Convergence rates for density estimators of weakly dependent time series

Nicolas Ragache[1] and Olivier Wintenberger[2]

[1] MAP5, Université René Descartes 45 rue des Saints-Pères, 75270 PARIS
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

Assuming that $(X_n)_{n\in\mathbb{Z}}$ is a vector valued time series with a common marginal distribution admitting a density f, our aim is to provide a wide range of consistent estimators of f. We consider different methods of estimation as kernel, projection or wavelets ones. Various cases of weakly dependent series are investigated including the Doukhan & Louhichi's η-weak dependence condition (see [DL99]) and the $\tilde{\phi}$-dependence of Dedecker & Prieur (see [DP04]). We thus obtain results for Markov chains, dynamical systems, bilinear models, non causal Moving Average... ¿From a moment inequality of Doukhan & Louhichi, we provide convergence rates of the term of $\mathbb{L}^q$-error or for the uniform bounds on compact sets, in mean or almost surely.

1 Introduction

Assume that $(X_n)_{n\in\mathbb{Z}}$ is a $\mathbb{R}^d$ valued random process such that each marginal X_i has the same distribution with a density f with respect to (w.r.t. in the sequel) Lebesgue measure. Note that stationarity is not required so that case of a sampled process $\{X_{i,n} = x_{h_n(i)}\}_{1\le i\le n}$ for any sequence of monotonic functions $(h_n(.))_{n\in\mathbb{Z}}$ and any stationary process $(x_n)_{n\in\mathbb{Z}}$ that admits a marginal density is included. This paper gives convergence rates for density estimation in very different cases. First, we shall consider two different frames of weak dependence:

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

Note that these two frames of dependence are associated with a large number of examples of time series (see section § 3). Secondly, following Doukhan (see [DO90]) we propose an unified study of linear density estimators $\hat{f}_n$ of the form

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

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 [TS04]). At a fixed point $x \in \mathbb{R}^d$:

$$\|\hat{f}_n(x) - f(x)\|_q = \mathcal{O}\left(n^{-\frac{\rho}{2\rho+d}}\right) . \tag{2}$$

The coefficient $\rho > 0$ corresponds to the regularity of f (see section § 2.2 for the definition of the notion of regularity). The same convergence rates also hold for the Mean Integrated Square Error (MISE, which writes as $\int \|\hat{f}_n(x) - f(x)\|_2^2 p(x)dx$ for some nonnegative and integrable function p). Uniformly on a compact set, a logarithmic loss appears. For all $M > 0$:

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

and

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

Those rates are optimal in the minimax sense, we thus have no hope to improve on them in the dependent setting. Relaxing the condition of independence, the optimal previous decay rates are still achieved for mixing coefficients (see [DO94]). A vast literature deals with density estimation for absolutely regular (or β-)dependent processes X_n. For instance, under the assumption $\beta_r = o\left(r^{-3-2d/\rho}\right)$, Ango Nze & Doukhan prove in [AD98] that (2), (3) and (4) still hold. A sharper condition $\sum_r |\beta_r| < \infty$ entails the optimal decay for the MISE (see [VI97]). Results on 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 in the η- and $\tilde{\phi}$-weak dependence contexts the equations (2), (3) and (4) .

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 $n \ge 2$:

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

where $\|X\|_q^q = \mathbb{E}|X|^q$ and for k integer such that $2 \le k \le q$,

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

if $t_1 \le \cdots \le t_k$ are such that $t_{p+1} - t_p = \sup_{1 \le i \le k-1} t_{i+1} - t_i = r$, and

$$V_{k,n} = n \sum_{r=0}^{n-1} (r+1)^{k-2} C_k(r) \text{ for } 2 \le k \le q \; .$$

In our framework, Z_i is defined such 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 of this part of the error depending on the covariances $C_k(r)$. The other part of the error, the bias, is treated in a deterministic way. In order to obtain suitable controls of the fluctuation term, we need two different 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 couple (X_j, X_k). In Doukhan & Louhichi (see [DL01]), rates are obtained when the coefficient η decays geometrically and the joint densities are bounded. We relax those conditions for cases of degenerated joint distributions and with η and $\tilde{\phi}$ decreasing more slowly (sub-geometric and Riemannian decays are considered).

