

## INTEGRAL ESTIMATION BASED ON MARKOVIAN DESIGN

ROMAIN AZAÏS,\* *Inria Nancy and Institut Élie Cartan de Lorraine*

BERNARD DELYON,\*\* *Univ Rennes, CNRS, IRMAR*

FRANÇOIS PORTIER,\*\*\*, *Télécom ParisTech and University of Paris-Saclay*

### Abstract

Suppose that a mobile sensor describes a Markovian trajectory in the ambient space and at each time the sensor measures an attribute of interest, e.g. the temperature. Using only the location history of the sensor and the associated measurements, we estimate the average value of the attribute over the space. In contrast to classical probabilistic integration methods, e.g. Monte Carlo, the proposed approach does not require any knowledge of the distribution of the sensor trajectory. We establish probabilistic bounds on the convergence rates of the estimator. These rates are better than the traditional ‘root  $n$ ’-rate, where  $n$  is the sample size, attached to other probabilistic integration methods. For finite sample sizes, we demonstrate the favorable behavior of the procedure through simulations and consider an application to the evaluation of the average temperature of oceans.

*Keywords:* Integral approximation; Markov chain; Nummelin splitting technique; invariant density estimation; kernel smoothing

2010 Mathematics Subject Classification: Primary 62M05  
Secondary 60J05

### 1. Introduction

In recent decades, climate scientists have been interested in the evolution of different physical attributes of the Earth in order to quantify the effects of global warming. For instance, attributes such as temperature, acidity and salinity of the oceans, or the concentration of greenhouse gases in the atmosphere are important indicators of global warming. These measurements are often obtained by sensors placed on drifting buoys in the oceans or weather balloons in the atmosphere, each describing an area or a volume. Whenever the data have been collected, a crucial quantity is the average of the measurements over a given space. As the sensors are eventually subjected to unpredictable effects such as marine currents or winds, their trajectories are modeled as random sequences. Our approach is concerned with trajectories satisfying the Markov property, meaning roughly that the distribution of the location at time  $t + 1$  is fully determined by the location at  $t$  and an independent random noise. For the sake of realism, the underlying transition probability and the invariant probability measure associated to the Markov chain are supposed unknown. In summary, we evaluate the average value of a physical quantity over some space when the measurements are taken along the path of a Markov chain.

---

Received 4 November 2016; revision received 4 June 2018.

The supplementary material for this article can be found at <http://doi.org/10.1017/apr.2018.38>.

\* Postal address: Team BIGS, Inria Nancy – Grand-Est Research Centre, 615 rue du Jardin Botanique, 54600 Villers-lès-Nancy, France.

\*\* Postal address: CNRS, IRMAR – UMR 6625, University of Rennes 1, F-35000 Rennes, France.

Email address: [bernard.delyon@univ-rennes1.fr](mailto:bernard.delyon@univ-rennes1.fr)

\*\*\* Postal address: LTCI, Télécom ParisTech, 46 rue Barrault, 75634 Paris, Cedex 13, France.

More formally, let  $Q$  denote a given bounded and open set of  $\mathbb{R}^d$  and suppose that  $\varphi: Q \rightarrow \mathbb{R}$  represents a physical attribute to each location in  $Q$ , e.g. the temperature in the air over a volume or the wind velocity on the sea over a surface. For simplicity, the Lebesgue measure of  $Q$  is set to be 1. Hence, we are interested in the average value of  $\varphi$  over  $Q$ , defined as

$$I_0 = \int_Q \varphi(x) \, dx.$$

In most examples of interest, the function  $\varphi$  is unknown and only some images of the function are obtained from measurement instruments. Suppose that we observe  $n \in \mathbb{N}^*$  points from the trajectory of a time-homogeneous Harris recurrent Markov chain  $X_1, X_2, \dots$  (see Meyn and Tweedie (2009)) with state space  $\mathcal{E} \supseteq Q$ . Suppose, moreover, that we know the associated images by the map  $\varphi$ , i.e.  $\varphi(X_1), \dots, \varphi(X_n)$ . Let  $\pi$  denote the density of the stationary measure of the chain. If  $\pi$  were known, it would be tempting to compute the Monte Carlo estimator of  $I_0$ , i.e.

$$\hat{I}_{\text{mc}} = n^{-1} \sum_{i=1}^n \frac{\varphi(X_i)}{\pi(X_i)},$$

which satisfies, under standard conditions (see Meyn and Tweedie (2009, Chapter 17)), a central limit theorem, i.e.  $n^{1/2}(\hat{I}_{\text{mc}} - I_0)$  converges weakly to a centered Gaussian distribution. As the previous estimator requires knowledge of  $\pi$ , which is not the case in our framework, we instead consider the following kernel smoothing estimator of  $I_0$ :

$$\hat{I}_{\text{ks}} = n^{-1} \sum_{i=1}^n \frac{\varphi(X_i)}{\hat{\pi}(X_i)},$$

where  $\hat{\pi}$  is the classical kernel estimator of the density (see Silverman (1986)), i.e.

$$\hat{\pi}(x) = (nh_n^d)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right), \quad x \in \mathbb{R}^d,$$

with  $K: \mathbb{R}^d \rightarrow \mathbb{R}$ , a symmetric function, called the kernel, that integrates to 1, and  $(h_n)_{n \in \mathbb{N}^*}$ , a sequence of positive numbers, called the bandwidth, that goes to 0 as  $n \rightarrow +\infty$ .

As the stationary measure is unknown, we cannot rely on Monte Carlo integration techniques, often used in simulation-based approximation such as importance sampling, control variates, or Metropolis–Hastings integration. We refer the reader to Evans and Swartz (2000) and Robert and Casella (2004) for more on integral approximation techniques.

The estimator  $\hat{I}_{\text{ks}}$  was introduced by Delyon and Portier (2016). The authors established bounds on the rate of convergence and in probability in the case of an independent and identically distributed (i.i.d.) sequence  $X_1, X_2, \dots$ . Their main observation was that the convergence rate of  $\hat{I}_{\text{ks}}$  to  $I_0$  is faster than the convergence rate of the Monte Carlo estimator  $\hat{I}_{\text{mc}}$  to  $I_0$  (even though  $\hat{I}_{\text{mc}}$  requires knowledge of  $\pi$ ). In contrast to standard Monte Carlo methods, the main ingredient of their proposal is the evaluation of the image of the design points by the kernel estimator, i.e.  $\hat{\pi}(X_1), \dots, \hat{\pi}(X_n)$ . These quantities capture an essential information: the isolation of each point. Basically, the more isolated  $X_i$ , the larger the weight  $1/\hat{\pi}(X_i)$  (and conversely). Hence, these weights realize an adaptation to the design points by attributing more weight to the lonely points.

Our main theoretical objective is to extend the results of Delyon and Portier (2016) when the sequence  $(X_i)_{i \in \mathbb{N}^*}$  is a time-homogeneous Harris recurrent Markov chain. Denote by  $s$

and  $r$  the (Nikolski) regularity of the functions  $\varphi$  and  $\pi$ , respectively. For any set  $B \subset \mathcal{E}$ , let  $\tau_B$  denote the return time of the chain to  $B$ . If there exist  $A \subset \mathcal{E}$  and  $p_0 > 3$  such that

$$\sup_{x \in A} \mathbb{E}_x[\tau_A^{p_0}] < +\infty, \tag{1}$$

where  $\mathbb{E}_x$  is the expectation for the Markov chain starting at  $X_0 = x$ , and if, as  $n \rightarrow +\infty$ ,

$$\frac{nh_n^{d(p_0/p_0-1)}}{|\log(n)|} \rightarrow +\infty,$$

we show (Theorem 3), under mild additional conditions, that as  $n \rightarrow +\infty$ ,

$$\hat{I}_{ks} - I_0 = O_{\mathbb{P}}(h_n^r + n^{-1/2}h_n^s + n^{-1}h_n^{-d}).$$

This is the same convergence rate as the one provided by Delyon and Portier (2016) for i.i.d. sequences  $(X_i)_{i \in \mathbb{N}^*}$ . The previous rate is better than the rate of  $\hat{I}_{mc}$  whenever  $n^{1/2}h_n^r \rightarrow 0$  and  $n^{-1/2}h_n^{-d} \rightarrow 0$  as  $n \rightarrow +\infty$ . Taking  $h_n \propto n^{-1/(r+d)}$ , we obtain a rate in  $n^{-r/(r+d)} + n^{-1/2}n^{-s/(r+d)}$  which is negligible before  $n^{-1/2}$  if and only if  $r > d$ . Consequently, in addition to being consistent when facing Markovian design, the kernel smoothing integral estimator might give an acceleration of the rate of convergence of the Monte Carlo estimator. This acceleration is unfortunately subject to the well-known curse of the dimension as one needs  $r > d$ . In contrast, a nice feature of the method is that only mild constraints are required on the regularity of  $\varphi$ . Finally, there exists a theoretical lower bound for random integration methods (see Novak (2016, Theorem 3)) which takes the form  $n^{-1/2}n^{-s/d}$ , while our proposal achieves  $n^{-1/2}n^{-s/2(s+d)} \geq n^{-1}$ . This gap in efficiency might be explained by the design distribution which is imposed in our framework.

The mathematical proofs follow from a mixture between the Nummelin splitting technique for Markov chains (see Nummelin (1978)), Hoeffding-type decompositions for  $U$ -statistics (see van der Vaart (1998, Section 11.4)), and uniform bounds for kernel density estimators in the case of independent observations (see Einmahl and Mason (2005)). More specifically, the Nummelin splitting technique, also called regeneration theory and presented in Section 2, allows the chain to be divided into  $l_n$  independent blocks. Assumption (1) implies that  $l_n$  and  $n$  have the same order allowing us to mimic the approach of Delyon and Portier (2016) taken in the independent case.

- (i) Linearize the terms  $1/\hat{\pi}(X_1), \dots, 1/\hat{\pi}(X_n)$  by the use of a Taylor expansion. This is typically used in semi-parametric problems such as, for instance, the single-index model; see Härdle and Stoker (1989) and Roget-Vial (2003).
- (ii) Find a probabilistic bound on some degenerate  $U$ -statistic depending on the sequence  $K((X_j - X_i)/h)/h^d, (i, j) \in \{1, \dots, n\}^2$ . We follow the approach of Bertail and Cléménçon (2011) by using an Hoeffding-type decomposition based on the blocks.
- (iii) We avoid letting the values in the denominator of  $\hat{I}_{ks}$  become small by showing that  $\inf_{x \in Q} \hat{\pi}(x)$  is bounded away from 0 with large probability. In particular, we show (Theorem 2) that as  $n \rightarrow +\infty$ ,

$$\sup_{x \in \mathbb{R}^d} |\hat{\pi}(x) - \pi_{h_n}(x)| \rightarrow 0 \quad \text{in probability,}$$

where  $\pi_{h_n}(x)$  is the expectation of  $\hat{\pi}(x)$  under stationarity. We rely on empirical process theory and, more precisely, on a formulation of Talagrand’s inequality established in Einmahl and Mason (2005). To the best of the authors’ knowledge, the previous result

in the case of general time-homogeneous Markov chains is new. Consistency results (nonuniform) for time-homogeneous Markov chains can be found in Roussas (1969). In the case of mixing-type dependency, uniform convergence rates can be found in Hansen (2008).

Steps (i) and (ii) are directly developed in the proof of Theorem 3, while the consistency result (iii) is presented in Section 3.

In contrast with the framework of Delyon and Portier (2016), in which the density is needed to be continuously differentiable on  $\mathbb{R}^d$ , we include density functions that possibly jump at the boundary of  $Q$  (see the discussion before the statement of Theorem 3).

To compute  $\hat{I}_{ks}$ , the bandwidth  $h_n$  and the kernel  $K$  need to be chosen. Preliminary numerical experiments show that  $\hat{I}_{ks}$  is quite sensible for the values of  $h_n$  chosen, whereas the choice of  $K$  has no strong influence. In Delyon and Portier (2016),  $h_n$  was chosen according to both the independent points of the design and the function  $\varphi$ . In this paper, we use the multivariate plug-in bandwidth selection developed by Chacón and Duong (2010). A simulation study shows the favorable behavior of the estimator with this choice of bandwidth in various settings.

