
Convergence rates for density estimators of weakly dependent time series

Nicolas Ragache¹ and Olivier Wintenberger²

¹ MAP5, Université René Descartes 45 rue des Saints-Pères, 75270 Paris, France
nicolas.ragache@ensae.fr

² SAMOS, Statistique Appliquée et MODélisation Stochastique, Université Paris 1,
Centre Pierre Mendès France, 90 rue de Tolbiac, F-75634 Paris Cedex 13,
France. olivier.wintenberger@univ-paris1.fr

1 Introduction

Assume that $(X_n)_{n \in \mathbb{Z}}$ is a sequence of \mathbb{R}^d valued random variables with common distribution which is absolutely continuous with respect to Lebesgue's measure, with density f . Stationarity is not assumed so that the case of a sampled process $\{X_{i,n} = x_{h_n(i)}\}_{1 \leq i \leq n}$ for any sequence of monotonic functions $(h_n(\cdot))_{n \in \mathbb{Z}}$ and any stationary process $(x_n)_{n \in \mathbb{Z}}$ that admits a marginal density is included. This paper investigates convergence rates for density estimation in different cases. First, we consider two concepts of weak dependence:

- Non-causal η -dependence introduced in [DL99] by Doukhan & Louhichi,
- Dedecker & Prieur's $\hat{\phi}$ -dependence (see [DP04]).

These two notions of dependence cover a large number of examples of time series (see section § 3). Next, following Doukhan (see [Dou90]) we propose a unified study of linear density estimators \hat{f}_n of the form

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n K_{m_n}(x, X_i), \quad (1)$$

where $\{K_{m_n}\}$ is a sequence of kernels. Under classical assumptions on $\{K_{m_n}\}$ (see section § 2.2), the results in the case of independent and identically distributed (i.i.d. in short) observations X_i are well known (see for instance [Tsy04]). At a fixed point $x \in \mathbb{R}^d$, the sequence m_n can be chosen such that

$$\|\hat{f}_n(x) - f(x)\|_q = O\left(n^{-\rho/(2\rho+d)}\right), \quad (2)$$

where $\|X\|_q^q = \mathbb{E}|X|^q$. The coefficient $\rho > 0$ measures the regularity of f (see Section 2.2 for the definition of the notion of regularity). The same rate of convergence also holds for the Mean Integrated Square Error (MISE), defined

as $\int \|\hat{f}_n(x) - f(x)\|_2^2 p(x) dx$ for some nonnegative and integrable function p . The rate of uniform convergence on a compact set incurs a logarithmic loss appears. For all $M > 0$ and for a suitable choice of the sequence m_n ,

$$\mathbb{E} \sup_{\|x\| \leq M} |\hat{f}_n(x) - f(x)|^q = \mathcal{O} \left(\frac{\log n}{n} \right)^{q\rho/(d+2\rho)}, \quad (3)$$

and

$$\sup_{\|x\| \leq M} |\hat{f}_n(x) - f(x)| =_{a.s.} \mathcal{O} \left(\frac{\log n}{n} \right)^{\rho/(d+2\rho)}. \quad (4)$$

These rates are optimal in the minimax sense. We thus have no hope to improve on them in the dependent setting. A wide literature deals with density estimation for absolutely regular or β -mixing processes (for a definition of mixing coefficients, see [Dou94]). For instance, under the assumption $\beta_r = o(r^{-3-2d/\rho})$, Ango Nze & Doukhan prove in [AD98] that (2), (3) and (4) still hold. The sharper condition $\sum_r |\beta_r| < \infty$ entails the optimal rate of convergence for the MISE (see [Vie97]). Results for the MISE have been extended to the more general $\tilde{\phi}$ - and η -dependence contexts by Dedecker & Prieur ([DP04]) and Doukhan & Louhichi in [DL01]. In this paper, our aim is to extend the bounds (2), (3) and (4) in the η - and $\tilde{\phi}$ -weak dependence contexts.

We use the same method as in [DL99] based on the following moment inequality for weakly dependent and centered sequences $(Z_n)_{n \in \mathbb{Z}}$. For each even integer q and for each integer $n \geq 2$:

$$\left\| \sum_{i=1}^n Z_i \right\|_q^q \leq \frac{(2q-2)!}{(q-1)!} \left\{ V_{2,n}^{q/2} \vee V_{q,n} \right\}, \quad (5)$$

where $\|X\|_q^q = \mathbb{E}|X|^q$ and for $k = 2, \dots, q$,

$$V_{k,n} = n \sum_{r=0}^{n-1} (r+1)^{k-2} C_k(r),$$

with

$$C_k(r) := \sup\{|\text{cov}(Z_{t_1} \cdots Z_{t_p}, Z_{t_{p+1}} \cdots Z_{t_k})|\}, \quad (6)$$

where the supremum is over all the ordered k -tuples $t_1 \leq \dots \leq t_k$ such that $\sup_{1 \leq i \leq k-1} t_{i+1} - t_i = r$.

We will apply this bound when the Z_i s are defined in such a way that $\sum_{i=1}^n Z_i$ is proportional to the fluctuation term $\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)$. The inequality (5) gives a bound for this part of the deviation of the estimator which depends on the covariance bounds $C_k(r)$. The other part of the deviation is the bias, which is treated by deterministic methods. In order to obtain suitable controls of the fluctuation term, we need two different type of bounds

for $C_k(r)$. Conditions on the decay of the weak dependence coefficients give a first bound. Another type of condition is also required to bound $C_k(r)$ for the smaller values of r ; this is classically achieved with a regularity condition on the joint law of the pairs (X_j, X_k) for all $j \neq k$. In Doukhan & Louhichi (see [DL01]), rates of convergence are obtained when the coefficient η decays geometrically fast and the joint densities are bounded. We relax these conditions to cover the case when the joint distributions are not absolutely continuous and when the η - and $\tilde{\phi}$ -dependence coefficients decrease slowly (sub-geometric and Riemannian decays are considered).

Under our assumptions, we prove that (2) still holds (see Theorem 1). Unfortunately, additional losses appear for the uniform bounds. When η_r or $\tilde{\phi}_r = O(e^{-ar^b})$ with $a > 0$ and $b > 0$, we prove in Theorem 2 that (3) and (4) hold with $\log(n)$ replaced by $\log^{2(b+1)/b}(n)$. If η_r or $\tilde{\phi}_r = O(r^{-a})$ with $a > 1$, Theorem 3 gives bounds similar to (3) and (4) with the right hand side replaced by $O(n^{-q\rho/\{d+2\rho+2d/(q_0+d)\}})$ and $O(\{\log^{q_0+d}(n)/n^{q_0-2}\}^{\rho/\{2\rho q_0+d(q_0+2)\}})$, respectively, and with $q_0 = 2\lceil(a-1)/2\rceil$ (by definition $\lceil x \rceil$ is the smallest integer larger than or equal to the real number x). As already noticed in [DL01], the loss w.r.t the i.i.d. case highly depends on the decay of the dependence coefficients. In the case of geometric decay, the loss is logarithmic while it is polynomial in the case of polynomial decays.

The paper is organized as follows. In Section 2.1, we introduce the notions of η and $\tilde{\phi}$ dependence. We give the notation and hypothesis in Section 2.2. The main results are presented in Section 2.3. We then apply these results to particular cases of weak dependence processes, and we provide examples of kernel K_m in Section 3. Section 4 contains the proof of the Theorems and three important lemmas.

2 Main results

We first describe the notions of dependence considered in this paper, then we introduce assumptions and formulate the main results of the paper (convergence rates).

2.1 Weak dependence

We consider a sequence $(X_i)_{i \in \mathbb{Z}}$ of \mathbb{R}^d valued random variables, and we fix a norm $\|\cdot\|$ on \mathbb{R}^d . Moreover, if $h : \mathbb{R}^{du} \rightarrow \mathbb{R}$ for some $u \geq 1$, we define

$$\text{Lip}(h) = \sup_{(a_1, \dots, a_u) \neq (b_1, \dots, b_u)} \frac{|h(a_1, \dots, a_u) - h(b_1, \dots, b_u)|}{\|a_1 - b_1\| + \dots + \|a_u - b_u\|}.$$

Definition 1 (η -dependence, Doukhan & Louhichi (1999)). *The process $(X_i)_{i \in \mathbb{Z}}$ is η -weakly dependent if there exists a sequence of non-negative real numbers $(\eta_r)_{r \geq 0}$ satisfying $\eta_r \rightarrow 0$ when $r \rightarrow \infty$ and*

$$|\text{cov}(h(X_{i_1}, \dots, X_{i_u}), k(X_{i_{u+1}}, \dots, X_{i_{u+v}}))| \leq (u\text{Lip}(h) + v\text{Lip}(k))\eta_r,$$

for all $(u+v)$ -tuples, (i_1, \dots, i_{u+v}) with $i_1 \leq \dots \leq i_u \leq i_u + r \leq i_{u+1} \leq \dots \leq i_{u+v}$, and $h, k \in \Lambda^{(1)}$ where

$$\Lambda^{(1)} = \left\{ h : \exists u \geq 0, h : \mathbb{R}^{du} \rightarrow \mathbb{R}, \text{Lip}(h) < \infty, \|h\|_\infty = \sup_{x \in \mathbb{R}^{du}} |h(x)| \leq 1 \right\}.$$

Remark The η -dependence condition can be applied to non-causal sequences because information “from the future” (i.e. on the right of the covariance) contributes to the dependence coefficient in the same way as information “from the past” (i.e. on the left). It is the non-causal alternative to the θ condition in [DD03] and [DL99].