We prove that (2) still holds (see Theorem 1). Unfortunately, additional losses appear for the uniform bounds. For the decays η_r or $\tilde{\phi}_r = \mathcal{O}\left(e^{-ar^b}\right)$ with $a > 0$ and $b > 0$, we prove in Theorem 2 that (3) and (4) hold replacing the logarithmic factor $\log n$ by the worse one $\log^{2\frac{b+1}{b}}(n)$. If now either η_r or $\tilde{\phi} = \mathcal{O}(r^{-a})$ with $a > 1$, Theorem 3 gives bounds for (3) and (4) where the right hand side now write as $\mathcal{O}\left(n^{-q\rho q_0/(2\rho q_0 + d(q_0+2))}\right)$ and $\mathcal{O}\left(\left(\log^{2+4/(q_0-2)} n/n\right)^{\frac{\rho(q_0-2)}{2\rho q_0 + d(q_0+2)}}\right)$, for $q_0 = 2\left\lceil \frac{a-1}{2} \right\rceil$ (by definition $\lceil x \rceil = -[-x]$). As already noticed in [DL01], the loss w.r.t the i.i.d. case highly depends on the decay of the coefficients. In geometric cases, the loss is logarithmic while it is polynomial in the Riemannian cases.

The paper is organized as follows. In section § 2.1 we introduce the notions of η-dependence and $\tilde{\phi}$ dependence. We give the notation and hypothesis in § 2.2. The main results are presented at the section § 2.3. We then apply those results on special weak dependence processes in section § 3. Section § 4 contains the proofs of the theorems with three important lemmas.

2 Main results.

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

2.1 Weak dependences.

We consider a sequence $(X_i)_{i\in\mathbb{Z}}$ with values in $\mathbb{R}^d$, and we fix a norm $\|\cdot\|$ on $\mathbb{R}^d$. Moreover, if $h:\mathbb{R}^{du}\to\mathbb{R}$ for some $u\geq 1$, we set

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

The η-dependence.

Definition 1 (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 $(\eta_r)_{r\geq 0}$ verifying $\eta_r\to 0$ when $r\to\infty$ and such that:*

$$\left|Cov\left(h\left(X_{i_1},\ldots X_{i_u}\right),k\left(X_{i_{u+1}},\ldots,X_{i_{u+v}}\right)\right)\right|\leq(u\text{Lip}(h)+v\text{Lip}(k))\eta_r\ ,$$

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

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

Remark: The η-dependence refers to non-causal situations because information 'from the future' (i.e. on the right of the covariance) contributes as much as information 'from the past' (i.e. on the left) in the dependence scheme. It is the non-causal alternative to the θ condition in [DD03] and [DL99].

The $\tilde{\phi}$-dependence. This condition is a causal one:

Definition 2 (Dedecker & Prieur (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 $\Lambda_{1,l}=\{h:\mathbb{R}^{dl}\mapsto\mathbb{R}/Lip\,(h)<1\}$. The sequence of coefficients $\tilde{\phi}_k(r)$ is then defined by

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

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

Remarks. As in the previous case, we control the covariance of such processes. For a Lipschitz function k and for an integrable function h:

$$\left|\text{Cov}\left(h\left(X_{i_1},\ldots,X_{i_u}\right),k\left(X_{i_{u+1}},\ldots,X_{i_{u+v}}\right)\right)\right|\leq v\mathbb{E}\left|h\left(X_{i_1},\ldots,X_{i_u}\right)\right|\text{Lip}\,(k)\tilde{\phi}(r)\ . \qquad (7)$$

Furthermore, the coefficient $\tilde{\phi}(r)$ reaches the least upper bound in (7).

2.2 Notations and settings.

Assume that $(X_n)_{n\in\mathbb{Z}}$ is a η- or $\tilde{\phi}$-weakly dependent time series of $\mathbb{R}^d$. We consider two types of decays for the coefficients. The geometric cases is referred to as one of the assumptions [H1] or [H1']:

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

We refer to the Riemannian cases if assumption [H2] or [H2'] holds:

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

As it classically appears in density estimation, we shall assume:

[H3]: All X_n, $n \in \mathbb{Z}$ have a common marginal distribution admitting a bounded density f. Furthermore, f is ρ-regular (noted $f \in \mathcal{C}_\rho$): for $\rho = \lceil\rho\rceil + b$ (then $0 < b \leq 1$ by defintion), f is $\lceil\rho\rceil$-times continuously differentiable and $|f^{(\lceil\rho\rceil)}(x) - f^{(\lceil\rho\rceil)}(y)| \leq A|x-y|^b$ with $A \geq 0$, $\forall(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$.

One technical assumption relies on the joint distributions (when their exist):

[H4]: the joint distributions $f_{j,k}$ of the couples (X_j, X_k) are uniformly bounded for $j \neq k$.