The organization of the paper is as follows. In Section 2 we present a brief overview of the regeneration approach for Markov chains. The notation and concepts introduced there will be useful in the rest of the paper. Section 3 is concerned with the uniform convergence of kernel density estimators for Markov chains. In Section 4 we provide the main theoretical statement of the paper which consists of a bound on the rate of convergence of  $\hat{I}_{ks} - I_0$ . In Section 5 we present a large simulation study as well as a real-data analysis performed from sea surface temperature data of the three major oceans. Technical details concerning regeneration-based bounds for expectations and of the initial measure, as well as the proofs of the results of Section 3, are presented in Sections A–C of the supplementary material; see Azaïs *et al.* (2018). For the sake of readability, results and equations from the supplementary material are prefixed with an asterisk.

## 2. Regeneration

In this section we give a short account of the regeneration theory, also referred to as the Nummelin splitting technique, as discovered by Athreya and Ney (1978) and Nummelin (1978), and extensively studied by Nummelin (1984) and Meyn and Tweedie (2009).

We consider a Markov chain  $X_0, X_1, X_2, \dots$  with state space  $\mathcal{E}$  and transition probabilities  $P(x, dy)$ . The notation  $\mathbb{E}_\nu$  denotes the expectation according to the chain under  $X_0 \sim \nu$ , and  $\mathbb{E}_x$  in the  $\nu = \delta_x$  case. The associated probabilities are denoted by  $\mathbb{P}_\nu$  and  $\mathbb{P}_x$ , respectively. We assume that for some set  $A$ , the hitting time

$$\tau_A = \min\{i \geq 1: X_i \in A\}$$

satisfies, for all  $x \in \mathcal{E}$ ,

$$\mathbb{P}_x(\tau_A < \infty) = 1, \tag{2}$$

$$\sup_{x \in A} \mathbb{E}_x[\tau_A] < \infty. \tag{3}$$

We assume also that for some probability measure  $\psi$ , some  $\lambda_0 > 0$ , and some  $m_0 \geq 1$ ,

$$P^{m_0}(x, B) \geq \lambda_0 \psi(B) \quad \text{for all } x \in A, B \text{ measurable.} \tag{4}$$

The previous equation means that  $A$  is a ‘petite set’ in the terminology of Meyn and Tweedie (2009, Section 5.5.2). In particular, the set  $A$  is  $\psi$ -communicating in the sense that (see

Nummelin (1984, Definition 2.2, p. 11)), for all  $x \in A$ ,  $B$  measurable,

$$\psi(B) > 0 \implies \text{there exists } m \geq 1, P^m(x, B) > 0.$$

From (2), we see that the time to reach  $A$  is finite with probability 1, and the chain is  $\psi$ -irreducible, i.e. the whole space  $\mathcal{E}$  is  $\psi$ -communicating. An irreducible Markov chain is called Harris recurrent if, for all  $B \subset \mathcal{E}$  such that  $\psi(B) > 0$  and for all  $x \in \mathcal{E}$ ,

$$\mathbb{P}_x(\{X_n \in B\} \text{ infinitely often}) = 1.$$

As a consequence of (2), for all  $x \in \mathcal{E}$ ,  $\mathbb{P}_x(\{X_n \in A\} \text{ infinitely often}) = 1$ . Starting from  $A$  and if  $\psi(B) > 0$ , we can use (4) to show that the chain reaches  $B$  with positive probability. Consequently, under (2) and (4), the chain is Harris recurrent (see Meyn and Tweedie (2009, Proposition 9.1.7) or Nummelin (1984, Proposition 4.8)). From Meyn and Tweedie (2009, Theorem 10.0.1) (see also Nummelin (1984, Corollary 5.3(ii))), the chain admits an invariant measure and (3) allows us to prove that this measure is finite.

If  $m_0 = 1$ , the regeneration theory (see below) allows us to split the chain into independent subsequences. This is obviously of great technical interest as many results can be adapted from the independent setting. The  $m_0 > 1$  case is somewhat different and we shall say a few words about it later.

When  $m_0 = 1$ ,

$$P(x, B) \geq \lambda_0 \psi(B) \quad \text{for all } x \in A, B \text{ measurable}, \tag{5}$$

i.e. each time the chain hits  $A$  it can be restarted with probability  $\lambda_0$  with the measure  $\psi$ . Note that this assumption is weaker than the well-known Doeblin condition which requires (5) to hold for every  $x \in \mathcal{E}$ . In order to make these regeneration times stopping times, the chain has to be extended and redefined as the so-called split chain  $Z_i = (X_i, Y_i)$ ,  $i = 1, 2, \dots$ , having the following transitions:

- generation of  $Y_i$ , given  $X_i$ ,

$$X_i \notin A \rightarrow Y_i = 0, \quad X_i \in A \rightarrow Y_i \sim \text{binomial}(1, \lambda_0);$$

- generation of  $X_{i+1}$ ,

$$\begin{aligned} X_i \notin A &\rightarrow X_{i+1} \sim P(X_i, dx), \\ X_i \in A, Y_i = 0 &\rightarrow X_{i+1} \sim (1 - \lambda_0)^{-1}(P(X_i, dx) - \lambda_0 \psi(dx)), \\ X_i \in A, Y_i = 1 &\rightarrow X_{i+1} \sim \psi(dx). \end{aligned}$$

It is easily checked that the chain  $X_0, X_1, X_2, \dots$  has the correct transition probability  $P$ . In addition, the set  $a = A \times \{1\}$  is now an atom for  $Z_0, Z_1, Z_2, \dots$  in the sense that (the transition probability of  $Z_0, Z_1, Z_2, \dots$  is abusively still denoted by  $P$ )

$$P(z, C) = \Psi(C) \quad \text{for all } z \in a, C \text{ measurable}, \tag{6}$$

where  $\Psi$  depends only on the measure  $\psi$  and  $\lambda_0$ . The measure  $\Psi$  is  $P(z, B \times \{1\}) = \int_B \psi(x) \lambda_0 \mathbf{1}_{\{x \in A\}} dx$  and  $P(z, B \times \{0\}) = \int_B \psi(x) (1 - \lambda_0 \mathbf{1}_{\{x \in A\}}) dx$ .) In particular, the chain regenerates as soon as it enters  $a$ , i.e. whenever  $Z_i \in a$ , the distribution of  $Z_{i+1}, Z_{i+2}, \dots$  is always the same. We denote the expectation under this measure as  $\mathbb{E}_a$ . We also set

$$\theta_a = \inf\{i \geq 1 : Z_i \in a\}.$$

As a consequence of (2) and (3) (see Lemma \*1 of the supplementary material),

$$\mathbb{P}_z(\theta_a < \infty) = 1, \quad \alpha_0 = \mathbb{E}_a[\theta_a] < \infty \quad \text{for all } z \in \mathcal{E} \times \{0, 1\}. \tag{7}$$

Two essential consequences of (6) and (7) are the following. Let  $\theta_a(k)$  denote the  $k$ th hitting time of  $a$  ( $\theta_a(1) = \theta_a > 0$ ). Then the variables

$$B_k = (Z_{\theta_a(k)+1}, \dots, Z_{\theta_a(k+1)}), \quad k \in \mathbb{N}^*, \tag{8}$$

form an i.i.d. sequence of random variables valued in  $\bigcup_{i \geq 1} \mathbb{R}^i$ . These random variables are called ‘blocks’. Second, the chain has a unique invariant probability  $\pi$  and we have the classical formula (see Nummelin (1984, Equation (5.7))), for any bounded function  $g$ ,

$$\mathbb{E}_x \left[ \sum_{i=1}^{\theta_a} g(Z_i) \right] = \alpha_0 \pi(g). \tag{9}$$

Based on this, many properties of independent sequences can be extended to Markov chains. As it is useful in our study, in Section A of the supplementary material, we derive a bound on the second-order moments of certain empirical sums over Markov chains satisfying (2), (3), and (5).

*Control of the recurrence.* As we see with (3) above, a key point for the application of this theory is the control of the moments of  $\tau_A$ . This can be classically achieved through the following result (see Jarner and Roberts (2002, Theorem 3.6)): if there exists a function  $V \geq 1$  such that, for all  $x \in \mathcal{E}$ ,

$$\mathbb{E}_x[V(X_1)] \leq V(x) - cV(x)^{1-1/p} + c^{-1}\mathbf{1}_A(x) \tag{10}$$

with  $c > 0$ , then for some  $c' > 0$  and for all  $x \in \mathcal{E}$ ,  $\mathbb{E}_x[\tau_A^p] \leq c'V(x)$ .

*The  $m_0 > 1$  case.* Consider, e.g. the chain  $X_i = (A_i, B_i)$ ,  $i \in \mathbb{N}$ , with the following transition: given  $X_{i-1}$ , draw  $U_i \sim \text{binomial}(1, \frac{1}{2})$ ,  $A'_i, B'_i \sim \mathcal{N}(0, 1)$ , and set  $X_i = (A'_i, B_{i-1})$  if  $U_i = 0$ , and otherwise  $X_i = (A_{i-1}, B'_i)$ . Then  $(X_i)_{i \in \mathbb{N}}$  does not satisfy (4) with  $m_0 = 1$ , but does with  $m_0 = 2$ . This may induce serious complications since the block theory actually fails for the chain  $(X_i)_{i \in \mathbb{N}}$ .

However, for  $k = 0, \dots, m_0 - 1$ , the chain  $(X_{im_0+k})_{i \in \mathbb{N}}$  satisfies (5). Consequently, some properties when  $m_0 > 1$  might be directly deduced from the  $m_0 = 1$  case, e.g. for obtaining bounds on empirical sums.

### 3. Convergence of density estimators

This section includes some results on kernel estimators of the density of the invariant measure associated to a Markov chain. We start by stating approximation results in  $L_p$ -spaces and then we consider the question of uniform convergence with the help of empirical process theory.

As the proofs of certain results are technical, we postpone their proofs to Section B of the supplementary material.

#### 3.1. Approximation in $L_p$ -spaces

We denote by  $\lfloor s \rfloor$  the greatest integer smaller than  $s$ , e.g.  $\lfloor 3 \rfloor = 2$ . Following Tsybakov (2009), we define the Nikol'ski class of functions  $H_q(s, M)$  of regularity  $s$  with constant  $M > 0$  and order  $q \geq 1$  as the set of functions  $\psi$  bounded by  $M$  and  $\lfloor s \rfloor$ -times differentiable whose derivatives of order  $\lfloor s \rfloor$  satisfy, for every  $u \in \mathbb{R}^d$ ,

$$\int |\psi^{(l)}(x+u) - \psi^{(l)}(x)|^q dx \leq M^q |u|_1^{q(s-\lfloor s \rfloor)}, \quad l = (l_1, \dots, l_d) \in \mathbb{N}^d, \quad \sum_{i=1}^d l_i \leq \lfloor s \rfloor,$$

where  $\psi^{(l)} = \partial_{x_1}^{l_1} \cdots \partial_{x_d}^{l_d} \psi$  and  $\|\cdot\|_1$  denotes the  $\ell_1$ -norm. Note that  $0 < s - \lfloor s \rfloor \leq 1$ . When  $s < 1$ , the Nikolski class contains discontinuous functions whereas the more classical Hölder regularity class does not (see Delyon and Portier (2016, Lemma 9)). As a result, the Nikolski class is too large to guarantee pointwise convergence of kernel density estimators. It still ensures the convergence in the  $L_q(\pi)$ -norm which is enough for our purpose. While the usual definition of the Nikolski class is with  $q = 2$ , considering different values of  $q$  helps when treating the bias of the density estimator along the blocks of the chain.

We say that  $K$  is a kernel with order  $p \in \mathbb{N}^*$  whenever  $K : \mathbb{R}^d \rightarrow \mathbb{R}$  is symmetric about 0, bounded, and satisfies

$$\int K(x) dx = 1, \quad \int x^l K(x) dx = 0, \quad l = (l_1, \dots, l_d), \quad 0 < \sum_{i=1}^d l_i \leq p - 1$$

with the notation  $x^l = (x_1^{l_1}, \dots, x_d^{l_d})$ .

