rn(a(x))$. The following figure illustrates this property:

Step 2 Define

$$\mathbb{A}_r = \left[0, r\frac{\sqrt{2}}{2}\right]^2$$

and consider the vector $\nu = (-1, -1)/\sqrt{2} \in \mathbb{S}^1$. Note that, for any $v \in \mathbb{S}^1$, there exists a unique rotation $\rho_v(\cdot) \in \mathrm{GL}_2(\mathbb{R})$ such that $\rho_v(\nu) = v$. In the following figure the sets $\mathbb{A}_r$ and $a + \rho_v(\mathbb{A}_r)$ are represented for some point a in $\mathbb{R}^2$ and vector v in $\mathbb{S}^1$.

Now, assume that for any $a \in \partial\mathbb{D}_r = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = r^2\}$, the outward-pointing unit vector is denoted by $v(a)$. Then $a + \rho_{v(a)}(\mathbb{A}_r) \subseteq \mathbb{D}_r$ and moreover $a + \rho_v(\mathbb{A}_r) \subseteq \mathbb{D}_r$ as soon as v is a unit vector such that the angle θ between v and $v(a)$ is less than or equals to $\theta_0 = \arccos(\sqrt{2}/4) - \pi/4$. This is the case if $|v(a) - v| \leq 2\sin(\theta_0/2) = \delta_0$. This result generalizes to any ball centered at $c \in \mathbb{R}^2$ in such a way: if $a \in \partial(c + \mathbb{D}_r)$ then

$$a \in (a + \rho_v(\mathbb{A}_r)) \subseteq (c + \mathbb{D}_r)$$

as soon as v is a unit vector such that $|v - v(a)| \leq \delta_0$. The following figure illustrates this property.

Step 3 Set $0 < r < r_0 \delta_0 < r_0/2$. Our goal is to prove the following property: for any $x \in \mathcal{C}(r/6)$, if y satisfies both $y \in \mathcal{C}(r/6)$ and $|x - y| \leq r/3$, then

$$y \in \big(y + \rho_{n(a(x))}(\mathbb{A}_r)\big) \subseteq \mathcal{D}.$$

Note that

$$|a(x) - a(y)| < |a(x) - x| + |x - y| + |y - a(y)|$$
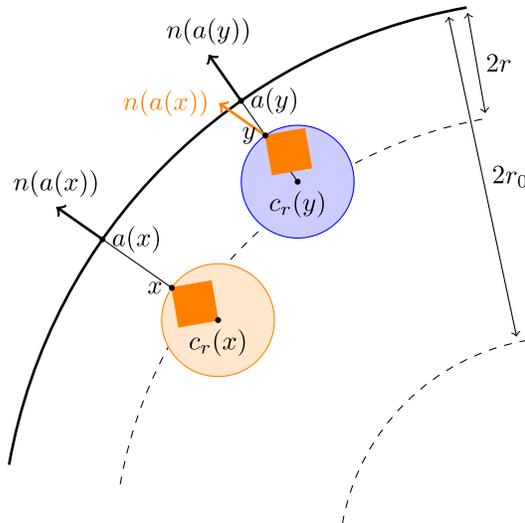$$< 2r/6 + r/3 + 2r/6 \leq r$$

which implies (using the Lipschitz property of the normal $n(\cdot)$) that:

$$|n(a(x)) - n(a(y))| < \frac{r}{r_0} \leq \delta_0.$$

Using the previous steps combined with this result, since $y \in \partial(c_r(y) + \mathbb{D}_r)$, we have

$$y \in \big(y + \rho_{n(a(x))}(\mathbb{A}_r)\big) \subseteq \big(c_r(y) + \mathbb{D}_r\big) \subseteq \mathcal{D}.$$

Last inclusion is straightforward and comes from the triangle inequality. This property is illustrated below:

Step 4 Set $0 < r < r_0\delta_0 < r_0/2$ and define for any $x \in \overline{\mathcal{C}(r/6)}$ the set

$$V(x) = \{y \in \mathcal{C}(r/6) : |x - y| < r/3\},$$

which is an open neighborhood of x. Note also that, using the property proved at the previous step, for any $y \in V(x)$ we have:

$$y + \rho_{n(a(x))}(\mathbb{A}_r) \subseteq \mathcal{D}$$

Finally $\{V(x) : x \in \overline{\mathcal{C}(r/6)}\}$ is a covering of the compact set $\overline{\mathcal{C}(r/6)}$ by open sets. Thus, there exists a finite number of points $x_1, \ldots, x_N$ such that $\{V(x_n) : n = 1, \ldots, N\}$ is also a covering of $\overline{\mathcal{C}(r/6)}$. The result follows easily.