Definition 2 ($\tilde{\phi}$ -dependence, Dedecker & Priour (2004)). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and \mathcal{M} a σ -algebra of \mathcal{A} . For any $l \in \mathbb{N}^*$, any random variable $X \in \mathbb{R}^{dl}$ we define:

$$\tilde{\phi}(\mathcal{M}, X) = \sup\{\|\mathbb{E}(g(X)|\mathcal{M}) - \mathbb{E}(g(X))\|_\infty, g \in \Lambda_{1,l}\},$$

where $\Lambda_{1,l} = \{h : \mathbb{R}^{dl} \mapsto \mathbb{R} / \text{Lip}(h) < 1\}$. The sequence of coefficients $\tilde{\phi}_k(r)$ is then defined by

$$\tilde{\phi}_k(r) = \max_{l \leq k} \frac{1}{l} \sup_{i+r \leq j_1 < j_2 < \dots < j_l} \tilde{\phi}(\sigma(\{X_j; j \leq i\}), (X_{j_1}, \dots, X_{j_l})).$$

The process is $\tilde{\phi}$ -dependent if $\tilde{\phi}(r) = \sup_{k>0} \tilde{\phi}_k(r)$ tends to 0 with r .

Remark The $\tilde{\phi}$ dependence coefficients provide covariance bounds. For a Lipschitz function k and a bounded function h ,

$$|\text{cov}(h(X_{i_1}, \dots, X_{i_u}), k(X_{i_{u+1}}, \dots, X_{i_{u+v}}))| \leq v\mathbb{E}|h(X_{i_1}, \dots, X_{i_u})| \text{Lip}(k)\tilde{\phi}(r). \quad (7)$$

2.2 Notations and definitions

Assume that $(X_n)_{n \in \mathbb{Z}}$ is an η or $\tilde{\phi}$ dependent sequence of \mathbb{R}^d valued random variables. We consider two types of decays for the coefficients. The geometric case is the case when Assumption [H1] or [H1'] holds.

$$[\text{H1}]: \eta_r = O(e^{-ar^b}) \text{ with } a > 0 \text{ and } b > 0,$$

$$[\text{H1}']: \tilde{\phi}(r) = O(e^{-ar^b}) \text{ with } a > 0 \text{ and } b > 0.$$

The Riemannian case is the case when Assumption [H2] or [H2'] holds.

$$[\text{H2}]: \eta_r = O(r^{-a}) \text{ with } a > 1,$$

$$[\text{H2}']: \tilde{\phi}(r) = O(r^{-a}) \text{ with } a > 1.$$

As usual in density estimation, we shall assume:

[H3]: The common marginal distribution of the random variables X_n , $n \in \mathbb{Z}$ is absolutely continuous with respect to Lebesgue's measure, with common bounded density f .

The next assumption is on the density with respect to Lebesgue's measure (if it exists) of the joint distribution of the pairs (X_j, X_k) , $j \neq k$.

[H4] The density $f_{j,k}$ of the joint distribution of the pair (X_j, X_k) is uniformly bounded with respect to $j \neq k$.

Unfortunately, for some processes, these densities may not even exist. For example, the joint distributions of Markov chains $X_n = G(X_{n-1}, \epsilon_n)$ may not be absolutely continuous. One of the simplest example is

$$X_k = \frac{1}{2} (X_{k-1} + \epsilon_k) , \quad (8)$$

where $\{\epsilon_k\}$ is an i.i.d. sequence of Bernoulli random variables and X_0 is uniformly distributed on $[0, 1]$. The process $\{X_n\}$ is strictly stationary but the joint distributions of the pairs (X_0, X_k) are degenerated for any k . This Markov chain can also be represented (through an inversion of the time) as a dynamical system $(T_{-n}, \dots, T_{-1}, T_0)$ which has the same law as (X_0, X_1, \dots, X_n) (T_0 and X_0 are random variables distributed according to the invariant measure, see [BGR00] for more details). Let us recall the definition of a dynamical system.

Definition 3 (dynamical system). *A one-dimensional dynamical system is defined by*

$$\forall k \in \mathbb{N}, T_k := F^k(T_0) , \quad (9)$$

where $F : I \rightarrow I$, I is a compact subset of \mathbb{R} and in this context, F^k denotes the k -th iterate of the application F : $F^1 = F$, $F^{k+1} = F \circ F^k$, $k \geq 1$. We assume that there exists an invariant probability measure μ_0 , i.e. $F(\mu_0) = \mu_0$, absolutely continuous with respect to Lebesgue's measure, and that T_0 is a random variable with distribution μ_0 .

We restrict our study to one-dimensional dynamical systems T in the class \mathcal{F} of dynamical systems defined by a transformation F that satisfies the following assumptions (see [Pri01]).

- $\forall k \in \mathbb{N}, \forall x \in \text{int}(I)$, $\lim_{t \rightarrow 0^+} F^k(x+t) = F^k(x^+)$ and $\lim_{t \rightarrow 0^-} F^k(x+t) = F^k(x^-)$ exist;
- $\forall k \in \mathbb{N}^*$, denoting $D_+^k = \{x \in \text{int}(I), F^k(x^+) = x\}$ and $D_-^k = \{x \in \text{int}(I), F^k(x^-) = x\}$, we assume $\lambda \left(\bigcup_{k \in \mathbb{N}^*} (D_+^k \cup D_-^k) \right) = 0$, where λ is the Lebesgue measure.

When the joint distributions of the pairs (X_j, X_k) are not assumed absolutely continuous (and then [H4] is not satisfied), we shall instead assume:

[H5] The dynamical system $(X_n)_{n \in \mathbb{Z}}$ belongs to \mathcal{F} .

We consider in this paper linear estimators as in (1). The sequence of kernels K_m is assumed to satisfy the following assumptions.

- (a) The support of K_m is a compact set with diameter $O(1/m^{1/d})$;
- (b) The functions $x \mapsto K_m(x, y)$ and $x \mapsto K_m(y, x)$ are Lipschitz functions with Lipschitz constant $O(m^{1+1/d})$;
- (c) For all x in the support of K_m , $\int K_m(x, y) dy = 1$;
- (d) The bias of the estimator \hat{f}_n defined in (1) is of order $m_n^{-\rho/d}$, uniformly on compact sets.

$$\sup_{\|x\| \leq M} \left| \mathbb{E}[\hat{f}_n(x)] - f(x) \right| = O(m_n^{-\rho/d}). \quad (10)$$

2.3 Results

In all our results we consider kernels K_m and a density estimator of the form (1) such that assumptions (a), (b), (c) and (d) hold.

Theorem 1 (\mathbb{L}^q -convergence).

Geometric case. Under Assumptions [H4] or [H5] and [H1] or [H1'], the sequence m_n can be chosen such that inequality (2) holds for all $0 < q < +\infty$.

Riemannian case. Under the assumptions [H4] or [H5], if additionally

- [H2] holds with $a > \max(1 + 2/d + (d+1)/\rho, 2 + 1/d)$ (η -dependence),
 - or [H2'] holds with $a > 1 + 2/d + 1/\rho$ (ϕ -dependence),
- then the sequence m_n can be chosen such that inequality (2) holds for all $0 < q \leq q_0 = 2 \lceil (a-1)/2 \rceil$.

Theorem 2 (Uniform rates, geometric decays). For any $M > 0$, under Assumptions [H4] or [H5] and [H1] or [H1'] we have, for all $0 < q < +\infty$, and for a suitable choice of the sequence m_n ,

$$\mathbb{E} \sup_{\|x\| \leq M} |\hat{f}_n(x) - f(x)|^q = O \left(\left(\frac{\log^{2(b+1)/b}(n)}{n} \right)^{q\rho/(d+2\rho)} \right),$$

$$\sup_{\|x\| \leq M} |\hat{f}_n(x) - f(x)| =_{a.s.} O \left(\left(\frac{\log^{2(b+1)/b}(n)}{n} \right)^{\rho/(d+2\rho)} \right).$$

Theorem 3 (Uniform rates, Riemannian decays). *For any $M > 0$, under Assumptions [H4] or [H5], [H2] or [H2'] with $a \geq 4$ and $\rho > 2d$, for $q_0 = 2\lceil(a-1)/2\rceil$ and $q \leq q_0$, the sequence m_n can be chosen such that*

$$\mathbb{E} \sup_{\|x\| \leq M} |\hat{f}_n(x) - f(x)|^q = O\left(n^{-\frac{q\rho}{d+2\rho+2d/(q_0+d)}}\right),$$

or such that

$$\sup_{\|x\| \leq M} |\hat{f}_n(x) - f(x)| =_{a.s.} O\left(\left(\frac{\log^{q_0+d}(n)}{n^{q_0-2}}\right)^{\frac{\rho}{d(q_0+2)+\rho(q_0+d)}}\right).$$

Remarks.