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

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

where the ϵ_k follows a binomial law and X_0 is uniformly distributed on $[0,1]$. The process X_n is strictly stationary but the joint laws of couples (X_0, X_k) are degenerated. The Markov chain can also be represented (through an inversion of the time) as a dynamical system $(T_{-n}, \ldots, T_{-1}, T_0)$ which has the same law than $(X_0, X_1, \ldots, X_n)$ (T_0 and X_0 are random variables distributed according to the invariant measure, see [BG00] for more details). Let us recall the definition of a dynamical system:

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

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

where $F : I \to I$ for I a compact of $\mathbb{R}$. F is a transformation admitting an invariant probability measure μ_0, that we assume to be Lebesgue dominated, and T_0 is a random variable distributed according to μ_0.

We restrict ourselves to one-dimensional dynamical systems $T_k \in \mathcal{F}$ where the class $\mathcal{F}$ is defined under assumptions on the transformation F (see [PR01]):

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

According to this discussion, when the joint laws are degenerated (and then [H4] is not satisfied), we shall assume:

[H5]: $(X_n)_{n\in\mathbb{Z}}$ is a dynamical system belonging to $\mathcal{F}$.

We consider in this paper linear estimator as in (1). Kernel, projections and wavelets estimators are written in such a linear form in section § 4.2. The kernels K_{m_n} have to satisfy the assumptions:

(a) They are supported on a compact of diameter $\mathcal{O}(1/m_n)$,
(b) $x \mapsto K_{m_n}(x,y)$ and $x \mapsto K_{m_n}(y,x)$ for all y are Lipschitz functions with constant $\mathcal{O}\left(m_n^{1/d}\right)$,
(c) $\int K_{m_n}(x,y)dy = 1$,
(d) K_{m_n} is uniformly bounded in n.

2.3 Results.

We now present three Theorems followed by some remarks:

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

Geometric cases. *Under the assumptions [H4] or [H5] and [H1] or [H1'], then inequality (2) holds for all* $0 < q < +\infty$.

Riemannian cases. *Under the assumptions [H4] or [H5] if additionally*

- *η-dependence holds with [H2],* $a > \max\left(1 + \frac{2}{d} + \frac{d+1}{\rho}, 2 + \frac{1}{d}\right)$, *or*
- *$\tilde{\phi}$-dependence holds with [H2'] and* $a > 1 + \frac{2}{d} + \frac{1}{\rho}$,

then inequality (2) holds for all $0 < q \le q_0 = 2\left\lceil \frac{a-1}{2} \right\rceil$.

Theorem 2 (Uniform rates, geometric decays). *For any* $M > 0$ *under the assumptions [H4] or [H5] and [H1] or [H1'] we have for all* $0 < q < +\infty$*:*

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

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

Theorem 3 (Uniform rates, Riemannian decays). *For any $M > 0$ under the assumptions [H4] or [H5], with additionally [H2] or [H2'], $a \ge 4$ and $\rho > 2d$, then, for $q_0 = 2\left\lceil \frac{a-1}{2} \right\rceil$ and $q \le q_0$:*

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

$$\sup_{\|x\| \le M} |\hat{f}_n(x) - f(x)| =_{a.s.} \mathcal{O}\left(\left(\frac{\log^{2+4/(q_0-2)} n}{n} \right)^{\frac{\rho(q_0-2)}{2\rho q_0 + d(q_0+2)}} \right).$$

Remarks.

- In Theorem 1, the optimal convergence rate of equation (2) still hold in our weak dependence context. In the Riemannian case, when $a \ge 4$, the conditions are always satisfied by assuming a sufficient regularity on the density f, i.e. $\rho > d + 1$.
- Losses appearing when we consider uniform convergence rates (Theorem 2 & 3) are due to the fact that the probability inequalities we obtain for dependent observations are not as good as the Bernstein one (Bernstein inequalities in weak dependence context are proved in [KN05]). Unlike in the independent case, the convergence decays are not the same according to the decays of the weak dependence coefficients.
- In Theorem 2 the loss is a logarithmic power. Let us remark that this loss disappears when b tends to infinity, or equivalently when we tends to the independent case. In the case of η-dependence and geometric decreasing, the same result is in [DL99] for the special case $b = 1$. In the case of $\tilde{\phi}$-dependence, we have for the first time uniform rates of convergence for density estimators.
- In Theorem 3, better is the mean-rate than the almost sure rate for technical raisons. Instead in the geometric case, our loss is no longer logarithmic but a power of n. We obtain asymptotically the optimal rate when $q_0 \to \infty$, or equivalently $a \to \infty$.
- We investigate for the first time convergence rates for Riemannian decay of those weak dependence coefficients. The condition we obtain on a is similar to the condition on β in [AD03]. Even if our rates are better than in [DL01], we have a huge loss comparatively with the mixing case. It would be interesting to have minimax results in this framework to know if we can achieve better convergence rates.