Appendix B: Simple polygons satisfy Assumption 5

Chazelle (1991) proved that any simple polygon can be triangulated (in linear time with respect to the number of vertices). In view of this result, our problem boils down to the case of a triangle. Let ABC be a triangle and let P_ε and Q_ε such that $AP_\varepsilon Q_\varepsilon$ and ABC are similar with homothetic ratio $1 - 2\varepsilon > 0$. Then for any point x in $AP_\varepsilon Q_\varepsilon$, the parallelogram generated by the vectors $\vec{u} = \varepsilon\overrightarrow{AB}$ and $\vec{v} = \varepsilon\overrightarrow{AC}$ with its origin located at x is included into the triangle ABC. This property is illustrated in the following figure.
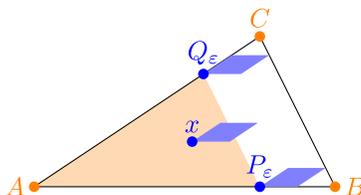
One can also prove that there exist S_ε, T_ε, U_ε and V_ε such that the triangles $BS_\varepsilon T_\varepsilon$ and $CU_\varepsilon V_\varepsilon$ satisfy similar properties. Now, it is easily seen that, for ε small enough, the union of the triangles $AP_\varepsilon Q_\varepsilon$, $BS_\varepsilon T_\varepsilon$ and $CU_\varepsilon V_\varepsilon$ equals to ABC (this property is illustrated below with $\varepsilon = 0.125$).
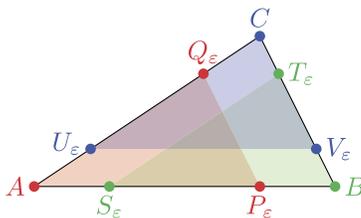
Using straightforward arguments, it can be proven that ABC satisfies Assumption 5 with $\kappa = 3$ and $r \leq \varepsilon\sqrt{2\,\text{Area}(ABC)}$. This ends the proof.

Appendix C: Classical lemmas

The following Lemma is the Bousquet inequality from Boucheron et al. (2004) (see also Theorem 12.5 in Boucheron et al., 2013).

Lemma 8. *Let $X_1, \ldots, X_n$ be independent identically distributed random vectors. Let $\mathcal{S}$ be a countable set of functions. Denote for $s \in \mathcal{S}$ $X_{i,s} = s(X_i)$ and $Z = \sup_{s \in \mathcal{S}} \sum_{i=1}^n s(X_i)$. Assume that for all $i = 1, \ldots, n$ and $s \in \mathcal{S}$ $\mathbf{E}X_{i,s} = 0$, and that $X_{i,s} \leq 1$. Assume also that $v = 2\mathbf{E}Z + \sup_{s \in \mathcal{S}} \sum_{i=1}^n \mathbf{E}(X_{i,s})^2 < \infty$. Then we have for all $t > 0$*

$$\mathbf{P}\left(Z - \mathbf{E}Z \geq t\right) \leq \exp\left\{-\frac{t^2}{2(v + \frac{t}{3})}\right\}.$$

The following lemma comes from Viennet (1997). See in particular Lemma 4.2.