- Theorem 1 shows that the optimal convergence rate of (2) still holds in the weak dependence context. In the Riemannian case, when $a \geq 4$, the conditions are satisfied if the density function f is sufficient regular, namely, if $\rho > d + 1$.
- The loss with respect to the i.i.d. case in the uniform convergence rates (Theorems 2 and 3) is due to the fact that the probability inequalities for dependent observations are not as good as Bernstein's inequality for i.i.d. random variables (Bernstein inequalities in weak dependence context are proved in [KN05]). The convergence rates depend on the decay of the weak dependence coefficients. This is in contrast to the case of independent observations.
- In Theorem 2 the loss is a power of the logarithm of the number of observations. Let us remark that this loss is reduced when b tends to infinity. In the case of η -dependence and geometric decreasing, the same result is in [DL99] for the special case $b = 1$. In the framework of $\tilde{\phi}$ -dependence, Theorem 2 seems to provide the first result on uniform rates of convergence for density estimators.
- In Theorem 3, the rate of convergence in the mean is better than the almost sure rate for technical reasons. Contrary to the geometric case, the loss is no longer logarithmic but is a power of n . The rate gets closer to the optimal rate as $q_0 \rightarrow \infty$, or equivalently $a \rightarrow \infty$.
- These results are new under the assumption of Riemannian decay of the weak dependence coefficients. The condition on a is similar to the condition on β in [AD03]. Even if the rates are better than in [DL01], there is a huge loss with respect to the mixing case. It would be interesting to know the minimax rates of convergence in this framework.

3 Models, applications and extensions

The class of weak dependent processes is very large. We apply our results to three examples: **two-sided moving averages**, **bilinear models** and **expanding maps**. The first two will be handled with the help of the coefficients η , the third one with the coefficients $\tilde{\phi}$.

3.1 Examples of η -dependent time series.

It is of course possible to define η -dependent random fields (see [DDLLLP04] for further details); for simplicity, we only consider processes indexed by \mathbb{Z} .

Definition 4 (Bernoulli shifts). *Let $H : \mathbb{R}^{\mathbb{Z}} \rightarrow \mathbb{R}$ be a measurable function. A Bernoulli shift is defined as $X_n = H(\xi_{n-i}, i \in \mathbb{Z})$ where $(\xi_i)_{i \in \mathbb{Z}}$ is a sequence of i.i.d random variables called the innovation process.*

In order to obtain a bound for the coefficients $\{\eta_r\}$, we introduce the following regularity condition on H . There exists a sequence $\{\delta_r\}$ such that

$$\sup_{i \in \mathbb{Z}} \mathbb{E} \left| H(\xi_{i-j}, j \in \mathbb{Z}) - H(\xi_{i-j} \mathbb{1}_{|j| < r}, j \in \mathbb{Z}) \right| \leq \delta_r,$$

Bernoulli shifts are η -dependent with $\eta_r = 2\delta_{r/2}$ (see [DL99]). In the following, we consider two special cases of **Bernoulli shifts**.

1. **Non causal linear processes.** A real valued sequence $(a_i)_{i \in \mathbb{Z}}$ such that $\sum_{j \in \mathbb{Z}} a_j^2 < \infty$ and the innovation process $\{\xi_n\}$ define a **non-causal linear process** $X_n = \sum_{-\infty}^{+\infty} a_i \xi_{n-i}$. If we control a moment of the innovations, the **linear process** (X_n) is η -dependent. The sequence $\{\eta_r\}_{r \in \mathbb{N}}$ is directly linked to the coefficients $\{a_i\}_{i \in \mathbb{Z}}$ and various types of decay may occur. We consider only Riemannian decays $a_i = \mathcal{O}(i^{-A})$ with $A \geq 5$ since results for geometric decays are already known. Here $\eta_r = \mathcal{O}\left(\sum_{|i| > r/2} a_i\right) = \mathcal{O}(r^{1-A})$ and [H2] holds. Furthermore, we assume that the sequence $(\xi_i)_{i \in \mathbb{Z}}$ is i.i.d. and satisfies the condition $|\mathbb{E}e^{iu\xi_0}| \leq C(1 + |u|)^{-\delta}$, for all $u \in \mathbb{R}$ and for some $\delta > 0$ and $C < \infty$. Then, the densities f and $f_{j,k}$ exist for all $j \neq k$ and they are uniformly bounded (see the proof in the causal case in Lemma 1 and Lemma 2 in [GKS96]); hence [H4] holds. If the density f of X_0 is ρ -regular with $\rho > 2$, our estimators converge to the density with the rates:
 - $n^{-\rho/(2\rho+1)}$ in \mathbb{L}^q -norm ($q \leq 4$) at each point x ,
 - $n^{-\rho/(2\rho+3/2)}$ in \mathbb{L}^q -norm ($q \leq 4$) uniformly on an interval,
 - $(\log^4(n)/n)^{\rho/(4\rho+3)}$ almost surely on an interval.

In the first case, the rate we obtain is the same as in the i.i.d. case. For such linear models, the density estimator also satisfies the Central Limit Theorem (see [HLT01] and [Ded98]).

2. **Bilinear model.** The process $\{X_t\}$ is a **bilinear model** if there exist two sequences $(a_i)_{i \in \mathbb{N}^*}$ and $(b_i)_{i \in \mathbb{N}^*}$ of real numbers and real numbers a and b such that:

$$X_t = \xi_t \left(a + \sum_{j=1}^{\infty} a_j X_{t-j} \right) + b + \sum_{j=1}^{\infty} b_j X_{t-j}. \quad (11)$$

Squared **ARCH**(∞) or **GARCH**(p, q) processes satisfy such an equation, with $b = b_j = 0$ for all $j \geq 1$. Define

$$\lambda = \|\xi_0\|_p \sum_{j=1}^{\infty} a_j + \sum_{j=1}^{\infty} b_j.$$

If $\lambda < 1$, then the equation (11) has a strictly stationary solution in L^p (see [DMR05]). This solution is a **Bernoulli shift** for which we have the behavior of the coefficient η :

- $\eta_r = O(\mathbb{E}^{-\lambda r})$ for some $\lambda > 0$ if there exists an integer N such that $a_i = b_i = 0$ for $i \geq N$.
- $\eta_r = O(\mathbb{E}^{-\lambda \sqrt{r}})$ for some $\lambda > 0$ if $a_i = O(\mathbb{E}^{-Ai})$ and $b_i = O(\mathbb{E}^{-Bi})$ with $A > 0$ and $B > 0$.
- $\eta_r = O(\{r/\log(r)\}^{-\lambda})$ for some $\lambda > 0$ if $a_i = O(i^{-A})$ and $b_i = O(i^{-B})$ with $A > 1$ and $B > 1$.

Let us assume that the i.i.d. sequence $\{\xi_t\}$ has a marginal density $f_\xi \in C_\rho$, for some $\rho > 2$. The density of X_t conditionally to the past can be written as a function of f_ξ . We then check recursively that the common density of X_t for all t , say f , also belongs to C_ρ . Furthermore, the regularity of f_ξ ensures that f and the joint densities $f_{j,k}$ for all $j \neq k$ are bounded (see [DMR05]) and [H4] holds. The assumptions of Theorem 1 are satisfied, and the estimator \hat{f}_n achieves the minimax bound (2) if either:

- There exists an integer N such that $a_i = b_i = 0$ for $i \geq N$;
- There exist $A > 0$ and $B > 0$ such that $a_i = O(e^{-Ai})$ and $b_i = O(e^{-Bi})$;
- There exist $A \geq 4$ and $B \geq 5$ such that $a_i = O(i^{-A})$ and $b_i = O(i^{-B})$. Then, this optimal bound holds only for $2 \leq q < q(A, B)$ where $q(A, B) = 2[(B-1) \wedge A]/2$.

Note finally that the rates of uniform convergence provided by Theorems 2 and 3 are sub-optimal.

3.2 Examples of $\tilde{\phi}$ -dependent time series.

Let us introduce an important class of **dynamical systems**:

Example 1. $(T_i = F^i(T_0))_{i \in \mathbb{N}}$ is an **expanding map** or equivalently F is a Lasota-Yorke function if it satisfies the three following criteria.

- (Regularity) There exists a grid $0 = a_0 \leq a_1 \leq \dots \leq a_n = 1$ such as $F \in \mathcal{C}_1$ and $|F'(x)| > 0$ on $]a_{i-1}, a_i[$ for each $i = 1, \dots, n$.
- (Expansivity) Let I_n be the set on which $(F^n)'$ is defined. There exists $A > 0$ and $s > 1$ such that $\inf_{x \in I_n} |(F^n)'| > As^n$.
- (Topological mixing) For any nonempty open sets U, V , there exists $n_0 \geq 1$ such as $F^{-n}(U) \cap V \neq \emptyset$ for all $n \geq n_0$.

Examples of **Markov chains** $X_n = G(X_{n+1}, \epsilon_n)$ associated to an **expanding map** $\{T_n\}$ belonging to \mathcal{F} are given in [BGR00] and [DP04]. The simplest one is $X_k = (X_{k-1} + \epsilon_k)/2$ where the ϵ_k follows a binomial law and X_0 is uniformly distributed on $[0, 1]$. We easily check that $F(x) = 2x \bmod 1$, the transformation of the associated **dynamical system** T_n , satisfies all the assumptions such as T_n is an **expanding map** belonging to \mathcal{F} .

The coefficients of $\tilde{\phi}$ -dependence of such a **Markov chain** satisfy $\tilde{\phi}(r) = O(e^{-ar})$ for some $a > 0$ (see [DP04]). Theorems 1 and 2 give the \mathbb{L}^q rate $n^{-\rho/(2\rho+1)}$, the uniform \mathbb{L}^q rate and the almost sure rate $(\log^4(n)/n)^{\rho/(2\rho+1)}$ of the estimators of the density of μ_0 .