For every  $h > 0$ , we introduce the notation

$$K_h(\cdot) = h^{-d} K\left(\frac{\cdot}{h}\right).$$

For any other function  $\psi : \mathbb{R}^d \rightarrow \mathbb{R}$ , the convolution between  $\psi$  and  $K_h$  is

$$\psi_h(x) = (\psi \star K_h)(x) = \int \psi(x - hu)K(u) du.$$

In the next lemma, we assert that for kernels with sufficiently high order, the larger the Nikolski regularity of  $\psi$  and  $\pi$  the better the rate of convergence of  $\psi_h$  to  $\psi$  in  $L_q(\pi)$ -norm. For any bounded real-valued function  $g$  defined on some space  $\mathcal{X}$ , we set

$$g_\infty = \sup_{x \in \mathcal{X}} |g(x)|.$$

**Lemma 1.** *Let  $s > 0$ ,  $q \geq 1$ , and suppose that  $K$  has order (strictly) greater than  $\lfloor s \rfloor$  such that  $\int |u|_1^s |K(u)| du < +\infty$  and  $\psi : \mathbb{R}^d \rightarrow \mathbb{R}$  belongs to  $H_q(s, M_1)$ . Then for any bounded density  $\pi$  on  $\mathbb{R}^d$  and every  $h > 0$ ,*

$$\|\psi - \psi_h\|_{L_q(\pi)} \leq C_1 M_1 \pi_\infty^{1/q} h^s, \tag{11}$$

where  $C_1$  depends on  $K$  and  $s$ . Suppose the previous assumptions hold with  $q = 1$ . Let  $r > 0$  and assume, moreover, that  $K$  has order (strictly) greater than  $\lfloor r \rfloor$  such that  $\int |u|_1^r |K(u)| du < +\infty$ ,  $\pi$  belongs to  $H_1(r, M_2)$ , and  $\int |\psi(x)| dx < +\infty$ . Then there exists  $C_2 > 0$  such that, for every  $h > 0$ ,

$$|\pi(\psi - \psi_h)| \leq C_2 (M_1 \pi_\infty + M_2 \psi_\infty) h^{r \vee s}, \tag{12}$$

where  $C_2$  depends on  $K$ ,  $s$ , and  $r$ .

### 3.2. Uniform concentration

Our approach is based on empirical process theory and more precisely on the result from Einmahl and Mason (2005). Given i.i.d. random variables  $\xi_1, \xi_2, \dots$ , it can be used to provide a bound on the expected value of

$$\sup_{f \in \mathcal{F}} \left| \sum_{i=1}^n (f(\xi_i) - \mathbb{E}[f(\xi_1)]) \right|,$$

whenever the class of functions  $\mathcal{F}$  is a Vapnik–Chervonenkis (VC) class of functions (see

Theorem 1 below). A class  $\mathcal{F}$  is VC whenever there exist  $A > 0$  and  $v > 0$  such that for every probability measure  $Q$  satisfying  $\|F\|_{L_2(Q)} < \infty$  and every  $0 < \varepsilon < 1$ ,

$$\mathcal{N}(\mathcal{F}, L_2(Q), \varepsilon \|F\|_{L_2(Q)}) \leq \left(\frac{A}{\varepsilon}\right)^v,$$

where  $F$  is an envelope for  $\mathcal{F}$ , i.e. for any  $f \in \mathcal{F}$ ,  $|f(x)| \leq F(x)$ , and  $\mathcal{N}(T, d, \varepsilon)$  denotes the  $\varepsilon$ -covering number of the metric space  $(T, d)$  (see van der Vaart and Wellner (1996)). Many classes of interest turn out to be VC, e.g. polynomials and indicators, and several preservation properties are available (see Propositions 1–3 below).

The following statement is actually a slight modification of Einmahl and Mason (2005, Proposition 1). Comments are given below.

**Theorem 1.** (Einmahl and Mason (2005).) *Let  $\xi_1, \dots, \xi_n$  be an i.i.d. sequence and  $\mathcal{F}$  be a VC class of functions with envelope  $F$  and characteristics  $(A, v)$  with  $A \geq e$  and  $v \geq 1$ . Set  $\beta^2 = \mathbb{E}[F(\xi_1)^2]$ . Let  $\sigma^2$  be such that*

$$\sigma^2 \geq \sup_{f \in \mathcal{F}} \mathbb{E}[f(\xi_1)^2], \quad \sigma^2 \geq 16vn^{-1} \log\left(A\left(\frac{\beta}{\sigma} \vee 1\right)\right) \sup_{f \in \mathcal{F}, x \in \mathcal{X}} f(x)^2, \quad (13)$$

then

$$\mathbb{E} \sup_{f \in \mathcal{F}} \left| \sum_{i=1}^n (f(\xi_i) - \mathbb{E}[f(\xi_1)]) \right| \leq C_0 \sqrt{vn\sigma^2 \log\left(A\left(\frac{\beta}{\sigma} \vee 1\right)\right)}, \quad (14)$$

where  $C_0$  is a universal constant.

In Einmahl and Mason (2005), the left-hand side was actually a Rademacher sum, but then (14) follows from the Symmetrization lemma, e.g. van der Vaart and Wellner (1996, Lemma 2.3.1). Another difference is that it is stated only in the  $\sigma \leq \beta$  case. But if  $\sigma \geq \beta$ , we can increase  $F$ , e.g.  $F \rightarrow a \vee F$ , in such a way that  $\beta$  will be equal to  $\sigma$  ( $A$  and  $v$  do not change) and apply the previous result; this leads to (14).

Preservation properties of the covering number’s size will be useful in the sequel to show that some classes are VC. In the following proposition we assert that locally Lipschitz transformations of VC classes are still VC. This result is a slight variation of van der Vaart and Wellner (1996, Theorem 2.10.20) in which the authors considered uniform entropy numbers with respect to discretely finite probability measures.

**Proposition 1.** *Let  $\mathcal{F}_1, \dots, \mathcal{F}_d$  be VC classes of functions defined on a common space  $\mathcal{X}$  such that each  $f \in \mathcal{F}_j$  is valued in the set  $I_j \subset \mathbb{R}$  and  $\mathcal{F}_j$  has envelope  $F_j$ . Let  $\Psi: I_1 \times \dots \times I_d \rightarrow \mathbb{R}$  be such that, for any  $A = (A_1, \dots, A_d) \in \mathbb{R}_+^d$ ,*

$$|\Psi(z) - \Psi(\tilde{z})| \sum_{j=1}^d C_j(A) |z_j - \tilde{z}_j| \quad (15)$$

for all  $z, \tilde{z} \in ([-A_1, A_1] \cap I_1) \times \dots \times ([-A_d, A_d] \cap I_d)$ , where  $C_j: \mathbb{R}^d \rightarrow \mathbb{R}$ ,  $j = 1, \dots, d$ , are nonnegative functions. Let  $\mathcal{G}$  denote the class of functions  $x \mapsto \Psi(f_1(x), \dots, f_d(x))$  when  $(f_1, \dots, f_d)$  ranges over  $\mathcal{F}_1 \times \dots \times \mathcal{F}_d$ . The class  $\mathcal{G}$  is a VC class of functions with envelope

$$G = |\Psi(f_0)| + 2 \sum_{j=1}^d (1 \vee F_j) C_j(F),$$

where  $F = (F_1, \dots, F_d)$ , and  $f_0$  is an arbitrary function in  $\mathcal{F}_1 \times \dots \times \mathcal{F}_d$ .



In the following proposition, which includes a result from Nolan and Pollard (1987), we have interesting examples of uniformly bounded VC classes of functions. We consider a kernel function  $K : \mathbb{R}^d \rightarrow \mathbb{R}$  that takes one of the two following forms:

$$K(x) = \begin{cases} K^{(0)}(|x|), & (16a) \\ \prod_{k=1}^d K^{(0)}(x_k), & (16b) \end{cases}$$

where  $K^{(0)}$  is a bounded real function of bounded variation. We denote by  $K_\infty$  the supremum of  $K$ .

**Proposition 2.** *The class of functions  $\{x \mapsto \mathbf{1}_{\{x \leq M\}} : M \in \mathbb{R}\}$  is a uniformly bounded VC class of functions. Assume that (16a) and (16b) hold. The class of functions  $\{x \mapsto K(h^{-1}(y - x)) : y \in \mathbb{R}^d, h > 0\}$  is a uniformly bounded VC class of functions.*

By applying Proposition 1 to the VC classes of the previous proposition, we establish the VC property for some class of functions which will be of great interest in the sequel.

**Proposition 3.** *Assume that (16a) and (16b) hold. The class of functions*

$$\{(t, x) \mapsto t \mathbf{1}_{\{t \leq M\}} K(h^{-1}(y - x)) : y \in \mathbb{R}^d, h > 0, M \in \mathbb{R}\} \tag{17}$$

*defined on  $\mathbb{R} \times \mathbb{R}^d$  is a VC class of functions with envelope  $(t, x) \mapsto 2((1 \vee K_\infty)|t| + (1 \vee |t|)K_\infty)$ .*

Based on Proposition 2, if the random variables  $X_1, X_2, \dots$  used in the construction of  $\hat{\pi}$  were independent, then we would have, under the assumptions of Theorem 1 and Proposition 2,

$$\sup_{y \in \mathbb{R}^d} |\hat{\pi}(y) - \pi_{h_n}(y)| = O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{nh_n^d}}\right),$$

whenever  $h_n \rightarrow 0$  and  $nh_n^d / \log(n) \rightarrow +\infty$  as  $n \rightarrow +\infty$ . For Markov chains, we require the stronger condition on the sequence of bandwidth, i.e.

$$h_n \rightarrow 0, \quad \frac{nh_n^{dp_0/(p_0-1)}}{\log(n)} \rightarrow +\infty \tag{18}$$

for some  $p_0 > 2$  such that

$$\xi(p_0) = \sup_{x \in A} \mathbb{E}_x[\tau_A^{p_0}] < +\infty. \tag{19}$$

In addition, our approach only permits us to obtain the convergence to 0 in probability, not any sharp bound on the rate of convergence.

**Theorem 2.** *Let  $(X_i)_{i \in \mathbb{N}}$  be a Markov chain satisfying (2), (5), and (19) for some  $p_0 > 2$ . Suppose that  $K$  satisfies (16a) and (16b), and that (18) holds for the same  $p_0 > 2$ . If  $\pi$  is bounded, and  $\int (|K(x)| + K(x)^2) dx < +\infty$ , we have*

$$\sup_{y \in \mathbb{R}^d} |\hat{\pi}(y) - \pi_{h_n}(y)| \rightarrow 0 \text{ in } \mathbb{P}_\pi\text{-probability.}$$

Further analysis on the difference between  $\pi$  and  $\pi_{h_n}$  leads to the following statement which prevents the estimated density from being close to 0.

**Corollary 1.** *Under the assumptions of Theorem 2, suppose that  $Q \subset \mathbb{R}^d$  is a compact set such that  $\pi$  is continuous on  $Q$  and  $\inf_{y \in Q} \pi(y) \geq b > 0$ . If  $K$  has bounded support and if there exist  $c > 0$  and  $h_0 > 0$  such that for every  $x \in Q, 0 < h < h_0$ , it holds that  $(\mathbf{1}_{\{Q\}} \star K_h)(x) \geq c$ , then*

$$\mathbb{P}_\pi \left( \inf_{y \in Q} \hat{\pi}(y) \geq \frac{cb}{2} \right) \rightarrow 1.$$

**4. Main result**

We now provide the rate of convergence of the estimator  $\hat{I}_{ks}$  of  $I_0$ . We rely largely on the regenerative framework described in the previous section. In particular, the following set of assumptions ensures the statements of Theorem 2 and Corollary 1.

- (H1) For some  $s > 0$  and  $M_1 > 0$ , the support of  $\varphi$  is a compact set  $Q \subset \mathbb{R}^d$  and  $\varphi$  belongs to  $H_q(s, M_1)$  for any  $q \geq 1$ .
- (H2) For some  $r > 0$  and  $M_2 > 0$ ,  $\pi$  is continuous, bounded on  $Q$ , and belongs to  $H_q(r, M_2)$  for any  $q \geq 1$ . Moreover, there exists  $b > 0$  such that  $\inf_{y \in Q} \pi(y) \geq b$ .
- (H3) Let  $K$  be a kernel satisfying (16a) and (16b) with order (strictly) greater than  $r$  and  $s$ . There exist  $c > 0$  and  $h_0 > 0$  such that for every  $x \in Q$  and  $0 < h < h_0$ ,

$$(\mathbf{1}_{\{Q\}} \star K_h)(x) \geq c.$$

- (H4) Let  $(X_i)_{i \in \mathbb{N}}$  be a Markov chain satisfying (2) and (5) with initial measure  $\nu$  absolutely continuous with respect to  $\pi$ . There exists  $p_0 > 3$  such that

$$\sup_{x \in A} \mathbb{E}_x[\tau_A^{p_0}] < +\infty,$$

where  $A$  is the recurrent set introduced in (2), and as  $n \rightarrow +\infty$ , the sequence of bandwidth  $(h_n)_{n \in \mathbb{N}^*}$  satisfies, as  $n \rightarrow +\infty$ ,

$$h_n \rightarrow 0, \quad \frac{nh_n^{dp_0/(p_0-1)}}{\log(n)} \rightarrow +\infty.$$

Most stable Markov chains satisfy (H4). This has been the subject of many studies; see Meyn and Tweedie (2009) where the drift condition (10) was used to bound the moments of the return times. Examples include, for instance, auto-regressive models (see Meyn and Tweedie (2009, Theorem 16.5.1, Equation (16.43))) or the Metropolis–Hastings algorithm (see Jarner and Roberts (2002, Example 5.2)). Since the invariant measure  $\pi$  is the solution to  $\pi(y) = \int \pi(x)P(x, y) dx$ , where  $P$  is the transition density, the smoothness of  $y \mapsto P(x, y)$  will essentially ensure the smoothness of  $\pi$  as required in (H2). Whenever  $\pi > 0$  on the support of  $\varphi$  (e.g. as soon as  $P(x, y) > 0$  for all  $(x, y)$ ) and continuous, the lower bound in (H2) holds.