Lemma 9. *Let $(\chi_i)_{i \in \mathbb{Z}}$ be a stationary β-mixing process with rate $(\beta(k))_{k \geq 0}$. Let ϕ a measurable function such that $\mathbf{E}\left(\phi^2(\chi_0)\right) < \infty$. There exists a sequence of random variables $(b_k(\chi_0))_{k \geq 0}$ such that $0 \leq b_k(\chi_0) \leq 1$ and $\mathbf{E}(b_k(\chi_0)) = \beta(k)$ that satisfies*

$$\mathrm{Cov}(\phi(\chi_0), \phi(\chi_k)) \leq 2\mathbf{E}\left\{b_k(\chi_0)\phi^2(\chi_0)\right\}$$

and

$$\mathrm{Var}\left(\sum_{i=1}^n \phi(\chi_i)\right) \leq 4n\mathbf{E}\left\{\left(\sum_{k=0}^n b_k(\chi_0)\right)\phi^2(\chi_0)\right\}.$$

For $p \geq 1$, p' such that $\frac{1}{p} + \frac{1}{p'} = 1$, assume $\sum_{k \geq 0}(k+1)^{p-1}\beta(k) < \infty$. Let ϕ a measurable function such that $\mathbf{E}(\phi^{2p'}(\chi_0)) < \infty$. We have

$$\mathrm{Var}\left(\sum_{i=1}^n \phi(\chi_i)\right) \leq 4n\left(\mathbf{E}\left(\phi^{2p'}(\chi_0)\right)\right)^{1/p'}\left(p\sum_{k \geq 0}(k+1)^{p-1}\beta(k)\right)^{1/p}.$$

References

Akakpo, N. (2012). Adaptation to anisotropy and inhomogeneity via dyadic piecewise polynomial selection. *Math. Methods Statist.*, 21(1):1–28. MR2901269

Arias-Castro, E. and Rodríguez-Casal, A. (2017). On estimating the perimeter using the alpha-shape. *Ann. Inst. Henri Poincaré Probab. Stat.*, 53(3):1051–1068. MR3689960

Barron, A., Birgé, L., and Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields*, 113(3):301–413. MR1679028

Baudry, J.-P., Maugis, C., and Michel, B. (2018). Slope heuristics: overview and implementation. *Stat. Comput.*, 22(2):455–470. MR2865029

Berbee, H. C. P. (1979). *Random walks with stationary increments and renewal theory*, volume 112 of *Mathematical Centre Tracts*. Mathematisch Centrum, Amsterdam. MR0547109

Bertin, K., El Kolei, S., and Klutchnikoff, N. (2018). Adaptive density estimation on bounded domains. *Accepted for publication in Annales de l'institut Henri Poincaré: Probabilités et Statistiques*. MR4029144

Birgé, L. and Rozenholc, Y. (2006). How many bins should be put in a regular histogram. *ESAIM Probab. Stat.*, 10:24–45. MR2197101

Bossy, M., Gobet, E., and Talay, D. (2004). A symmetrized Euler scheme for an efficient approximation of reflected diffusions. *J. Appl. Probab.*, 41(3):877–889. MR2074829

Botev, Z. I., Grotowski, J. F., and Kroese, D. P. (2010). Kernel density estimation via diffusion. *Ann. Statist.*, 38(5):2916–2957. MR2722460

Boucheron, S., Lugosi, G., and Bousquet, O. (2004). Concentration inequalities. In *Advanced Lectures on Machine Learning*, pages 208–240. Springer.

Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration inequalities*. Oxford University Press, Oxford. A nonasymptotic theory of independence, With a foreword by Michel Ledoux. MR3185193

Bouezmarni, T. and Rombouts, J. V. K. (2010). Nonparametric density estimation for multivariate bounded data. *J. Statist. Plann. Inference*, 140(1):139–152. MR2568128

Castellan, G. (2000). Sélection d'histogrammes à l'aide d'un critère de type Akaike. *C. R. Acad. Sci. Paris Sér. I Math.*, 330(8):729–732. MR1763919

Castellan, G. (2003). Density estimation via exponential model selection. *IEEE Trans. Inform. Theory*, 49(8):2052–2060. MR2004713

Cattiaux, P., León, J. R., and Prieur, C. (2017). Invariant density estimation for a reflected diffusion using an Euler scheme. *Monte Carlo Methods Appl.*, 23(2):71–88. MR3657943

Chazelle, B. (1991). Triangulating a simple polygon in linear time. *Discrete Comput. Geom.*, 6(5):485–524. MR1115104

Chen, S. X. (1999). Beta kernel estimators for density functions. *Comput. Statist. Data Anal.*, 31(2):131–145. MR1718494

Cholaquidis, A., Fraiman, R., Mordecki, E., and Papalardo, C. (2016). Level sets and drift estimation for reflected brownian motion with drift. *arXiv preprint* arXiv:1611.09588.