3.3 Sampled process

Since we do not assume stationarity of the observed process, the following observation scheme is covered by our results. Let $(x_n)_{n \in \mathbb{Z}}$ be a stationary process whose marginal distribution is absolutely continuous, let $(h_n)_{n \in \mathbb{Z}}$ be a sequence of monotone functions and consider the sampled process $\{X_{i,n}\}_{1 \leq i \leq n}$ defined by $X_{i,n} = x_{h_n(i)}$. The dependence coefficients of the sampled process may decay to zero faster than the underlying unobserved process. For instance, if the dependence coefficients of the process $(x_n)_{n \in \mathbb{Z}}$ have a Riemannian decay, those of the sampled process $\{x_{h_n(i)}\}$ with $h_n(i) = i2^n$ decay geometrically fast. The observation scheme is thus a crucial factor that determines the rate of convergence of density estimators.

3.4 Density estimators and bias

In this section, we provide examples of kernels K_m and smoothness assumptions on the density f such that assumptions (a), (b), (c) and (d) of subsection 2.2 are satisfied.

Kernel estimators The kernel estimator associated to the bandwidth parameter m_n is defined by:

$$\hat{f}_n(x) = \frac{m_n}{n} \sum_{i=1}^n K\left(m_n^{1/d}(x - X_i)\right).$$

We briefly recall the classical analysis for the deterministic part R_n in this case (see [Tsy04]). Since the sequence $\{X_n\}$ has a constant marginal distribution, we have $\mathbb{E}[\hat{f}_n(x)] = f_n(x)$ with $f_n(x) = \int_D K(s)f\left(x - s/m_n^{1/d}\right) ds$. Let us assume that K is a Lipschitz function compactly supported in $D \subset \mathbb{R}^d$. For $\rho > 0$, let K satisfy, for all $j = j_1 + \dots + j_d$ with $(j_1, \dots, j_d) \in \mathbb{N}^d$:

$$\int x_1^{j_1} \dots x_d^{j_d} K(x_1, \dots, x_d) dx_1 \dots dx_d = \begin{cases} 1 & \text{if } j = 0, \\ 0 & \text{for } j \in \{1, \dots, \lceil \rho - 1 \rceil - 1\}, \\ \neq 0 & \text{if } j = \lceil \rho - 1 \rceil. \end{cases}$$

Then the kernels $K_m(x, y) = mK(m^{1/d}(x - y))$ satisfy (a), (b) and (c). Assumption (d) holds and if $f \in \mathcal{C}_\rho$, where \mathcal{C}_ρ is the class of function f such that for $\rho = \lceil \rho - 1 \rceil + c$ with $0 < c \leq 1$, f is $\lceil \rho - 1 \rceil$ -times continuously differentiable and there exists $A > 0$ such that $\forall (x, y) \in \mathbb{R}^d \times \mathbb{R}^d$, $|f^{(\lceil \rho - 1 \rceil)}(x) - f^{(\lceil \rho - 1 \rceil)}(y)| \leq A|x - y|^c$.

Projection estimators We only consider in this section the case $d = 1$. Under the assumption that the family $\{1, x, x^2, \dots\}$ belongs to $L^2(I, \mu)$, where I is a bounded interval of \mathbb{R} and μ is a measure on I , an orthonormal basis of $L^2(I, \mu)$ can be defined which consists of polynomials $\{P_0, P_1, P_2, \dots\}$. We assume that f belongs to a class \mathcal{C}'_ρ which is slightly more restrictive than the class \mathcal{C}_ρ (see Theorem 6.23 p.218 in [DS01] for details). Then for any $f \in L^2(I, \mu) \cap \mathcal{C}'_\rho$, there exists a function $\pi_{f, m_n} \in V_{m_n}$ such that $\sup_{x \in I} |f(x) - \pi_{f, m_n}(x)| = O(m_n^{-\rho})$. Consider then the projection $\pi_{m_n} f$ of f on the subspace $V_{m_n} = \text{Vect}\{P_0, P_1, \dots, P_{m_n}\}$. It can be expressed as

$$\pi_{m_n} f(x) = \sum_{j=0}^{m_n} \left\{ \int_I P_j(s) f(s) d\mu(s) \right\} P_j(x).$$

The projection estimator of the density f of the real valued random variables $\{X_i\}_{1 \leq i \leq n}$ is naturally defined as

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n K_{m_n}(x, X_i) = \frac{1}{n} \sum_{i=1}^n \sum_{j=0}^{m_n} P_j(X_i) P_j(x).$$

Then $\mathbb{E} \hat{f}_n(x) = \pi_{m_n} f(x)$ is an approximation of $f(x)$ in V_{m_n} . The fact that I is compact and the Christoffel-Darboux formula and its corollary (see [Sze33]) ensure properties (a) and (b) for the kernels K_m . We easily check that properties (c) also holds. Unfortunately, the optimal rate $(m_n^{-\rho})$ does not necessarily hold. We then have to consider the weighted kernels $K_m^a(x, y)$ defined by:

$$K_m^a(x, y) = \sum_{j=0}^m a_{m,j} \sum_{k=0}^j P_k(x) P_k(y),$$

where $\{a_{m,j}; m \in \mathbb{N}, 0 \leq j \leq m\}$ is a weight sequence satisfying $\sum_{j=0}^m a_{m,j} = 1$ and for all j : $\lim_{m \rightarrow \infty} a_{m,j} = 0$. If the sequence $\{a_{m,j}\}$ is such that K_m^a is a nonnegative kernel then $\|K_m^a\|_1 = \int_I K_m^a(x, s) d\mu(s) = 1$ and the kernel K_m^a satisfies (a), (b) and (c). Moreover, the uniform norm of the operator $f \mapsto K_m^a * f(x)$ is $\sup_{\|f\|_\infty=1} \|K_m^a * f\|_\infty = \|K_m^a\|_1 = 1$. The linear estimator built with this kernel is

$$\hat{f}_n^a(x) = \frac{1}{n} \sum_{i=1}^n \sum_{j=0}^{m_n} a_{m_n,j} \sum_{k=0}^j P_k(X_i) P_k(x),$$

and its bias has the optimal rate:

$$\begin{aligned}
|\mathbb{E}\hat{f}_n^a(x) - f(x)| &= |K_{m_n}^a * f(x) - \pi_{f,m_n} f(x) + \pi_{f,m_n} f(x) - f(x)|, \\
&\leq |K_{m_n}^a * (f(x) - \pi_{f,m_n} f(x)) + \pi_{f,m_n} f(x) - f(x)|, \\
&\leq (\|K_{m_n}^a\|_1 + 1)m_n^{-\rho} = \mathcal{O}(m_n^{-\rho}).
\end{aligned}$$

Such an array $\{a_{m,j}\}$ cannot always be defined. We give an example where it is possible.

Example 2 (Fejer kernel). For the trigonometric basis $\{\cos(nx), \sin(nx)\}_{n \in \mathbb{N}}$, we can find a 2π -periodic function $f \in \mathcal{C}'_1$ such that $\sup_{x \in [-\pi; \pi]} |f(x) - \pi_m f(x)| = O(m^{-1} \log m)$. The associated estimator reads:

$$\hat{f}_n(x) = \frac{1}{2\pi} + \frac{1}{n\pi} \sum_{i=1}^n \sum_{k=1}^{m_n} \cos(kX_i) \cos(kx) + \sin(kX_i) \sin(kx).$$

We remark that $\mathbb{E}\hat{f}_n$ is the Fourier series of f truncated at order m_n :

$$D_{m_n} f(x) = \frac{1}{2\pi} \int_0^{2\pi} f(t) D_{m_n}(x-t) dt.$$

where

$$D_m(x) = \sum_{k=-m}^m \mathbb{E}^{ikx} = \frac{\sin(\{2m+1\}x/2)}{\sin(x/2)}$$

is (the symmetric) Dirichlet's kernel. Recall that Fejer's kernel is defined as

$$F_m(x) = \frac{1}{m} \sum_{k=0}^{m-1} D_k(x) = \sum_{k=-(m-1)}^{m-1} \left(1 - \frac{|k|}{m}\right) \mathbb{E}^{ikx} = \frac{\sin^2(mx/2)}{m \sin^2(x/2)}.$$

The kernel F_m is a nonnegative weighted kernel corresponding to Dirichlet's kernel and the sequence of weights $a_{m,j} = 1/m$ and satisfies (a), (b) and (c). The estimator associated to the Fejer's kernels is defined by

$$\tilde{f}_n(x) = \frac{1}{2\pi} + \frac{1}{n\pi} \sum_{i=1}^n \sum_{j=1}^{m_n} \frac{1}{m_n} \sum_{k=1}^j \cos kX_i \cos kx + \sin kX_i \sin kx,$$

If the common density f is 2π -periodic and belongs to \mathcal{C}'_1 , then assumption (d) holds.

Using general Jackson's kernels (see [DS01]), we can find an estimator such that $R_n = O(m_n^{-\rho/d})$ for other values of ρ , but the weight sequence $a_{m,j}$ highly depends of the value of ρ .

Wavelet estimation Wavelet estimation is a particular case of projection estimation. For the sake of simplicity, we restrict hte study to $d = 1$.

Definition 5 (Scaling function [Dou88]). A function $\phi \in L^2(\mathbb{R})$ is called a scaling function if the family $\{\phi(\cdot - k); k \in \mathbb{Z}\}$ is orthonormal.