High-order kernels can be constructed using the radial kernel (16a) or using the product-type kernel (16b) following, for instance, Gasser *et al.* (1985) or Li and Racine (2007, Section 1.11). The condition that  $(\mathbf{1}_{\{Q\}} \star K_h)$  is lower bounded (uniformly for  $x$  and small  $h$ ) is applicable in Corollary 1 which is a key ingredient to control the small values of  $\hat{\pi}$ . This condition cannot be trivially verified as it involves the boundary of  $Q$  and the regions where  $K < 0$ . A first example is when  $Q$  is the hypercube and  $K$  is a product-type kernel with initial kernel  $K^{(0)}$  such that  $\int_{-x}^{+\infty} K^{(0)}(u) du > 0$  for all  $x > 0$ . A second example is when the boundary of  $Q$  is smooth and  $K$  is such that  $\int_{\mathcal{H}} K(u) du > 0$  for every half-space  $\mathcal{H}$  containing  $0$ .

In the following theorem we extend the results of Delyon and Portier (2016) for independent sequences of random variables to Harris recurrent Markov chains. A secondary improvement with respect to Delyon and Portier (2016) concerns the requirements on the regularity of  $\pi$ . In Delyon and Portier (2016), the density  $\pi$  was assumed to be at least continuously differentiable on  $\mathbb{R}^d$  and bounded away from 0 on  $Q$ , excluding the case where  $\pi$  is supported on  $Q$ , and possibly discontinuous on the boundary. In our approach, we include such cases by supposing that  $\pi$  is in some Nikolski regularity class. This enables us to see the effect of jumps in the shape of  $\pi$ . As the Nikolski regularity of such functions is smaller than  $\frac{1}{2}$  (see Delyon and Portier (2016, Lemma 11)), a bias term in  $h_n^{1/2}$  will appear in the asymptotic decomposition.

**Theorem 3.** *If, moreover, (H1)–(H4) hold, we have for every initial measure,*

$$\hat{I}_{ks} - I_0 = O_{\mathbb{P}_\nu}(h_n^r + n^{-1}h_n^{-d} + n^{-1/2}h_n^s).$$

*Proof.* We consider the split chain  $(Z_i)_{i \in \mathbb{N}}$  introduced in Section 2 with initial distribution  $\nu$ . We are interested in showing that  $\mathbb{E}_\nu \mathbf{1}_{\{|\hat{I}_{ks} - I_0| > a_n\}} \rightarrow 0$  for some sequence  $a_n \rightarrow 0$ . By applying Lemma \*5 of the supplementary material, it suffices to prove the result in the case when  $\nu$  is equal to  $\pi$ .

From (\*11), it follows that  $l_n/n$  converges to its expectation  $\alpha_0^{-1} > 0$ . We use the fact that  $n/l_n = O_{\mathbb{P}_\pi}(1)$  and that the product of two  $O_{\mathbb{P}_\pi}(1)$  remains  $O_{\mathbb{P}_\pi}(1)$ .

Without loss of generality, we can assume that  $l_n > 2$ . Indeed, the complementary event occurs with probability going to 0 as  $n$  increases.

A convenient scaling in the sequel is to put  $\alpha_0(l_n - 1)$  and  $\alpha_0(l_n - 2)$  instead of  $n$ , in some places, since it simplifies many terms of our expansion. Hence, instead of  $\hat{I}_{ks}$ , we instead study

$$\tilde{I}_{ks} = (l_n - 1)^{-1} \alpha_0^{-1} \sum_{i=1}^n \frac{\varphi(X_i)}{\hat{\pi}_i}$$

with

$$\hat{\pi}_i = \alpha_0^{-1} (l_n - 2)^{-1} \sum_{j=1}^n K_{ij} \quad \text{and} \quad K_{ij} = K_{h_n}(X_i - X_j).$$

Since  $\hat{I}_{ks} = ((l_n - 1)/(l_n - 2)) \tilde{I}_{ks}$  and  $((l_n - 1)/(l_n - 2)) = O_{\mathbb{P}_\pi}(1)$ , the rates of convergence of  $\tilde{I}_{ks}$  and  $\hat{I}_{ks}$ , in probability, are the same.

We now introduce the notation

$$\psi_q(x) = \frac{\varphi(x)}{\pi(x)^q}, \quad q \in \mathbb{N}.$$

The development (reminiscent of the Taylor expansion of  $\hat{\pi}_i$  around  $\pi(X_i)$ )

$$\frac{1}{\hat{\pi}_i} = \frac{2}{\pi(X_i)} - \frac{\hat{\pi}_i}{\pi(X_i)^2} + \frac{(\pi(X_i) - \hat{\pi}_i)^2}{\hat{\pi}_i \pi(X_i)^2}$$

allows us to expand  $\tilde{I}_{ks}$  as

$$\tilde{I}_{ks} = (\alpha_0(l_n - 1))^{-1} \sum_{i=1}^n 2\psi_1(X_i) - (\alpha_0^2(l_n - 1)(l_n - 2))^{-1} \sum_{i=1}^n \sum_{j=1}^n \psi_2(X_i) K_{ij} + R_{1,n}$$

with

$$R_{1,n} = (\alpha_0(l_n - 1))^{-1} \sum_{i=1}^n \frac{\psi_2(X_i)(\pi(X_i) - \hat{\pi}_i)^2}{\hat{\pi}_i}.$$

Reorganizing the first two terms according to the blocks leads to

$$\tilde{I}_{ks} = -(\alpha_0^2(l_n - 1)(l_n - 2))^{-1} \sum_{k=0}^{l_n} \sum_{l=0}^{l_n} H_{kl} + (\alpha_0(l_n - 1))^{-1} \sum_{k=0}^{l_n} 2G_k + R_{1,n}$$

with, for any  $(k, l) \in \mathbb{N}^2$ ,

$$H_{kl} = \sum_{i \in B_k, j \in B_l} \psi_2(X_i) K_{ij}, \quad G_k = \sum_{i \in B_k} \psi_1(X_i).$$

The notation  $i \in B_k$  is short for  $Z_i \in B_k$  and the block  $B_0$  is the first (incomplete) block  $Z_1, Z_2, \dots, Z_{\theta_a(1)}$ . Diagonal terms of the above  $U$ -statistic and terms related to the first and last block are treated as a remainder, and we write

$$\tilde{I}_{ks} = -(\alpha_0^2(l_n - 1)(l_n - 2))^{-1} \sum_{k=1}^{l_n-1} \sum_{l=k+1}^{l_n-1} \{H_{kl}^*\} + (\alpha_0(l_n - 1))^{-1} \sum_{k=1}^{l_n-1} \{2G_k + R_{1,n} + R_{2,n}\} \tag{20}$$

with  $H_{kl}^* = H_{kl} + H_{lk}$  and

$$R_{2,n} = -(\alpha_0^2(l_n - 1)(l_n - 2))^{-1} \left( H_{00} + H_{l_n l_n} + H_{0l_n}^* + \sum_{k=1}^{l_n-1} \{H_{0k}^* + H_{l_n k}^* + H_{kk}\} \right) + (\alpha_0(l_n - 1))^{-1} 2(G_0 + G_{l_n}).$$

The first term in (20) is a  $U$ -statistic whose fluctuations can be controlled by using an Hoeffding-type decomposition with respect to the blocks. Denoting

$$\tilde{H}_k^* = \mathbb{E}_a[H_{1k}^* \mid B_k],$$

we can write

$$\tilde{I}_{ks} - I_\varphi = U_n + M_n + B_n + R_{1,n} + R_{2,n}$$

with (we use the fact that  $\sum_{1 \leq k < l \leq l_n-1} \{\tilde{H}_k^* + \tilde{H}_l^*\} = (l_n - 2) \sum_{1 \leq k \leq l_n-1} \tilde{H}_k^*$  and we underbrace terms which have been deliberately introduced and removed)

$$U_n = -(\alpha_0^2(l_n - 1)(l_n - 2))^{-1} \sum_{k=1}^{l_n-1} \sum_{l=k+1}^{l_n-1} \underbrace{\{H_{kl}^* - \tilde{H}_k^* - \tilde{H}_l^*\}}_1 + \underbrace{\mathbb{E}_a[H_{12}^*]}_2,$$

$$M_n = (\alpha_0(l_n - 1))^{-1} \sum_{k=1}^{l_n-1} \{2G_k - \underbrace{\alpha_0^{-1} \tilde{H}_k^*}_1 - \underbrace{\mathbb{E}_a(2G_1 - \alpha_0^{-1} \tilde{H}_1^*)}_3\},$$

$$B_n = \alpha_0^{-1} \underbrace{\mathbb{E}_a(2G_1 - \alpha_0^{-1} \tilde{H}_1^*)}_3 + \underbrace{(2\alpha_0^2)^{-1} \mathbb{E}_a[H_{12}^*]}_2 - \int \varphi(x) dx.$$

The notation follows from the fact that  $U_n$  is a  $U$ -statistic,  $M_n$  is a martingale,  $B_n$  is a bias term (nonrandom),  $R_{1,n}$  comes from the remainder of the Taylor expansion, and  $R_{2,n}$  corresponds to uncompleted blocks and diagonal terms. We now compute bounds for each term separately.

Step 1:  $U_n = O_{\mathbb{P}}(n^{-1}h_n^{-d/2})$ . Let

$$\tilde{U}_n = \left( \frac{\alpha_0^2(l_n - 1)(l_n - 2)}{n^2} \right) U_n.$$

Since  $l_n \leq n$ , we have,

$$|\tilde{U}_n| \leq n^{-2} \max_{1 \leq L \leq n} |S_L|$$

with  $S_L = \sum_{1 \leq k < l \leq L} \{H_{kl}^* - \tilde{H}_k^* - \tilde{H}_l^* + \mathbb{E}_a[H_{12}^*]\}$ . The independence between the blocks  $(B_k)_{k=1, \dots, n}$ , defined in (8), implies that the process  $L \mapsto S_L$  is a martingale. Then by Doob's inequality, we know that

$$\mathbb{P}_\pi(|\tilde{U}_n| > \varepsilon) \leq \frac{\mathbb{E}_a S_n^2}{\varepsilon^2 n^4},$$

and it remains to develop the squared sum inside the expectation. By construction, the terms in the sum defining  $S_L$  are all orthogonal. As a consequence, we have

$$\mathbb{P}_\pi(|\tilde{U}_n| > \varepsilon) \leq \frac{n(n - 1)\mathbb{E}_a[\{H_{12}^* - \tilde{H}_1^* - \tilde{H}_2^* + \mathbb{E}_a[H_{12}^*]\}^2]}{2\varepsilon^2 n^4} \leq \frac{\mathbb{E}_a[H_{12}^{*2}]}{2\varepsilon^2 n^2}.$$

Due to the symmetry of  $K$  and the boundedness of  $\psi_2$  and  $K$ , we have, denoting by  $\Delta_k$  the length of  $B_k$  for  $k \in \mathbb{N}$  (as introduced in the proof of Theorem 2),

$$\begin{aligned} \mathbb{E}_a[H_{12}^2] &= \mathbb{E}_a \left( \sum_{i \in B_1, j \in B_2} \{\psi_2(X_i)K_{ij} + \psi_2(X_j)K_{ji}\} \right)^2 \\ &\leq \psi_{2,\infty}^2 \mathbb{E}_a \left( \sum_{i \in B_1, j \in B_2} |K_{ij}| + |K_{ji}| \right)^2 \\ &\leq \psi_{2,\infty}^2 K_\infty h_n^{-d} \mathbb{E}_a \left( \Delta_1 \Delta_2 \sum_{i \in B_1, j \in B_2} |K_{ij}| + |K_{ji}| \right) \\ &\leq \psi_{2,\infty}^2 K_\infty h_n^{-d} \mathbb{E}_a \left( (\Delta_1^2 + \Delta_2^2) \sum_{i \in B_1, j \in B_2} |K_{ij}| \right) \\ &= 2\psi_{2,\infty}^2 K_\infty h_n^{-d} \mathbb{E}_a \left( \Delta_1^2 \sum_{i \in B_1, j \in B_2} |K_{ij}| \right). \end{aligned}$$

The independence between the blocks allows us to integrate with respect to  $B_2$ , knowing  $B_1$ , and using (9) this yields

$$\begin{aligned} \mathbb{E}_a \left( \Delta_1^2 \sum_{i \in B_1, j \in B_2} |K_{ij}| \right) &= \alpha_0 \mathbb{E}_a \left( \theta_a^2 \sum_{i \in B_1} \int |K_{h_n}(X_i - y)| \pi(y) dy \right) \\ &\leq \alpha_0 \pi_\infty \int |K(x)| dx \mathbb{E}_a[\theta_a^3]. \end{aligned}$$

From Lemma \*1 and assumption (H4),  $\mathbb{E}_a \theta_a^3$  is finite. We conclude using the fact that  $U_n = O_p(1)\tilde{U}_n$ .