Cline, D. and Hart, J. (1991). Kernel estimation of densities with discontinuities or discontinuous derivatives. *Statistics*, 22(1):69–84. MR1097362

Comte, F., Prieur, C., and Samson, A. (2017). Adaptive estimation for stochastic damping Hamiltonian systems under partial observation. *Stochastic Process. Appl.*, 127(11):3689–3718. MR3707242

Delyon, B. and Juditsky, A. (1996). On minimax wavelet estimators. *Appl. Comput. Harmon. Anal.*, 3(3):215–228. MR1400080

Doukhan, P. (1994). Mixing. In *Mixing*, pages 15–23. Springer. MR1312160

Goldenshluger, A. and Lepski, O. (2014). On adaptive minimax density estimation on R^d. *Probab. Theory Related Fields*, 159(3-4):479–543. MR3230001

Ibragimov, I. A. (1962). Some limit theorems for stationary processes. *Theory of Probability & Its Applications*, 7(4):349–382. MR0148125

Ibragimov, I. A. and Hasminskii, R. Z. (1980). An estimate of the density of a distribution. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)*,

98:61–85, 161–162, 166. Studies in mathematical statistics, IV. MR0591862

Jones, M. C. (1993). Simple boundary correction for kernel density estimation. *Statistics and Computing*, 3(3):135–146. MR1125651

Lejeune, M. and Sarda, P. (1992). Smooth estimators of distribution and density functions. *Comput. Statist. Data Anal.*, 14(4):457–471. MR1192215

Marron, J. S. and Ruppert, D. (1994). Transformations to reduce boundary bias in kernel density estimation. *J. Roy. Statist. Soc. Ser. B*, 56(4):653–671. MR1293239

Marshall, J. C. and Hazelton, M. L. (2010). Boundary kernels for adaptive density estimators on regions with irregular boundaries. *J. Multivariate Anal.*, 101(4):949–963. MR2584911

Müller, H.-G. (1991). Smooth optimum kernel estimators near endpoints. *Biometrika*, 78(3):521–530. MR1130920

Müller, H.-G. and Stadtmüller, U. (1999). Multivariate boundary kernels and a continuous least squares principle. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 61(2):439–458. MR1680306

Schuster, E. F. (1985). Incorporating support constraints into nonparametric estimators of densities. *Comm. Statist. A—Theory Methods*, 14(5):1123–1136. MR0797636

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London. MR0848134

Tshipa, A., Valls-Fox, H., Fritz, H., Collins, K., Sebele, L., Mundy, P., and Chamaillé-Jammes, S. (2017). Partial migration links local surface-water management to large-scale elephant conservation in the world's largest transfrontier conservation area. *Biological Conservation*, 215:46–50.

Tsybakov, A. (2009). *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats. MR2724359

Valls-Fox, H., De Garine-Wichatitsky, M., Fritz, H., and Chamaillé-Jammes, S. (2018). Resource depletion versus landscape complementation: habitat selection by a multiple central place forager. *Landscape Ecology*, 33(1):127–140.

Viennet, G. (1997). Inequalities for absolutely regular sequences: application to density estimation. *Probab. Theory Related Fields*, 107(4):467–492. MR1440142

Volkonskii, V. and Rozanov, Y. A. (1959). Some limit theorems for random functions. i. *Theory of Probability & Its Applications*, 4(2):178–197. MR0121856

Walther, G. (1999). On a generalization of Blaschke's rolling theorem and the smoothing of surfaces. *Math. Methods Appl. Sci.*, 22(4):301–316. MR1671447

Wikelski, M. and Kays, R. (2018). Movebank: archive, analysis and sharing of animal movement data. *World Wide Web electronic publication*.