We choose the bandwidth parameter $m_n = 2^{j(n)}$ and define $V_j = \text{Vect}\{\phi_{j,k}, k \in \mathbb{Z}\}$, where $\phi_{j,k} = 2^{j/2}\phi(2^j(x-k))$. Under the assumption that ϕ is compactly supported, we define (the sum over the index k is in fact finite):

$$\hat{f}_n(x) = \frac{1}{n} \sum_{k=-\infty}^{\infty} \sum_{i=1}^n \phi_{j(n),k}(X_i) \phi_{j(n),k}(x) .$$

The wavelets estimator is of the form (1) with $K(x, y) = \sum_{k=-\infty}^{\infty} \phi(y-k)\phi(x-k)$ and $K_m(x, y) = mK(mx, my)$. Under the additional assumption that $\sum_{k \in \mathbb{Z}} \phi(x-k) = 1$ for almost all x , we can write:

$$\begin{aligned} \left| \mathbb{E}(\hat{f}_n(x) - f(x)) \right| &\leq \left| \int K_{m_n}(y, x) f(y) dy - f(x) \right| , \\ &= \left| \int m_n K(m_n y, m_n x) (f(y) - f(x)) dy \right| , \\ &= \left| \int m_n K(m_n x + t, m_n x) (f(x + t/m_n) - f(x)) dt \right| . \end{aligned}$$

If ϕ is a Lipschitz function such that $\int \phi(x)x^j dx = 0$ if $0 < j < \lceil \rho - 1 \rceil$ and $\int \phi(x)x^{\lceil \rho - 1 \rceil} dx \neq 0$, then the kernel K_m satisfy properties (a), (b) and (c). If $f \in C_\rho$, then Assumption (d) holds.

4 Proof of the Theorems

The proof of our results is based on the decomposition:

$$\hat{f}_n(x) - f(x) = \underbrace{\hat{f}_n(x) - \mathbb{E}(\hat{f}_n(x))}_{FL_n(x)=\text{fluctuation}} + \underbrace{\mathbb{E}(\hat{f}_n(x)) - f(x)}_{\text{bias}} . \quad (12)$$

The bias term is of order $m_n^{-\rho/d}$ by Assumption (d). We now present three lemmas useful to derive the rate of the fluctuation term.

Lemma 1 (Moment inequalities). *For each even integer q , under the assumption [H4] or [H5] and if moreover one of the following assumption holds:*

- [H1] or [H1'] holds (geometric case);
- [H2] holds, $m_n = n^\delta \log(n)^\gamma$ with $\delta > 0$, $\gamma \in \mathbb{R}$ and

$$a > \max \left(q - 1, \frac{(q-1)\delta(4+2/d)}{q-2+\delta(4-q)}, 2 + \frac{1}{d} \right) ,$$

- [H2'] holds, $m_n = n^\delta \log(n)^\gamma$ with $\delta > 0$ and $\gamma \in \mathbb{R}$ and

$$a > \max \left(q - 1, \frac{(q-1)\delta(2+2/d)}{q-2+\delta(4-q)}, 1 + \frac{1}{d} \right) .$$

Then, for each $x \in \mathbb{R}^d$,

$$\limsup_{n \rightarrow \infty} (n/m_n)^{q/2} \|FL_n(x)\|_q^q < +\infty .$$

Lemma 2 (Probability inequalities).

- *Geometric case.* Under Assumptions [H4] or [H5] and [H1] or [H1'] there exist positive constants C_1, C_2 such that

$$\mathbb{P} \left(|FL_n(x)| \geq \epsilon \sqrt{m_n/n} \right) C_1 \leq \exp\{-C_2 \epsilon^{b/(b+1)}\} .$$

- *Riemannian case.* Under Assumptions [H4] or [H5], if $m_n = n^\delta \log(n)^\gamma$ and if one of the following assumptions holds:
 - [H2] with $a > \max\{1 + 2(\delta + 1/d)/(1 - \delta), 2 + 1/d\}$,
 - [H2'] with $a > \max(1 + 2\{1/d(1 - \delta)\}, 1 + 1/d)$,
 then,

$$\mathbb{P} \left(|FL_n(x)| \geq \epsilon \sqrt{m_n/n} \right) \leq C \epsilon^{-q_0} ,$$

with $q_0 = 2 \lceil (a - 1)/2 \rceil$.

Lemma 3 (Fluctuation rates). Under the assumptions of Lemma 2, we have for any $M > 0$,

- *Geometric case.*

$$\sup_{\|x\| \leq M} |FL_n(x)| =_{a.s.} O \left(\sqrt{\frac{m_n}{n}} \log^{(b+1)/b}(n) \right) ;$$

- *Riemannian case.*

$$\sup_{\|x\| \leq M} |FL_n(x)| =_{a.s.} O \left(\left(\frac{m_n^{1+2/q_0}}{n^{1-2/q_0}} \right)^{\frac{1}{1+d/q_0}} \log n \right) ,$$

with $q_0 = 2 \lceil (a - 1)/2 \rceil$.

Remarks.

- In Lemma 1, we improve the moment inequality of [DL01], where the condition in the case of coefficient η is $a > 3(q - 1)$, which is always stronger than our condition.
- In the i.i.d. case a Bernstein type inequality is available:

$$\mathbb{P} \left(|FL_n(x)| \geq \epsilon \sqrt{\frac{m}{n}} \right) \leq C_1 \exp(-C_2 \epsilon^2) ,$$

Lemma 2 provides a weaker inequality for dependent sequences. Other probability inequalities for dependent sequences are presented in [DP04] and [KN05].

- Lemma 3 gives the almost sure bounds for the fluctuation. It is derived directly from the two previous lemmas.

Proof of the lemmas

Proof (Proof of Lemma 1). Let x be a fixed point in \mathbb{R}^d . Denote $Z_i = u_n(X_i) - \mathbb{E}u_n(X_i)$ where $u_n(\cdot) = K_{m_n}(\cdot, x)/\sqrt{m_n}$. Then

$$\sum_{i=1}^n Z_i = \sum_{i=1}^n u_n(X_i) - \mathbb{E}u_n(X_i) = \frac{n}{\sqrt{m_n}}(\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)) = \frac{n}{\sqrt{m_n}}FL_n(x). \quad (13)$$

The order of magnitude of the fluctuation $FL_n(x)$ is obtained by applying the inequality (5) to the centered sequence $\{Z_i\}_{1 \leq i \leq n}$ defined above. We then control the normalized fluctuation of (13) with the covariance terms $C_k(r)$ defined in equation (6). Firstly, we bound the covariance terms:

- **Case $r = 0$.** Here $t_1 = \dots = t_k = i$. Then we get:

$$C_k(r) = |\text{cov}(Z_{t_1} \cdots Z_{t_p}, Z_{t_{p+1}} \cdots Z_{t_k})| \leq 2\mathbb{E}|Z_i|^k.$$

By definition of Z_i :

$$\mathbb{E}|Z_i|^k \leq 2^k \mathbb{E}|u_n(X_i)|^k \leq 2^k \|u_n\|_\infty^{k-1} \mathbb{E}|u_n(X_0)|. \quad (14)$$

- **Case $r > 0$.** $C_k(r) = |\text{cov}(Z_{t_1} \cdots Z_{t_p}, Z_{t_{p+1}} \cdots Z_{t_k})|$ is bounded in different ways, either using weak-dependence property or by direct bound.
 - **Weak-dependence bounds:**
 - *η -dependence:* Consider the following application:

$$\phi_p : (x_1, \dots, x_p) \mapsto (u_n(x_1) \cdots u_n(x_p)).$$

Then $\|\phi_p\|_\infty \leq 2^p \|u_n\|_\infty^p$ and $\text{Lip } \phi_p \leq 2^p \|u_n\|_\infty^{p-1} \text{Lip } u_n$. Thus by η -dependence, for all $k \geq 2$ we have:

$$\begin{aligned} C_k(r) &\leq (p2^p \|u_n\|_\infty^{p-1} + (k-p)2^{p-k} \|u_n\|_\infty^{p-k-1}) \text{Lip } u_n \eta(r), \\ &\leq k2^k \|u_n\|_\infty^{k-1} \text{Lip } u_n \eta(r). \end{aligned} \quad (15)$$

- *φ -dependence:* We use the inequality (7). Using the bound

$$\mathbb{E}|\phi_p(X_1, \dots, X_p)| \leq \|u_n\|_\infty^{p-1} \mathbb{E}|u_n(X_0)|,$$

we derive a bound for the covariance terms:

$$C_k(r) \leq k2^k \|u_n\|_\infty^{k-2} \mathbb{E}|u_n(X_0)| \text{Lip } u_n \tilde{\varphi}(r). \quad (16)$$

- **Direct bound:** Triangular inequality implies for $C_k(r)$:

$$|\text{cov}(Z_{t_1} \cdots Z_{t_p}, Z_{t_{p+1}} \cdots Z_{t_k})| \leq \underbrace{\left| \mathbb{E} \prod_{i=1}^k Z_{t_i} \right|}_A + \underbrace{\left| \mathbb{E} \prod_{i=1}^p Z_{t_i} \right|}_{B_p} \underbrace{\left| \mathbb{E} \prod_{i=p+1}^k Z_{t_i} \right|}_{B_{k-p}},$$

$$\begin{aligned}
A &= |\mathbb{E}(u_n(X_{t_1}) - \mathbb{E}u_n(X_{t_1})) \cdots (u_n(X_{t_k}) - \mathbb{E}u_n(X_{t_k}))| , \\
&= |\mathbb{E}u_n(X_0)|^k + |\mathbb{E}(u_n(X_{t_1}) \cdots u_n(X_{t_k}))| \\
&\quad + \sum_{s=1}^{k-1} |\mathbb{E}u_n(X_0)|^{k-s} \sum_{t_{i_1} \leq \cdots \leq t_{i_s}} |\mathbb{E}(u_n(X_{t_{i_1}}) \cdots u_n(X_{t_{i_s}}))| .
\end{aligned}$$