Step 2:  $M_n = O_{\mathbb{P}}(n^{-1/2}h_n^{s \wedge r})$ . Consider

$$\tilde{M}_n = \left( \frac{\alpha_0(l_n - 1)}{n} \right) M_n.$$

We have

$$|\tilde{M}_n| \leq n^{-1} \max_{1 \leq L \leq n} \left| \sum_{1 \leq k \leq L} \{2G_k - \alpha_0^{-1} \tilde{H}_k^* - \mathbb{E}_a(2G_k - \alpha_0^{-1} \tilde{H}_k^*)\} \right|,$$

and Doob’s inequality yields

$$\begin{aligned} \mathbb{P}_\pi(|\tilde{M}_n| > \varepsilon) &\leq \frac{\mathbb{E}_a(\sum_{1 \leq k \leq n} \{2G_k - \alpha_0^{-1} \tilde{H}_k^* - \mathbb{E}_a(2G_1 - \alpha_0^{-1} \tilde{H}_1^*)\})^2}{\varepsilon^2 n^2} \\ &= \frac{\mathbb{E}_a(2G_1 - \alpha_0^{-1} \tilde{H}_1^* - \mathbb{E}_a(2G_1 - \alpha_0^{-1} \tilde{H}_1^*))^2}{\varepsilon^2 n} \\ &\leq \frac{\mathbb{E}_a(2G_1 - \alpha_0^{-1} \tilde{H}_1^*)^2}{\varepsilon^2 n}. \end{aligned}$$

Applying (9), we have

$$\begin{aligned} \alpha_0^{-1} \tilde{H}_1^* &= \sum_{i \in B_1} \int (\psi_2(X_i) K_{h_n}(X_i - y) + \psi_2(y) K_{h_n}(y - X_i)) \pi(y) dy \\ &= \sum_{i \in B_1} \{\psi_2(X_i) \pi_{h_n}(X_i) + \psi_{1h_n}(X_i)\}; \end{aligned} \tag{21}$$

hence, it holds that

$$2G_1 - \alpha_0^{-1} \tilde{H}_1^* = \sum_{i \in B_1} \{\psi_2(X_i)(\pi(X_i) - \pi_{h_n}(X_i)) + (\psi_1(X_i) - \psi_{1h_n}(X_i))\}. \tag{22}$$

Then from Minkowski’s inequality and Lemma \*3, for some  $2 < p < p_0 - 1$  (see assumption (H4)) we obtain

$$\begin{aligned} \|2G_1 - \alpha_0^{-1} \tilde{H}_1^*\|_2 &\leq \psi_{2,\infty} \left\| \sum_{i \in B_1} |\pi(X_i) - \pi_{h_n}(X_i)| \right\|_2 + \left\| \sum_{i \in B_1} |\psi_1(X_i) - \psi_{1h_n}(X_i)| \right\|_2 \\ &\leq C(\|(\pi(X_0) - \pi_{h_n}(X_0))\tau_A^{p/2}\|_2 + \|(\psi_1(X_0) - \psi_{1h_n}(X_0))\tau_A^{p/2}\|_2), \end{aligned}$$

where  $C$  is a constant that depends on  $p$  and on the chain and  $\|\cdot\|_2$  denotes the  $L_2(\pi)$ -norm. Now we use Hölder’s inequality with conjugates  $u$  and  $v$  to obtain

$$\|2G_1 - \alpha_0^{-1} \tilde{H}_1^*\|_2 \leq C \|\tau_A^{p/2}\|_{2v} (\|\pi(X_0) - \pi_{h_n}(X_0)\|_{2u} + \|\psi_1(X_0) - \psi_{1h_n}(X_0)\|_{2u}).$$

Now choose  $v$  sufficiently close to 1 to ensure that, using (H4) and (\*6) and (\*2),  $\mathbb{E}_\pi[\tau_A^{pv}] \leq \mathbb{E}_\pi[\tau_A^{p_0-1}] \leq \mathbb{E}_\pi[\theta_a^{p_0-1}] < +\infty$ . We then use Lemma 1 to obtain the desired rate,  $h_n^r + h_n^s$ , for the two other quantities.

Step 3:  $B_n = O(h_n^r)$ . Using (9) and (22), we have

$$\alpha_0^{-1} \mathbb{E}_a(2G_1 - \alpha_0^{-1} \tilde{H}_1^*) = \int \psi_2(x)(\pi(x) - \pi_{h_n}(x))\pi(x) dx + \int (\psi_1(x) - \psi_{1h_n}(x))\pi(x) dx,$$

and then using (\*9) and the definition of  $\psi_2$  yields

$$\alpha_0^{-1} \mathbb{E}_a(2G_1 - \alpha_0^{-1} \tilde{H}_1^*) = 2\pi(\psi_1 - \psi_{1h_n}).$$

Similarly from (\*9), (9), and (21), it follows that

$$(2\alpha_0^2)^{-1}\mathbb{E}_a[H_{12}^*] = (2\alpha_0^2)^{-1}\mathbb{E}_a[\tilde{H}_1^*] = \frac{1}{2} \int (\psi_2(x)\pi_{h_n}(x) + \psi_{1h_n}(x)\pi(x)) dx = \pi(\psi_{1h_n}).$$

Since  $\int \varphi(x) dx = \pi(\psi_1)$ , this yields

$$B_n = \pi(\psi_1 - \psi_{1h_n}).$$

Since there exists  $M$  such that  $\psi_1$  belongs to  $H_1(r \wedge s, M)$ , applying Lemma 1 yields a bound in  $h_n^{r \vee \min(r,s)} = h_n^r$  for  $B_n$ .

Step 4:  $R_{1,n} = O_{\mathbb{P}}(h_n^{2r} + n^{-1}h_n^{-d})$ . From Corollary 1, and since  $n(\alpha_0(l_n - 1))^{-1} = O_{\mathbb{P}}(1)$ , we obtain

$$\begin{aligned} R_{1,n} &\leq O_{\mathbb{P}}(1) \left\{ n^{-1} \sum_{i=1}^n (\pi(X_i) - \hat{\pi}_i)^2 \right\} \\ &\leq O_{\mathbb{P}}(1) \left\{ n^{-1} \sum_{i=1}^n (\pi(X_i) - \pi_{h_n}(X_i))^2 + (\pi_{h_n}(X_i) - \hat{\pi}_i)^2 \right\}. \end{aligned}$$

We compute the expectation of the first term inside the brackets. Using Lemma 1, we obtain a bound  $O_{\mathbb{P}}(h_n^{2r})$ . To treat the second term inside the bracket, denote  $J^{(-i)} = \{1 \leq k \leq l_n - 1 : i \notin B_k\}$ ,  $l(i) = \{k \in \mathbb{N} : i \in B_k\}$ , and  $K(i, B) = \sum_{j \in B} K_{h_n}(X_i - X_j)$ , and write ( $r_{1,n}$  and  $r_{2,n}$  are specified below)

$$\begin{aligned} &\sum_{i=1}^n (\pi_{h_n}(X_i) - \hat{\pi}_i)^2 \\ &= \sum_{i=\theta_a(1)+1}^{\theta_a(l_n)} (\pi_{h_n}(X_i) - \hat{\pi}_i)^2 + r_{1,n} \\ &= r_{1,n} + \sum_{i=\theta_a(1)+1}^{\theta_a(l_n)} \left( \pi_{h_n}(X_i) - (\alpha_0(l_n - 2))^{-1} \left\{ K(i, B_0) + K(i, B_{l_n}) + K(i, B_{l(i)}) \right. \right. \\ &\qquad \qquad \qquad \left. \left. + \sum_{k \in J^{(-i)}} K(i, B_k) \right\} \right)^2 \\ &\leq 2(l_n - 2)^{-2} \sum_{i=\theta_a(1)+1}^{\theta_a(l_n)} \left( \sum_{k \in J^{(-i)}} \{\pi_{h_n}(X_i) - \alpha_0^{-1} K(i, B_k)\} \right)^2 + r_{1,n} + r_{2,n} \\ &\leq 2(l_n - 2)^{-2} \sum_{i=1}^n \left( \sum_{k \in J^{(-i)}} \{\pi_{h_n}(X_i) - \alpha_0^{-1} K(i, B_k)\} \right)^2 + r_{1,n} + r_{2,n} \end{aligned}$$

with

$$r_{1,n} = \sum_{i=1}^n (\pi_{h_n}(X_i) - \hat{\pi}_i)^2 (\mathbf{1}_{\{i \leq \theta_a(1)\}} + \mathbf{1}_{\{i > \theta_a(l_n)\}}) \leq \sup_{y \in \mathbb{R}^d} |\hat{\pi}(y) - \pi_{h_n}(y)| (\theta_a(1) + \Delta_{l_n})$$

and

$$\begin{aligned}
 r_{2,n} &= 2(\alpha_0(l_n - 2))^{-2} \sum_{i=\theta_a(1)+1}^{\theta_a(l_n)} (K(i, B_0) + K(i, B_{l_n}) + K(i, B_{l(i)}))^2 \\
 &\leq 2(\alpha_0(l_n - 2))^{-2} K_\infty^2 h_n^{-2d} \sum_{i=\theta_a(1)+1}^{\theta_a(l_n)} (\Delta_0 + \Delta_{l_n} + \Delta_{l(i)})^2 \\
 &\leq 6(\alpha_0(l_n - 2))^{-2} K_\infty^2 h_n^{-2d} \sum_{i=\theta_a(1)+1}^{\theta_a(l_n)} (\Delta_0^2 + \Delta_{l_n}^2 + \Delta_{l(i)}^2) \\
 &\leq 6(\alpha_0(l_n - 2))^{-2} K_\infty^2 h_n^{-2d} \left( n(\Delta_0^2 + \Delta_{l_n}^2) + \sum_{i=\theta_a(1)+1}^{\theta_a(l_n)} \Delta_{l(i)}^2 \right).
 \end{aligned}$$

Since  $\sum_{i=\theta_a(1)+1}^{\theta_a(l_n)} \Delta_{l(i)}^2 = \sum_{k=1}^{l_n-1} \Delta_k^3 \leq \sum_{k=1}^n \Delta_k^3$ , we find that the above term between the parentheses has expectation of order  $n(\mathbb{E}_\pi \theta_a^2 + \mathbb{E}_a \theta_a^2 + \mathbb{E}_a \theta_a^3)$ . Since, from Lemma \*1 and assumption (H4), the previous expectations are bounded, it follows that  $r_{2,n} = O_{\mathbb{P}_\pi}(n(nh_n^d)^{-2})$  has a contribution  $O_{\mathbb{P}_\pi}((nh_n^d)^{-2})$  to  $R_{1,n}$ . Moreover, we have  $r_{1,n} = O_{\mathbb{P}_\pi}(1)$  by Theorem 2, which yields a contribution  $o_{\mathbb{P}_\pi}(n^{-1})$  to  $R_{1,n}$ . Regarding the aim of this step,  $r_{1,n}$  and  $r_{2,n}$  are negligible, so that we can concentrate on

$$\sum_{i=1}^n \left( \sum_{k \in J^{(-i)}} \{ \pi_{h_n}(X_i) - \alpha_0^{-1} K(i, B_k) \} \right)^2.$$