Firstly, with $k \geq 2$:

$$|\mathbb{E}u_n(X_0)|^k \leq \|u_n\|_\infty^{k-2} (\mathbb{E}|u_n(X_0)|)^2 .$$

Secondly, if $1 \leq s \leq k-1$:

$$\begin{aligned}
|\mathbb{E}(u_n(X_{t_{i_1}}) \cdots u_n(X_{t_{i_s}}))| &\leq \mathbb{E}|u_n(X_{t_{i_1}}) \cdots u_n(X_{t_{i_s}})| , \\
&\leq \|u_n\|_\infty^{s-1} \mathbb{E}|u_n(X_0)| , \\
|\mathbb{E}u_n(X_0)|^{k-s} &\leq \|u_n\|_\infty^{k-s-1} \mathbb{E}|u_n(X_0)| .
\end{aligned}$$

Thirdly there is at least two different observations with a gap of $r > 0$ among X_{t_1}, \dots, X_{t_k} so for any integer $k \geq 2$:

$$|\mathbb{E}(u_n(X_{t_1}) \cdots u_n(X_{t_k}))| \leq \|u_n\|_\infty^{k-2} \mathbb{E}|u_n(X_0)u_n(X_r)| .$$

Then, collecting the last four inequations yields:

$$\begin{aligned}
A &\leq \|u_n\|_\infty^{k-2} (\mathbb{E}|u_n(X_0)|)^2 \\
&\quad + (\mathbb{E}|u_n(X_0)|)^2 \sum_{s=1}^{k-1} C_s^k \|u_n(X_0)\|_\infty^{k-2} + \|u_n\|_\infty^{k-2} \mathbb{E}|u_n(X_0)u_n(X_r)| .
\end{aligned}$$

So:

$$A \leq \|u_n\|_\infty^{k-2} ((2^k - 1)(\mathbb{E}|u_n(X_0)|)^2 + \mathbb{E}|u_n(X_0)u_n(X_r)|) . \quad (17)$$

Now, we bound B_i with $i < k$. As before:

$$\begin{aligned}
B_i &= |\mathbb{E}(u_n(X_{t_1}) - \mathbb{E}u_n(X_{t_1})) \cdots (u_n(X_{t_i}) - \mathbb{E}u_n(X_{t_i}))| , \\
&= \sum_{s=0}^i |\mathbb{E}(u_n(X_0))|^{i-s} \sum_{t_{j_1} \leq \cdots \leq t_{j_s}} |\mathbb{E}(u_n(X_{t_{j_1}}) \cdots u_n(X_{t_{j_s}}))| , \\
&\leq 2^i \|u_n\|_\infty^{i-2} (\mathbb{E}|u_n(X_0)|)^2 .
\end{aligned}$$

Then:

$$B_p \times B_{k-p} \leq 2^k \|u_n\|_\infty^{k-4} (\mathbb{E}|u_n(X_0)|)^4 \leq 2^k \|u_n\|_\infty^{k-2} (\mathbb{E}|u_n(X_0)|)^2 . \quad (18)$$

Another interesting bound for $r > 0$ follows, because according to inequalities (17) and (18) we have:

$$C_k(r) \leq \|u_n\|_\infty^{k-2} ((2^{k+1} - 1)(\mathbb{E}|u_n(X_0)|)^2 + \mathbb{E}|u_n(X_0)u_n(X_r)|) .$$

Noting $\gamma_n(r) = \mathbb{E}|u_n(X_0)u_n(X_r)| \vee (\mathbb{E}|u_n(X_0)|)^2$, we have:

$$C_k(r) \leq 2^{k+1} \|u_n\|_\infty^{k-2} \gamma_n(r) . \quad (19)$$

We now use the different values of the bounds in inequalities (14), (15), (16) and (19). If we define the sequence $(w_r)_{0 \leq r \leq n-1}$ as:

- $w_0 = 1$,
- $w_r = \gamma_n(r) \wedge \|u_n\|_\infty \text{Lip } u_n \eta(r) \wedge \mathbb{E}|u_n(X_0)| \text{Lip } u_n \varphi(r)$,

then, for all r such that $0 \leq r \leq n-1$ and for all $k \geq 2$:

$$C_k(r) \leq k2^k \|u_n\|_\infty^{k-2} w_r .$$

We derive from this inequality and from (5):

$$\begin{aligned} \left\| \sum_{i=1}^n Z_i \right\|_q^q &\leq \frac{(2q-2)!}{(q-1)!} \left\{ \left(n \sum_{r=0}^{n-1} C_2(r) \right)^{q/2} \vee n \sum_{r=0}^{n-1} (r+1)^{q-2} C_q(r) \right\} , \\ &\leq (4q\sqrt{n})^q \left\{ \left(\sum_{r=0}^{n-1} w_r \right)^{q/2} \vee \left(\frac{\|u_n\|_\infty}{\sqrt{n}} \right)^{q-2} \sum_{r=0}^{n-1} (r+1)^{q-2} w_r \right\} . \end{aligned}$$

In order to control w_r , we give bounds for the terms $\gamma_n(r) = \mathbb{E}|u_n(X_0)u_n(X_r)| \vee (\mathbb{E}|u_n(X_0)|)^2$:

- In the case of [H4], we have:

$$\begin{aligned} \mathbb{E}|u_n(X_0)u_n(X_r)| &\leq \sup_{j,k} \|f_{j,k}\|_\infty \|u_n\|_1^2 , \\ (\mathbb{E}|u_n(X_0)|)^2 &\leq \|f\|_\infty^2 \|u_n\|_1^2 . \end{aligned}$$

- In the case of [H5], Lemma 2.3 of [Pri01] proves that $\mathbb{E}|u_n(X_0)u_n(X_r)| \leq (\mathbb{E}|u_n(X_0)|)^2$ for n sufficiently large and the same bound as above remains true for the last term.

In both cases, we conclude that $\gamma_n(r) \preceq \|u_n\|_1^2$. The properties (a), (b) and (c) of section 2.2 and the expression $u_n = K_{m_n}/\sqrt{m_n}$ ensure that $\|u_n\|_1 \preceq \frac{1}{\sqrt{m_n}}$ and $\text{Lip } u_n \preceq m_n^{1/2+1/d}$. We deduce using the compact support of K that $\|u_n\|_\infty \preceq \sqrt{m_n}$. We then have for $r \geq 1$:

$$w_r \preceq \frac{1}{m_n} \wedge m_n^{1+1/d} \eta(r) \wedge m_n^{1/d} \varphi(r) . \quad (20)$$

In order to prove Lemma 1, it remains to control the sums

$$\left(\frac{\|u_n\|_\infty}{\sqrt{n}} \right)^{k-2} \sum_{r=0}^{n-1} (r+1)^{k-2} \frac{1}{m_n} \wedge m_n^{1+1/d} \eta(r) \wedge m_n^{1/d} \varphi(r) , \quad (21)$$

for $k = 2$ and $k = q$ in both Riemannian and geometric cases.

- **Geometric case.**

Under [H1] or [H1']: We remark that $a \wedge b \leq a^\alpha b^{1-\alpha}$ for all $\alpha \in [0; 1]$. Using (20), we obtain first that $w_r \preceq (\eta(r) \wedge \varphi(r))^\alpha m_n^{\alpha(1+1/d)-(1-\alpha)}$ for n sufficiently large. Then for $0 < \alpha \leq \frac{d}{2d+1}$ we bound w_r independently of m_n : $w_r \preceq (\eta(r) \wedge \varphi(r))^\alpha$. For all even integer $k \geq 2$ we derive from the form of $\eta(r) \wedge \phi_r$ that (in the third inequality $u = ar^b$):

$$\begin{aligned} \sum_{r=1}^{n-1} (r+1)^{k-2} w_r &\preceq \sum_{r=0}^{n-1} (r+1)^{k-2} \exp(-\alpha ar^b), \\ &\preceq \int_0^\infty r^{k-2} \exp(-\alpha ar^b) dr, \\ &\preceq \frac{1}{ba^{\frac{k-1}{b}}} \int_1^\infty u^{\frac{k-1}{b}-1} \exp(-u) du, \\ &\preceq \frac{1}{ba^{\frac{k-1}{b}}} \Gamma\left(\frac{k-1}{b}\right). \end{aligned}$$

Using the Stirling formula, we can find a constant B such that, for the special cases $k = 2$ and $k = q$:

$$\sum_{r=1}^{n-1} (r+1)^{k-2} w_r \preceq \frac{1}{ba^{\frac{k-1}{b}}} \Gamma\left(\frac{k-1}{b}\right) \preceq (Bk)^{\frac{k}{b}}.$$

- **Riemannian case.**

Under [H6] and [H2]: Let us recall that [H6] implies that $m_n \leq n^\delta$ for n sufficiently large and $0 < \delta < 1$ and that the assumption of Lemma 1 implies that:

$$a > \max\left(q-1, \frac{\delta(q-1)(4+2/d)}{q-2+\delta(4-q)}, 2+\frac{1}{d}\right).$$