We use the independence between the blocks to compute

$$\begin{aligned}
 &\mathbb{E}_\pi \sum_{i=1}^n \left( \sum_{k \in J^{(-i)}} \{ \pi_{h_n}(X_i) - \alpha_0^{-1} K(i, B_k) \} \right)^2 \\
 &= n \mathbb{E}_\pi \left( \sum_{k=1}^{l_n-2} \{ \pi_{h_n}(X_0) - \alpha_0^{-1} K(0, B_k) \} \right)^2 \\
 &\leq n \mathbb{E}_\pi \left( \max_{1 \leq l \leq n} \left| \sum_{k=1}^l \{ \pi_{h_n}(X_0) - \alpha_0^{-1} K(0, B_k) \} \right| \right)^2.
 \end{aligned}$$

Since  $l \mapsto \sum_{k=1}^l \{ \pi_{h_n}(X_0) - \alpha_0^{-1} K(0, B_k) \}$  is a martingale, using Doob's inequality, we obtain

$$\begin{aligned}
 \mathbb{E}_\pi \sum_{i=1}^n \left( \sum_{k \in J^{(-i)}} \{ \pi_{h_n}(X_i) - \alpha_0^{-1} K(i, B_k) \} \right)^2 &\leq 4n \mathbb{E}_\pi \left[ \left( \sum_{k=1}^n \{ \pi_{h_n}(X_0) - \alpha_0^{-1} K(0, B_k) \} \right)^2 \right] \\
 &= 4n^2 \mathbb{E}_\pi [ \{ \pi_{h_n}(X_0) - \alpha_0^{-1} K(0, B_1) \}^2 ] \\
 &\leq 4n^2 \alpha_0^{-2} \mathbb{E}_\pi [ K(0, B_1)^2 ] \\
 &\leq 4n^2 \alpha_0^{-2} h_n^{-d} K_\infty \mathbb{E}_\pi \left\{ \Delta_1 \sum_{j \in B_1} |K_{0j}| \right\}.
 \end{aligned}$$

Here we use the independence between  $B_1$  and  $X_0$  to write

$$\mathbb{E}_\pi \left\{ \Delta_1 \sum_{j \in B_1} |K_{0j}| \right\} = \mathbb{E}_a \left\{ \theta_a \sum_{j \in B_1} \int \pi(X_j - h_n u) |K(u)| du \right\} \leq \pi_\infty \int |K(x)| dx \mathbb{E}_a [\theta_a^2].$$

This leads to the contribution  $O_{\mathbb{P}_\pi}(n^{-1} h_n^{-d})$  to  $R_{1,n}$ .



Step 5:  $R_{2,n} = O_{\mathbb{P}}(n^{-1}h_n^{-d})$ . Recall that

$$R_{2,n} = -(\alpha_0^2(l_n - 1)(l_n - 2))^{-1} \left( H_{00} + H_{l_n l_n} + H_{0l_n}^* + \sum_{k=1}^{l_n-1} \{H_{0k}^* + H_{l_n k}^* + H_{kk}\} \right) + (\alpha_0(l_n - 1))^{-1} 2(G_0 + G_{l_n})$$

with  $H_{kl} = \sum_{i \in B_k, j \in B_l} \psi_2(X_i) K_{ij}$ ,  $H_{kl}^* = H_{kl} + H_{lk}$ , and  $G_k = \sum_{i \in B_k} \psi_1(X_i)$ . First, the boundedness of  $\psi_1$  yields

$$\mathbb{E}_{\pi} |G_0| \leq \psi_{1,\infty} \mathbb{E}_{\pi} \theta_a \quad \text{and} \quad \mathbb{E}_{\pi} |G_{l_n}| \leq \psi_{1,\infty} \mathbb{E}_a \theta_a,$$

leading to a contribution of order  $O_{\mathbb{P}_{\pi}}(n^{-1}) \ll O_{\mathbb{P}_{\pi}}((nh_n^d)^{-1})$ . Second, we have

$$\begin{aligned} \mathbb{E}_{\pi} \left| \sum_{k=1}^{l_n-1} (H_{0k}^* + H_{l_n k}^* + H_{kk}) \right| &\leq n \mathbb{E}_{\pi} (|H_{01}^*| + |H_{l_n 1}^*| + |H_{11}|) \\ &\leq n K_{\infty} \psi_{2,\infty} h_n^{-d} (\mathbb{E}_a \theta_a \mathbb{E}_{\pi} \theta_a + (\mathbb{E}_a \theta_a)^2 + \mathbb{E}_a \theta_a^2), \end{aligned}$$

involving a  $(nh_n^d)^{-1}$  in the  $O_{\mathbb{P}}$ . In a similar fashion, the term  $H_{00} + H_{l_n l_n} + H_{0l_n}^*$  has order  $(n^2 h_n^d)^{-1} \ll (nh_n^d)^{-1}$ . □

### 5. Numerical experiments

#### 5.1. Estimation algorithm

We first recall the framework under investigation. We consider the estimation of the integral of a function  $\varphi$  over  $Q$  from a dataset  $(X_i, \varphi(X_i))_{1 \leq i \leq n}$  when the  $X_i$  form a Markov chain. The estimator  $\hat{I}_{ks}$  of  $I_0 = \int_Q \varphi(x) dx$  is

$$\hat{I}_{ks} = n^{-1} \sum_{i=1}^n \frac{\varphi(X_i)}{\hat{\pi}(X_i)}.$$

As noted by Delyon and Portier (2016) for independent data, the crucial factor for the estimation of  $I_0$  is to select the optimal bandwidth parameter  $h_n$  appearing in the estimator  $\hat{\pi}$  of the design distribution

$$\hat{\pi}(x) = (nh_n^d)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right), \quad x \in \mathbb{R}^d.$$

In this paper we propose to use the multivariate plug-in bandwidth selection developed by Chacón and Duong (2010). More precisely, we exploit the implementation of this algorithm in the R package ks (see R Development Core Team (2007)); see Duong (2007) for a presentation of a preliminary version. This approach ensures better results than the  $\varphi$ -based method proposed by Delyon and Portier (2016) in both the independent and the Markov frameworks. Moreover, this method is simpler because it provides an optimal bandwidth that only depends on the design (and not on  $\varphi$ ) contrary to the aforementioned competitive strategy. It is a particularly interesting procedure to integrate several functions from the same design points, e.g. temperature and salinity, since it requires only one selection of the bandwidth. We strongly recommend the use of this method rather than the one proposed previously by Delyon and Portier (2016). Such a choice of the bandwidth does not fit the theoretical framework of Theorem 3 (as it depends

on the design points) but does not require any knowledge on the regularity of the functions  $\varphi$  and  $f$ .

Delyon and Portier (2016) introduced a corrected version  $\hat{I}_{ks}^c$  of the integral estimator  $\hat{I}_{ks}$  that results in both smaller bias and variance in numerical experiments, i.e.

$$\hat{I}_{ks}^c = \frac{1}{n} \sum_{i=1}^n \frac{\varphi(X_i)}{\hat{\pi}(X_i)} \left( 1 - \frac{\hat{v}(X_i)}{\hat{\pi}(X_i)^2} \right),$$

where

$$\hat{v}(x) = \frac{1}{n(n-1)} \sum_{i=1}^n \left[ \frac{1}{h_n^d} K\left(\frac{x - X_i}{h_n}\right) - \hat{\pi}(x) \right]^2, \quad x \in \mathbb{R}^d.$$

This new estimator is chosen in order to eliminate the leading term in the expansion of the estimation error in the independent case. With function  $\hat{v}$  being positive,  $\hat{I}_{ks}^c$  is lower than  $\hat{I}_{ks}$ , which tends to have a positive bias. In the sequel, we compute both  $\hat{I}_{ks}$  and  $\hat{I}_{ks}^c$  from the same bandwidth  $h_n$  depending only on the design points, and obtained as aforementioned.

**5.2. Simulation study**

We consider three models. For each model, the function  $\varphi$  will be integrated on its support  $Q = [0, 1]^d$ .

*Model  $\mathcal{M}_1$ .* We have

$$\varphi(x_1, \dots, x_d) = \prod_{i=1}^d [2 \sin(\pi x_i)^2 \mathbf{1}_{[0,1]}(x_i)].$$

*Model  $\mathcal{M}_2$ .* We have

$$\varphi(x_1, \dots, x_d) = \prod_{i=1}^d \left[ \frac{1 + \pi^2}{\pi(1 + \exp(1))} \sin(\pi x_i) \exp(x_i) \mathbf{1}_{[0,1]}(x_i) \right].$$

*Model  $\mathcal{M}_3$ .* We have

$$\varphi(x_1, \dots, x_d) = \prod_{i=1}^d \left[ \frac{\pi}{2} \sin(\pi x_i) (1 + \cos(5\pi x_i)) \mathbf{1}_{[0,1]}(x_i) \right].$$

For improved comparability, the normalizing constant of each model is chosen in such a way that  $I_0 = 1$ . In Figure 1, we present the one-dimensional shape function of each model. The three models are continuous but have their own features. Model  $\mathcal{M}_1$  is symmetric centered on the center of  $Q$ , while  $\mathcal{M}_2$  has a negative skewness. Finally,  $\mathcal{M}_3$  has three distinct modes. Consequently, one may expect that the models are in some way sorted by increasing difficulty in numerical integration.

For each model  $\mathcal{M}_i, 1 \leq i \leq 3$ , we compute the estimator and its corrected version, presented in Section 5.1, from independent design (data with uniform distribution on  $Q$  is denoted by  $\mathcal{U}_Q$ ) and from Markov design. In the Markov case, the dataset is generated according to the Metropolis–Hastings algorithm with proposition kernel

$$P_r(x, dy) = \mathcal{U}_{[x-\varepsilon, x+\varepsilon]^d}(dy)$$

with  $\varepsilon = 0.2$  and target measure  $\mathcal{U}_Q$ . The Markov chain that results from this Metropolis–Hastings algorithm satisfies (H4). Indeed, first note that the kernel  $P^6(x, dy)$  has a density

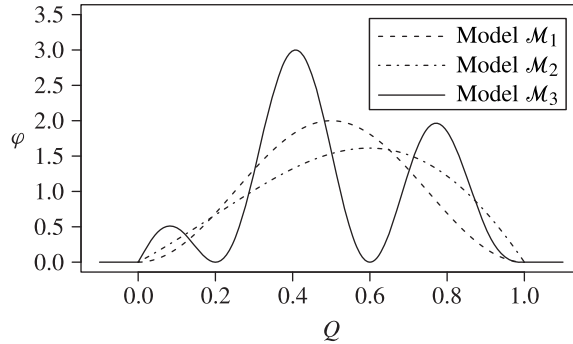


FIGURE 1: Shape of function  $\varphi$  for each model  $\mathcal{M}_i$ ,  $1 \leq i \leq 3$ , in dimension  $d = 1$ .

which is lower bounded on  $[0, 1]^{2d}$  by a positive number (since  $6\epsilon > 1$  and  $\varphi(x)/\varphi(y)$  is lower bounded on  $[0, 1]^{2d}$ ), i.e. starting from any  $x$  and waiting long enough will guarantee that any region is attained by the chain with positive probability. In other words, the uniform Doeblin condition holds for the chains  $(X_{mk+1})_{k \geq 1}$  with the Lebesgue measure. Applying Meyn and Tweedie (2009, Theorem 16.0.2, p. 394), we find that the return time to  $A$  of this chain, which is longer than the return time of the initial chain, has an exponential moment.

Independent and Markov designs have thus not been generated according to the same simulation model but share the same distribution, which makes them comparable. This will allow us to evaluate how the Markovian dependency affects the performance of the methods. For the sake of reference, we also compute the Monte Carlo estimator

$$\hat{I}_{mc} = n^{-1} \sum_{i=1}^n \frac{\varphi(X_i)}{\pi(X_i)},$$

which can only be carried out in a simulation study when the distribution  $\pi$  is known, and not from real data. Furthermore, we investigate various sample sizes ( $n = 500$ ,  $n = 1000$ , and  $n = 2000$ ) and different dimensions ( $d = 1$ ,  $d = 2$ , and  $d = 3$ ). All the numerical results over 50 independent replicates are presented in Figures 2–4 (models  $\mathcal{M}_1$ – $\mathcal{M}_3$ ). In order to make this numerical study reproducible, the R scripts implemented to generate datasets and estimate the integrals of interest are available upon request from the authors.