Then, we have $a > \max\left(k-1, \frac{\delta(k-1)(4+2/d)}{k-2+\delta(4-k)}\right)$ for both cases $k = q$ or $k = 2$. This assumption on a implies that:

$$\frac{(k+2/d)\delta+2-k}{2(a-k+1)} < \frac{(4-k)\delta+k-2}{2(k-1)}.$$

Furthermore, reminding that $0 < \delta < 1$:

$$0 < \frac{(4-k)\delta+k-2}{2(k-1)} = 1 - \frac{k(1+\delta)-4\delta}{2(k-1)} \leq 1.$$

We derive from the two previous inequalities that there exists $\zeta_k \in]0, 1[$ verifying

$$\frac{(k+2/d)\delta+2-k}{2(a-k+1)} < \zeta_k < \frac{(4-k)\delta+k-2}{2(k-1)}.$$

For $k = q$ or $k = 2$, we now use Tran's technique as in [ABD02]. We divide the sum (21) in two parts in order to bound it by sequences tending to 0, due to the choice of ζ_k :

$$\begin{aligned} \left(\sqrt{\frac{m_n}{n}}\right)^{k-2} \sum_{r=0}^{\lfloor n^{\zeta_k} \rfloor - 1} (r+1)^{k-2} w_r &\leq \left(\sqrt{\frac{m_n}{n}}\right)^{k-2} \frac{\lfloor n^{\zeta_k} \rfloor^{k-1}}{m_n}, \\ &\leq n^{(2\zeta_k(k-1) - ((4-k)\delta + k - 2))/2}, \\ &= O(1), \\ \left(\sqrt{\frac{m_n}{n}}\right)^{k-2} \sum_{r=\lfloor n^{\zeta_k} \rfloor}^{n-1} (r+1)^{k-2} w_r &\leq \left(\sqrt{\frac{m_n}{n}}\right)^{k-2} m_n^{1+1/d} \lfloor n^{\zeta_k} \rfloor^{k-1-a}, \\ &\leq n^{(-2\zeta_k(a-k-1) + ((k+2/d)\delta + 2 - k))/2}, \\ &= O(1). \end{aligned}$$

Under [H6] and [H2']: Under the assumption of Lemma 1:

$$a > \max\left(q-1, \frac{\delta(q-1)(2+2/d)}{q-2+\delta(4-q)}, 1+\frac{1}{d}\right),$$

we derive exactly as in the previous case that there exists $\zeta_k \in]0; 1[$ for $k = q$ or $k = 2$ such that

$$\frac{(k-2+2/d)\delta+2-k}{2(a-k+1)} < \zeta_k < \frac{(4-k)\delta+k-2}{2(k-1)}.$$

We then apply again the Tran's technique that bound the sum (21) in that case.

Lemma 1 directly follow from (13). \square

Remarks. We have in fact proved the following sharper result. In the geometric case, there exists a constant $C > 0$ such that for every even integer q and all x in the support of K_{m_n} :

$$\left\| \frac{1}{\sqrt{nm_n}} \sum_{i=1}^n K_{m_n}(X_i, x) - \mathbb{E}K_{m_n}(X_i, x) \right\|_q^q \leq (Cq^{1+1/b})^q. \quad (22)$$

Proof (Proof of Lemma 2). The cases of Riemannian or geometric decay of the dependence coefficients are considered separately.

- **Geometric decay** We present a technical lemma useful to deduce exponential probabilities from moment inequalities at any even order.

Lemma 4. *If the variables $\{V_n\}_{n \in \mathbb{Z}}$ satisfies, for all $k \in \mathbb{N}^*$*

$$\|V_n\|_{2k} \leq \phi(2k), \quad (23)$$

where ϕ is an increasing function with $\phi(0) = 0$. Then:

$$\mathbb{P}(|V_n| \geq \epsilon) \leq e^2 \exp(-\phi^{-1}(\epsilon/e)).$$

Proof. By Markov's inequality and Assumption (23), we obtain

$$\mathbb{P}(|V_n| \geq \epsilon) \leq \left(\frac{\phi(2k)}{\epsilon} \right)^{2k}.$$

With the convention $0^0 = 1$, the inequality is true for all $k \in \mathbb{N}$. Reminding that $\phi(0) = 0$, there exists an integer k_0 such that $\phi(2k_0) \leq \epsilon/e < \phi(2(k_0+1))$. Noting ϕ^{-1} the generalized inverse of ϕ , we have:

$$\begin{aligned} \mathbb{P}(|V_n| \geq \epsilon) &\leq \left(\frac{\phi(2k_0)}{\epsilon} \right)^{2k_0} \leq e^{-2k_0} = e^2 e^{-2(k_0+1)}, \\ &\leq e^2 \exp(-\phi^{-1}(\epsilon/e)). \end{aligned}$$

□

We rewrite the inequality (22): $\left\| \sqrt{\frac{n}{m_n}} FL_n \right\|_{2k} \leq \phi(2k)$ with $\phi(x) = Cx^{\frac{b+1}{b}}$ for a convenient constant C . Applying Lemma 4 to $V_n = \sqrt{\frac{n}{m_n}} FL_n$ we obtain:

$$\mathbb{P}\left(|FL_n| \geq \epsilon \sqrt{\frac{m_n}{n}}\right) \leq e^2 \exp(-\phi^{-1}(\epsilon/e)),$$

and we obtain the result of the Lemma 2.

- **Riemannian decay** In this case, the result of Lemma 1 is obtained only for some values of q depending of the value of the parameter a :
 - In the case of η -dependence:

$$a > \max\left(q-1, \frac{1+\delta+2/d}{1-\delta}, 2+\frac{1}{d}\right).$$

- In the case of $\tilde{\phi}$ -dependence:

$$a > \max\left(q-1, 1+\frac{2}{d(1-\delta)}, 1+\frac{1}{d}\right).$$

We consider that the assumptions of the Lemma 2 on a are satisfied in both cases of dependence. Then $q_0 = 2 \lceil \frac{a-1}{2} \rceil$ is the even integer such that $a-1 \leq q_0 < a+1$. It is the largest order such that the assumptions of Lemma 1 (recalled above) are verified and then the Lemma 1 gives us

directly the rate of the moment: $\limsup_{n \rightarrow \infty} \left(\frac{n}{m_n} \right)^{q_0/2} \|FL_n(x)\|_{q_0}^{q_0} < +\infty$.

We apply Markov to obtain the result of Lemma 2:

$$\mathbb{P}\left(|FL_n(x)| \geq \epsilon \sqrt{\frac{m_n}{n}}\right) \leq \frac{\left(\sqrt{\frac{n}{m_n}} \|FL_n(x)\|_{q_0} \right)^{q_0}}{\epsilon^{q_0}}.$$

□

Proof (Proof of Lemma 3). We follow here Liebscher's strategy as in [AD03]. We recover $B := B(0, M)$, the ball of center 0 and radius M , by at least $(4M\mu + 1)^d$ balls $B_j = B(x_j, 1/\mu)$. Then, under the assumption that $K_m(\cdot, y)$ is supported on a compact of diameter proportional smaller than $1/m^{1/d}$, we have, for all j :

$$\sup_{x \in B_j} |FL_n(x)| \leq |\hat{f}_n(x_j) - \mathbb{E}\hat{f}_n(x_j)| + C \frac{m_n^{1/d}}{\mu} (|\tilde{f}_n(x_j) - \mathbb{E}\tilde{f}_n(x_j)| + 2|\mathbb{E}\tilde{f}_n(x_j)|), \quad (24)$$

with C a constant and $\tilde{f}_n(x) = \frac{1}{n} \sum_{i=1}^n \tilde{K}_{m_n}(x, X_i)$ where \tilde{K}_{m_n} is a kernel of type $\tilde{K}_m(x, y) = K_0 m k(x, y, x_j, 1/m^{1/d})$. The $1/b$ -Lipschitz function $k(x, y, a, b)$ is equal to 1 on $B(a, b)$ and null outside $B(a, b + 1/b)$. The constant K_0 is fixed in order that \tilde{K}_{m_n} satisfies properties (a), (b) and (c) of section 2.2. Then using (24) and with obvious short notation:

$$\begin{aligned} \mathbb{P} \left(\sup_{\|x\| \leq M} |FL_n(x)| > \epsilon \sqrt{\frac{m_n}{n}} \right) &\leq \sum_{j=1}^{(4M\mu+1)^d} \mathbb{P} \left(\sup_{x \in B_j} |FL_n(x)| > \epsilon \sqrt{\frac{m_n}{n}} \right), \\ &\leq (4M\mu + 1)^d \left[\sup_{x \in B_j} \mathbb{P} \left(|FL_n(x_j)| > \epsilon \sqrt{\frac{m_n}{n}} \right) \right. \\ &\quad \left. + \mathbb{P} \left(C \frac{m_n^{1/d}}{\mu} |\tilde{FL}_n(x)| > \epsilon \sqrt{\frac{m_n}{n}} \right) \right. \\ &\quad \left. + \mathbb{P} \left(2C \frac{m_n^{1/d}}{\mu} |\mathbb{E}\tilde{f}_n(x_j)| > \epsilon \sqrt{\frac{m_n}{n}} \right) \right]. \end{aligned}$$

Using the fact that f is bounded, $\mathbb{E}\tilde{f}_n = \int \tilde{K}_{m_n}(x_j, s) f(s) ds$ is bounded independently of n . Then taking $\mu = m_n^{1/d-1/2} n^{1/2} L(n)/\epsilon$ ensures that $\mathbb{P} \left(2C \frac{m_n^{1/d}}{\mu} |\mathbb{E}\tilde{f}_n(x_j)| > \epsilon \sqrt{\frac{m_n}{n}} \right)$ is null for n sufficiently large. Applying Lemma 2 on f and \tilde{f} , uniform probability inequality in both cases of geometric and Riemannian decays become:

$$\mathbb{P} \left(\sup_{\|x\| \leq M} |FL_n(x)| \geq \epsilon_n \sqrt{\frac{m_n}{n}} \right) \leq \mu^d \exp \left(-C \epsilon_n^{\frac{b}{b+1}} \right), \quad (25)$$

$$\mathbb{P} \left(\sup_{\|x\| \leq M} |FL_n(x)| \geq \epsilon_n \sqrt{\frac{m_n}{n}} \right) \leq \mu^d \epsilon_n^{-q_0}. \quad (26)$$

In the geometric case, fix ϵ_n as $G(\log n)^{(b+1)/b}$ such that the bound becomes $\mu^d n^{-GC}$. Reminding that $\mu \leq n$, the sequence $\mu^d n^{-GC}$, bounded by n^{d-GC} , is summable for a conveniently chosen constant G . Borel-Cantelli's Lemma then concludes the proof in this case.