First, estimators  $\hat{I}_{ks}$  and  $\hat{I}_{ks}^c$  have similar dispersions, but the corrected version is more accurate in most cases and should be the version of choice. Unsurprisingly, the results are better in terms of bias and variance when estimation of the design distribution is computed from independent data rather than from Markov data. In addition, the accuracy deteriorates when the dimension increases. When the sample size is small, the integral is underestimated (see, e.g. models  $\mathcal{M}_1$  and  $\mathcal{M}_2$  in dimensions 2 and 3), in particular, in the Markov framework (see model  $\mathcal{M}_3$  in dimension 3). Numerical results are quite similar for models  $\mathcal{M}_1$  and  $\mathcal{M}_2$ , which indicate that the method is not sensitive to skewness. As expected, the quality is a little lower for  $\mathcal{M}_3$ . In the three models under consideration, the Monte Carlo estimator  $\hat{I}_{mc}$  presents no bias but a large dispersion in comparison with  $\hat{I}_{ks}$  and  $\hat{I}_{ks}^c$ , especially in the Markov framework where the dataset does not exactly follow the distribution  $\pi$ . The results of the numerical study indicate that the methodology is very efficient and applicable in various contexts, in particular, compared to Monte Carlo methods which yield less favorable results in terms of variance and

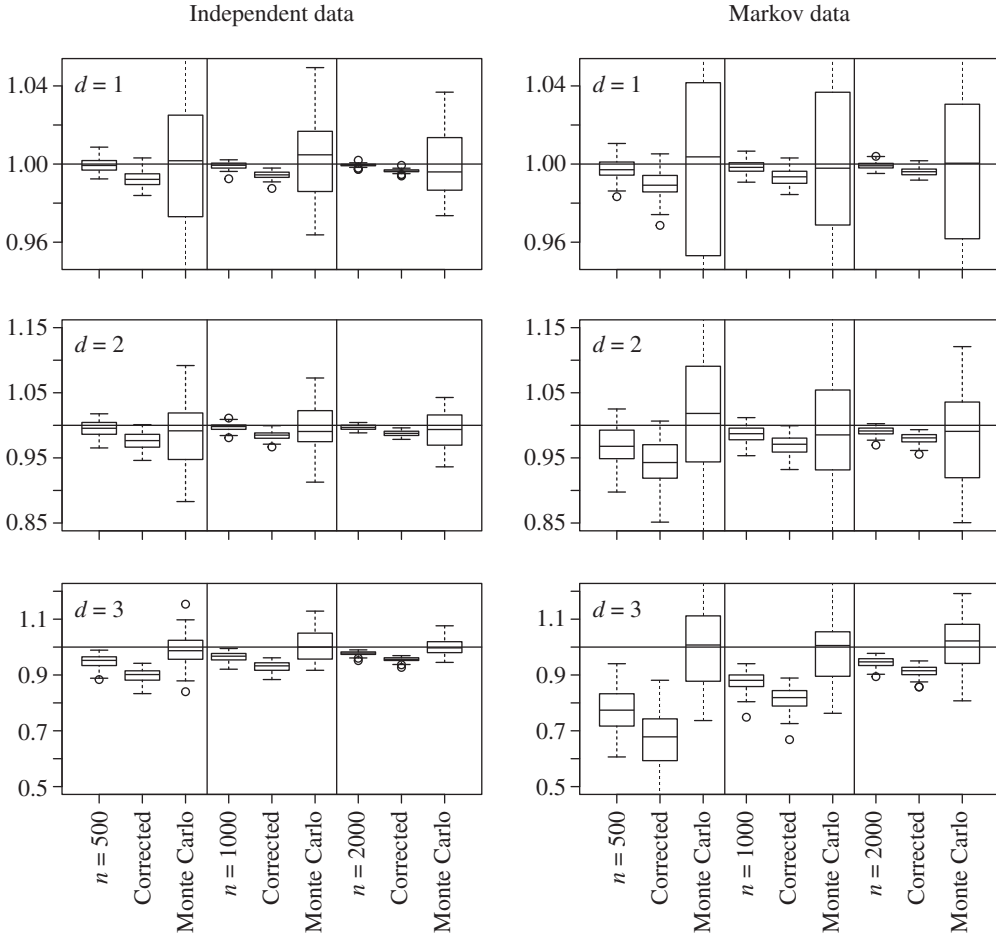


FIGURE 2: Plots of  $\hat{I}_{ks}$ ,  $\hat{I}_{ks}^c$ , and  $\hat{I}_{mc}$  computed from 50 replicates for model  $\mathcal{M}_1$  in dimension  $d = 1$  (upper),  $d = 2$  (middle), and  $d = 3$  (lower) from independent data (left) and Markov data (right).

cannot be applied in a statistical framework. Nevertheless, the results of additional numerical experiments indicate that both estimators present some bias when function  $\varphi$  is not continuous.

As stated in Theorem 3, the shape of the function  $\pi$  (and secondarily  $\varphi$ ) plays an important role in the convergence rate of  $\hat{I}_{ks}$ : the smoother the better. Hence, the situation when  $\pi$  is the uniform density on  $Q$  is far from being straightforward (as the function is not even continuous). Continuity of  $\pi$  is no remedy as it implies the cancellation of  $\pi$  at the border and, therefore, provides insufficient points near the border. One solution is to consider points that lie slightly outside  $Q$ , say in  $\tilde{Q} \supset Q$ , in order to stabilize the estimation of  $\pi$  at the border of  $Q$ . Then compute the kernel estimator  $\tilde{\pi}$  using all these points, and finally calculate

$$\tilde{I}_{ks} = n^{-1} \sum_{i=1}^n \frac{\varphi(X_i) \mathbf{1}_{\{X_i \in Q\}}}{\tilde{\pi}(X_i)}.$$

In the applications where only points in  $Q$  are given, one might prefer to consider a different set  $\tilde{Q}$ , slightly smaller than the original, in order to implement the previous method. If collecting

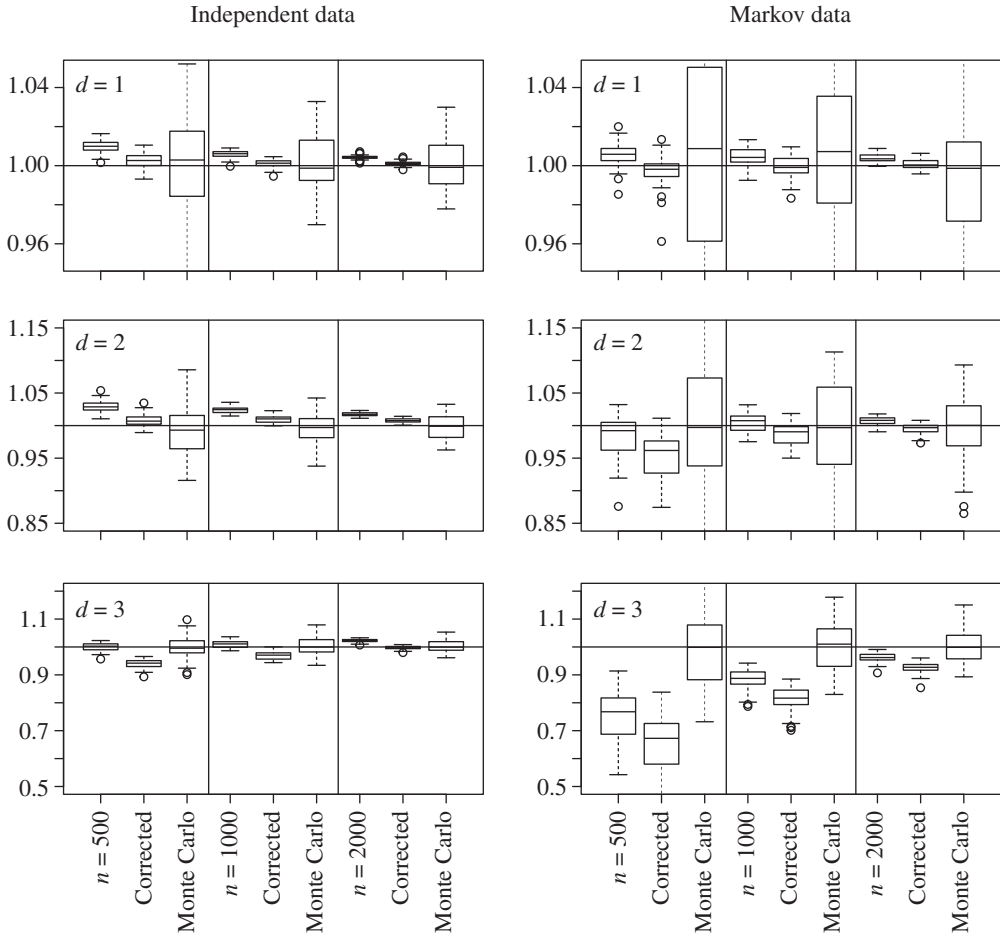


FIGURE 3: Plots of  $\hat{I}_{ks}$ ,  $\hat{I}_{ks}^c$ , and  $\hat{I}_{mc}$  computed from 50 replicates for model  $\mathcal{M}_2$  in dimension  $d = 1$  (upper),  $d = 2$  (middle), and  $d = 3$  (lower) from independent data (left) and Markov data (right).

the points has not been carried out, it might be appropriate to allow the sensor capturing data to leave  $Q$ .

### 5.3. Real-data analysis

The U.S. National Centers for Environmental Information (NCEI) are part of the National Oceanic and Atmospheric Administration (NOAA). NCEI form the world’s largest provider of weather and climate data. The real-data analysis presented in the present paper is based on sea surface temperatures obtained all around the world between 2005 and 2015 from profiling floats (PFL dataset) and is available on the NCEI’s website (World Ocean Database Search and Select, last consulted in July 2016, and available at <https://www.nodc.noaa.gov/cgi-bin/OC5/SELECT/builder.pl>). Sea surface temperatures have a large influence on climate and weather and are therefore used in analyses of climate change. The dataset investigated in this article contains about 1.3M data and is fully described in Table 1 and Figure 5. Data preprocessing has been implemented in Python, while estimation and data analysis have been made with R.

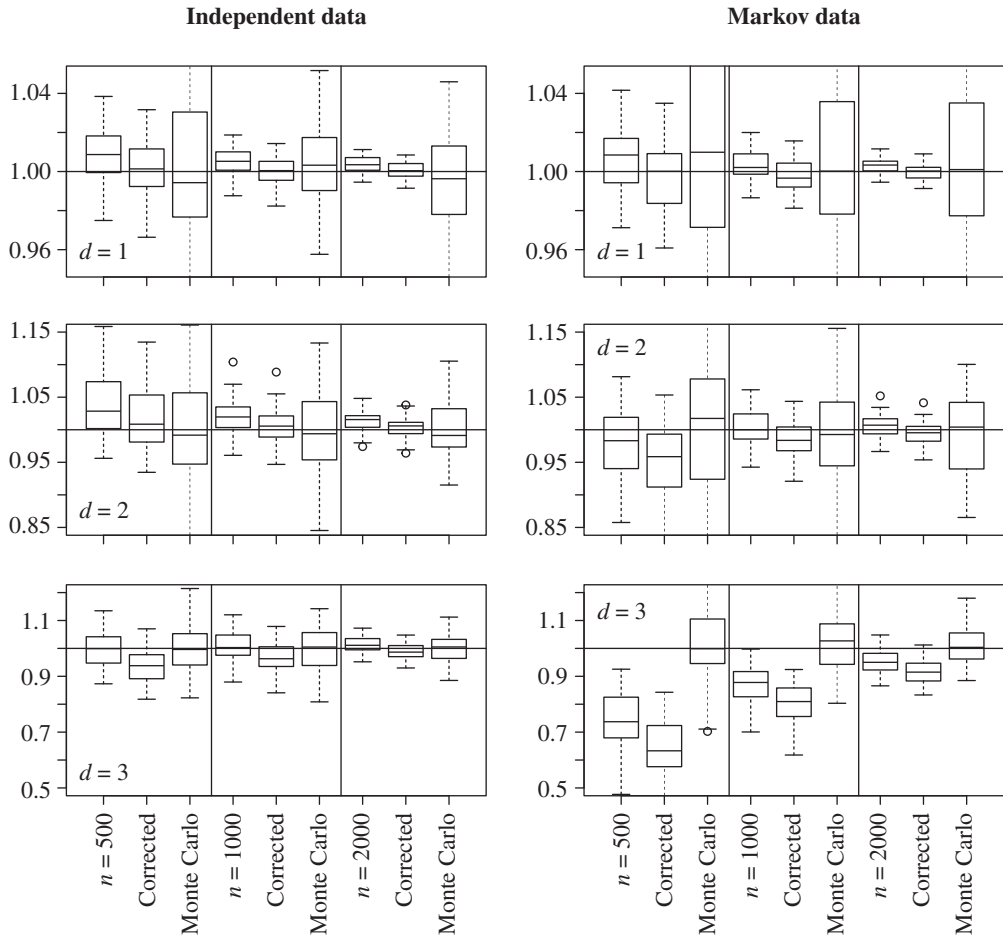


FIGURE 4: Plots of  $\hat{I}_{ks}$ ,  $\hat{I}_{ks}^c$ , and  $\hat{I}_{mc}$  computed from 50 replicates for model  $\mathcal{M}_3$  in dimension  $d = 1$  (upper),  $d = 2$  (middle), and  $d = 3$  (lower) from independent data (left) and Markov data (right).