In the Riemannian case, take $\epsilon_n = (m_n^{1-d/2} n^{1+d/2})^{\frac{1}{q_0+d}} \log n$ such that the bound becomes $n^{-1} \log^{-q_0} nL(n)$. Reminding that $q_0 \geq 2$, this sequence is summable and here again we conclude by applying Borel-Cantelli's Lemma. \square

Proof of the theorems

The order of magnitude of the bias is given by Assumption (d) and the Lemmas provide bounds for fluctuation term. There only remain to determine the optimal bandwidth m_n in each case.

Proof (Proof of Theorem 1). Applying Lemma 1 yields Theorem 1 when q is an even integer. For any real q , Lemma 1 with $2(\lceil q/2 \rceil + 1) \geq 2$ and Jensen's inequalities yields:

$$\begin{aligned} \left(\frac{n}{m_n}\right)^{q/2} \mathbb{E}|FL_n(x)|^q &= \left(\frac{n}{m_n}\right)^{q/2} \mathbb{E} \left(FL_n(x)^{2(\lceil q/2 \rceil + 1)} \right)^{q/\{2(\lceil q/2 \rceil + 1)\}}, \\ &\leq \left(\left(\frac{n}{m_n}\right)^{\lceil q/2 \rceil + 1} \mathbb{E} FL_n(x)^{2(\lceil q/2 \rceil + 1)} \right)^{q/\{2(\lceil q/2 \rceil + 1)\}}. \end{aligned}$$

Plugging this bound and the bound for the bias in (12), we obtain a bound for the \mathbb{L}^q -error of estimation:

$$\|\hat{f}_n(x) - f(x)\|_q \leq \|FL_n(x)\|_q + |R_n(x)| = O\left(\sqrt{\frac{m_n}{n}} + m_n^{-\rho/d}\right).$$

The optimal bandwidth $m_n^* = n^{\frac{d}{2\rho+d}}$ is the same as in the i.i.d. case. Thus [H6] holds with $\delta = \frac{d}{2\rho+d}$. For this valued of δ , the conditions on the parameter a of Lemma 2 are equivalent to those of Theorem 1. \square

Proof (Proof of Theorem 2). Applying the probability inequality (25) in the proof of Lemma 3 and the identity $\mathbb{E}|Y|^q = \int_0^{+\infty} \mathbb{P}(|Y| \geq t^{1/q}) dt$, we obtain

$$\mathbb{E} \sup_{\|x\| \leq M} |\hat{f}_n(x) - f(x)|^q = O\left(\left\{\sqrt{\frac{m_n}{n}} \log^{(b+1)/b}(n)\right\}^q + m_n^{-q\rho/d}\right).$$

Lemma 3 gives the rate of almost sure convergence:

$$\sup_{\|x\| \leq M} |\hat{f}_n(x) - f(x)| =_{a.s.} O\left(\sqrt{\frac{m_n}{n}} \log^{\frac{b+1}{b}} n + m_n^{-\rho/d}\right).$$

In both cases, the optimal bandwidth is $m_n^* = (n/\log^{2(b+1)/b}(n))^{d/(2\rho+d)}$, which yields the rates claimed in Theorem 2. \square

Proof (Proof of Theorem 3). Applying the probability inequality (26) and the same line of reasoning as in the previous proof, we obtain

$$\mathbb{E} \sup_{\|x\| \leq M} |\hat{f}_n(x) - f(x)|^q = O \left(\left(\sqrt{\frac{m_n}{n}} m_n^{\frac{2}{q_0+d}} \right)^q + m_n^{-q\rho/d} \right),$$

where $q_0 = 2 \lceil (a-1)/2 \rceil$. The optimal bandwidth $m_n^* = n^{d/(d+2\rho+2d/(q_0+d))}$ implies [H6] with $\delta = d/(d+2\rho+2d/(q_0+d))$. For this value of δ , the conditions on a of Lemma 2 are satisfied as soon as $a \geq 4$ and $\rho > 2d$.

Lemma 3 gives the rate for the fluctuation in the almost sure case. This leads the optimal bandwidth

$$m_n^* = \left(n^{q_0-2} / \log^{q_0+d}(n) \right)^{\frac{d}{d(q_0+2)+\rho(q_0+d)}}.$$

We then deduce the two different rates of Theorem 3, either in the almost sure or in the \mathbb{L}^q framework. \square

References

- [AD98] P. ANGO NZE and P. DOUKHAN (1998), *Functional estimation for time series: uniform convergence properties*, Journal of Statistical Planning and Inference, vol. 68, pp. 5-29.
- [ABD02] P. ANGO NZE, P. BÜHLMANN and P. DOUKHAN (2002), *Weak dependence beyond mixing and asymptotics for nonparametric regression*, Annals of Statistics, vol. 30, n. 2, pp. 397-430.
- [AD03] P. ANGO NZE and P. DOUKHAN (2003), *Weak Dependence: Models and Applications to econometrics*, Econometric Theory, vol. 20, n. 6, pp. 995-1045.
- [BGR00] A.D. BARBOUR, R.M. GERRARD and G. REINERT (2000), *Iterates of expanding maps*, Probability Theory and Related Fields, vol. 116, pp. 151-180.
- [Ded98] J. DEDECKER (1998), *A central limit theorem for random fields*, Probability Theory and Related Fields, vol. 110, pp. 397-426.
- [DD03] J. DEDECKER and P. DOUKHAN (2003), *A new covariance inequality and applications*, Stochastic Processes and their Applications, vol. 106, n. 1, pp. 63-80.
- [DP04] J. DEDECKER and C. PRIEUR (2004), *New dependence coefficients. Examples and applications to statistics*, To appear in Probability Theory and Related Fields.
- [DDLLP04] J. DEDECKER, P. DOUKHAN, G. LANG, J.R. LEON, S. LOUHICHI and C. PRIEUR (2004), *Weak dependence: models, theory and applications*, Merida, XVII escuela venezolana de matematicas.
- [Dou88] P. DOUKHAN (1988), *Formes de Toeplitz associées à une analyse multi-échelle*, Compte rendus des Séances de l'Académie des Sciences, Série I. Mathématique. 306, vol. 84, n. 15, pp. 663-666.

- [Dou90] P. DOUKHAN (1991), *Consistency of delta-sequence estimates of a density or of a regression function for a weakly dependent stationary sequence*, Séminaire de statistique d'Orsay, Estimation Fonctionnelle 91-55.
- [Dou94] P. DOUKHAN (1994) *Mixing: properties and examples*, Lecture Notes in Statistics, vol. 85, Springer-Verlag.
- [DL99] P. DOUKHAN and S. LOUHICHI (1999), *A new weak dependence condition and applications to moment inequalities*, Stochastic Process and their Applications, vol. 84, pp. 313-342.
- [DL01] P. DOUKHAN and S. LOUHICHI (2001), *Functional estimation for weakly dependent stationary time series*, Scandinavian Journal of Statistics, vol. 28, n. 2, pp. 325-342.
- [DMR05] P. DOUKHAN H. MADRE and M. ROSENBAUM (2005). *ARCH type bilinear weakly dependent models*, submitted.
- [DS01] P. DOUKHAN and J.C. SIFRE (2001), *Cours d'analyse - Analyse réelle et intégration*, Dunod.
- [GKS96] L. GIRAITIS, H.L. KOUL and D. SURGAILIS (1996), *Asymptotic normality of regression estimators with long memory errors*, Statistics & Probability Letters, vol. 29, pp. 317-335.
- [HLT01] M. HALLIN, Z. LU, L.T. TRAN (2001), *Density estimation for spatial linear processes*, Bernoulli, pp. 657-668.
- [KN05] R. S. KALLABIS and M. H. NEUMANN (2005), *A Bernstein inequality under weak dependence*, prepublication.
- [Pri01] C. PRIEUR (2001), *Density Estimation For One-Dimensional Dynamical Systems*, ESAIM , Probability & Statistics, pp. 51-76.
- [Sze33] G. SZEGÖ (1933), *Orthogonal polynomials*, American Mathematical Society Colloquium Publication, vol. 23.
- [Tsy04] A.B. TSYBAKOV (2004), *Introduction à l'estimation non-paramétrique*, Springer.
- [Vie97] G. VIENNET (1997), *Inequalities for absolutely regular sequences : application to density estimation*, Probability Theory and Related Fields, vol. 107, pp. 467-492.