TABLE 1: Size of the sub-datasets extracted from PFL dataset between 2005 and 2015. The total dataset is 1 343 094.

Year	Pacific Ocean (dataset)		Atlantic Ocean (dataset)		Indian Ocean (dataset)	
	727 135		336 180		279 779	
	min	max	min	max	min	max
2005	35 773	–	16 242	–	14 134	–
2015	–	86 961	–	45 488	–	33 049

The database of interest consists of spatiotemporal data obtained from measure instruments with unpredictable trajectories, which makes them hardly tractable. We focus here on the estimation of the average sea surface temperature for a given period of time, between 2005 and 2015, and for some given areas in the three major oceans. Areas considered in this paper are delimited by the latitude: more than 50° (North region), [30°, 50°] and [10°, 30°] (North Tropical

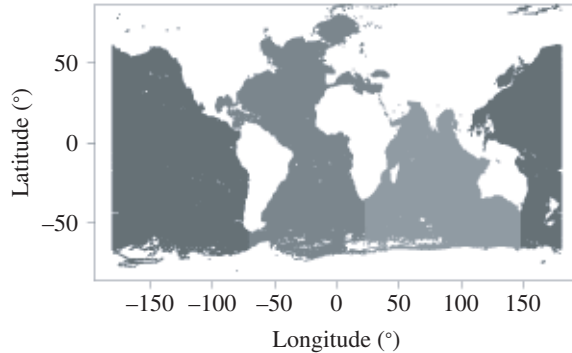


FIGURE 5: Visualization of the 1 343 094 points of the PFL dataset between 2005 and 2015. Oceans are shaded (dark to light: Pacific, Atlantic, Indian).

region),  $[-10^\circ, 10^\circ]$  (Equatorial region),  $[-30^\circ, -10^\circ]$  (South Tropical region),  $[-30^\circ, -50^\circ]$ , and less than  $-50^\circ$  (South region). For each mentioned spatial region, we estimate the average sea surface temperature over each month by the corrected algorithm presented in Section 5.1. This technique is fully adapted to the problem at hand because measurement locations follow erratic trajectories with unknown distribution.

We present the local average sea surface temperatures for the three oceans in Figure 6. Temperature patterns are obtained according to the location on the North–South axis. We observe that the variability of sea surface temperatures in a given region over 11 years is weak compared to the variations in latitude, especially for the Pacific Ocean. In other words, the temperature mainly depends on the latitude, rather on the period of the year. Unsurprisingly, sea surface temperatures are the highest under the equator and near the tropics, where the Earth receives the most direct sunlight.

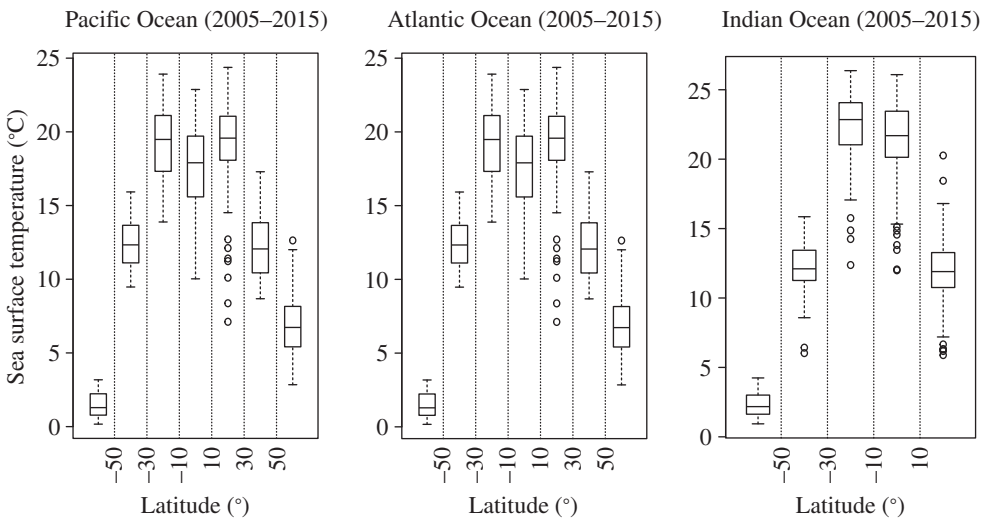


FIGURE 6: Average sea surface temperatures according to the latitude of the considered area for the three major oceans. Each plot has been computed from  $11 \times 12 = 132$  estimates of the average temperature for each month of each year between 2005 and 2015.

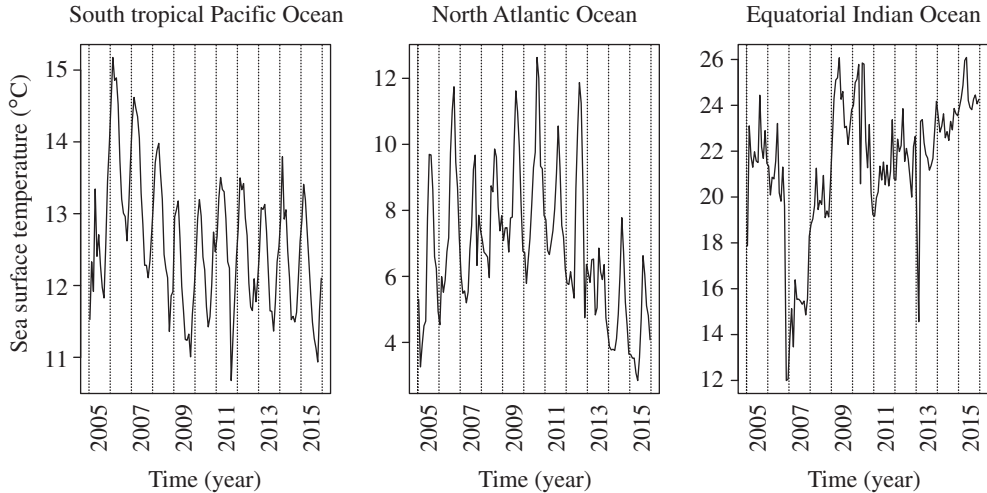


FIGURE 7: Times series of sea surface temperature in some specific areas of the three major oceans between 2005 and 2015. Latitude between  $-30^\circ$  and  $-10^\circ$  for the south tropical Pacific Ocean (*left*),  $50^\circ$  and  $60^\circ$  for the North Atlantic Ocean (*middle*), and  $-10^\circ$  and  $10^\circ$  for the equatorial Indian Ocean (*right*).

In Figure 7 we present time series over 11 years of average sea surface temperatures in three regions: south tropical Pacific Ocean (latitude between  $-30^\circ$  and  $-10^\circ$ ), North Atlantic Ocean (latitude between  $50^\circ$  and  $60^\circ$ ), and equatorial Indian Ocean (latitude between  $-10^\circ$  and  $10^\circ$ ). First, it should be noted that we observe an expected seasonal effect on sea surface temperatures of South Pacific and North Atlantic Oceans: the highest temperatures occur in January and February in the southern hemisphere, while they occur in August and September for the North Atlantic Ocean. In addition, we note a general decrease in sea surface temperature in the southern Pacific between 2006 and 2009 followed by a stable period. This phenomenon has been taken into account in simulations proposed by Kosaka and Xie (2013). In particular, they showed that recent cooling in the Pacific Ocean is tied to the recent global warming hiatus. We also remark that the temperature in the North Atlantic Ocean has decreased recently. Indeed, there is a region of cooling in the Northern Atlantic. Rahmstorf *et al.* (2015) suggested that this cooling may be due to changes in the Atlantic meridional overturning circulation in the late twentieth century. Finally, we point out that the equatorial Indian Ocean has tended to warm for at least 10 years. According to Roxy *et al.* (2014), this warming began more than a century ago and is linked to the El Niño–Southern Oscillation periodical phenomenon.

### Acknowledgements

F. P. gratefully acknowledges support from the Fonds de la Recherche Scientifique (FNRS) A4/5 FC 2779/2014–2017 (grant number 22342320) and the contract Project d'Actions de Recherche Concertées (grant number 12/17–045) of the Communauté française de Belgique.

### References

- [1] ATHREYA, K. B. AND NEY, P. (1978). A new approach to the limit theory of recurrent Markov chains. *Trans. Amer. Math. Soc.* **245**, 493–501.
- [2] AZAÏS, R., DELYON, B. AND PORTIER, F. (2018). Integral estimation based on Markovian design. Supplementary material. Available at <http://doi.org/10.1017/apr.2018.38>.



- [3] BERTAIL, P. AND CLÉMENÇON, S. (2011). A renewal approach to Markovian  $U$ -statistics. *Math. Meth. Statist.* **20**, 79–105.
- [4] CHACÓN, J. E. AND DUONG, T. (2010). Multivariate plug-in bandwidth selection with unconstrained pilot bandwidth matrices. *TEST* **19**, 375–398.
- [5] DELYON, B. AND PORTIER, F. (2016). Integral approximation by kernel smoothing. *Bernoulli* **22**, 2177–2208.
- [6] DUONG, T. (2007). Ks: kernel density estimation and kernel discriminant analysis for multivariate data in R. *J. Statist. Software* **21**, 16pp.
- [7] EINMAHL, U. AND MASON, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.* **33**, 1380–1403.
- [8] EVANS, M. AND SWARTZ, T. (2000). *Approximating Integrals via Monte Carlo and Deterministic Methods*. Oxford University Press.
- [9] GASSER, T., MULLER, H.-G. AND MAMMITZSCH, V. (1985). Kernels for nonparametric curve estimation. *J. R. Statist. Soc. B* **47**, 238–252.
- [10] HANSEN, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* **24**, 726–748.
- [11] HÄRDLE, W. AND STOKER, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *J. Amer. Statist. Assoc.* **84**, 986–995.
- [12] JARNER, S. F. AND ROBERTS, G. O. (2002). Polynomial convergence rates of Markov chains. *Ann. Appl. Prob.* **12**, 224–247.
- [13] KOSAKA, Y. AND XIE, S.-P. (2013). Recent global-warming hiatus tied to equatorial Pacific surface cooling. *Nature* **501**, 403–407.
- [14] LI, Q. AND RACINE, J. S. (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.
- [15] MEYN, S. AND TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd edn. Cambridge University Press.
- [16] NOLAN, D. AND POLLARD, D. (1987).  $U$ -processes: rates of convergence. *Ann. Statist.* **15**, 780–799.
- [17] NOVAK, E. (2016). Some results on the complexity of numerical integration. In *Monte Carlo and Quasi-Monte Carlo Methods*, Springer, Cham, pp. 161–183.
- [18] NUMMELIN, E. (1978). A splitting technique for Harris recurrent Markov chains. *Z. Wahrscheinlichkeitsth.* **43**, 309–318.
- [19] NUMMELIN, E. (1984). *General Irreducible Markov Chains and Nonnegative Operators*. Cambridge University Press.
- [20] R DEVELOPMENT CORE TEAM (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. Available at <http://www.R-project.org/>.
- [21] RAHMSTORF, S. *et al.* (2015). Exceptional twentieth-century slowdown in Atlantic Ocean overturning circulation. *Nature Climate Change* **5**, 475–480.
- [22] ROBERT, C. P. AND CASELLA, G. (2004). *Monte Carlo Statistical Methods*, 2nd edn. Springer, New York.
- [23] ROGET-VIAL, C. (2003). Deux contributions à l'étude semi-paramétrique d'un modèle de régression. Doctoral Thesis, University of Rennes 1.
- [24] ROUSSAS, G. G. (1969). Nonparametric estimation of the transition distribution function of a Markov process. *Ann. Math. Statist.* **40**, 1386–1400.
- [25] ROXY, M. K., RITIKA, K., TERRAY, P. AND MASSON, S. (2014). The curious case of Indian Ocean warming. *J. Climate* **27**, 8501–8509.
- [26] SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London.
- [27] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York.
- [28] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.
- [29] VAN DER VAART, A. W. AND WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York.