3 Models, applications and extensions.

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

3.1 Examples η-dependent time series.

η-dependent random fields may also be defined (see [DD04] for further details), for simplicity, here the index set is always $\mathbb{Z}$.

Definition 4 (Bernoulli shifts). *Let $H : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ 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 variables called the innovation process.*

Thanks to the following regularity condition on H:

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

the real sequence $\{\delta_r\}_{r\in\mathbb{N}}$ gives an expression of the η_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.** For a real sequence $(a_i)_{i\in\mathbb{Z}}$, $X_n = \sum_{-\infty}^{+\infty} a_i\xi_{n-i}$ is a real ($d = 1$) **non-causal linear process**. 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 restrict to Riemannian decays $a_i = \mathcal{O}\left(i^{-A}\right)$ with $A \ge 5$ since geometric decays yield already known results. 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. satisfying the condition $|\mathbb{E}e^{iu\xi_0}| \le C(1+|u|)^{-\delta}$, for all $u \in \mathbb{R}$ and for some $\delta > 0$ and $C < \infty$. Then, both densities f and $f_{j,k}$ exist for $j \ne k$ and they are uniformly bounded (see the proof in the causal case in lemma 1 and lemma 2 in [GK96]) ; 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:
   - $n^{-\frac{\rho}{2\rho+1}}$ in $\mathbb{L}^q$-norm ($q \le 4$) at each point x,
   - $n^{-\frac{\rho}{2\rho+3/2}}$ in $\mathbb{L}^q$-norm ($q \le 4$) uniformly on an interval,
   - $\left(\log^4 n/n\right)^{\frac{\rho}{4\rho+3}}$ 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 Central Limit Theorems (see [HL01] and [DE98]).

2. **Bilinear model.** The process X_t follows a **bilinear model** if and only if there exist two sequences $(a_i)_{i\in\mathbb{N}^*}$ and $(b_i)_{i\in\mathbb{N}^*}$ of real numbers, and a and b two non-negative numbers such that:
$$X_t = \xi_t \left(a + \sum_{j=1}^{\infty} a_j X_{t-j} \right) + b + \sum_{j=1}^{\infty} b_j X_{t-j} \ . \tag{10}$$
Real ARCH(∞) and **GARCH(p,q)** are the most famous **bilinear models**. Assuming $\lambda = \|\xi_0\|_{L^p}\|a\|_1 + \|b\|_1 < 1$, then the equation (10) has a strictly stationary solution in L^p (see [DM05]). This solution is a special **Bernoulli shift** for which we have the behavior of the coefficient η:
   - $\eta_r = \mathcal{O}\left(e^{-\lambda r}\right)$, $\lambda > 0$ when there exists an integer N such that $a_i = b_i = 0$ for $i \geq N$.
   - $\eta_r = \mathcal{O}\left((r/\log r)^{-\lambda}\right)$, $\lambda > 0$ under Riemannian decay, i.e. we have $a_i = \mathcal{O}(i^{-A})$ and $b_i = \mathcal{O}(i^{-B})$ with $A > 1$ and $B > 1$.
   - $\eta_r = \mathcal{O}\left(e^{-\lambda\sqrt{r}}\right)$ under geometric decay, i.e. we have $A > 0$ and $B > 0$ such that $a_i = \mathcal{O}(e^{-Ai})$ and $b_i = \mathcal{O}(e^{-Bi})$.

   Let us furthermore assume that the independent innovations ξ_t have the same marginal density $f_\xi \in C^\rho$, for $\rho > 2$. The density of X_t conditionally to the past write simply as a function of f_ξ. We then verify recursively that the common density of X_t for all t, f, is also C^ρ. Furthermore, the regularity on ξ ensures that f and the joint densities $f_{j,k}$ for all $j \neq k$ are bounded (see [DM05]) and [H4] holds. Theorem 1 implies the minimax bound (2) if either:
   - There exists an integer N such that $a_i = b_i = 0$ for $i \geq N$ (we are in fact in the case of **ARCH(p)** models).
   - We have A and B such that $a_i = \mathcal{O}(e^{-Ai})$ and $b_i = \mathcal{O}(e^{-Bi})$ with $A > 0$ and $B > 0$ (which includes **GARCH(p,q)** models and certain **ARCH(∞)** models).
   - We have $A \geq 4$ and $B \geq 5$ such that $a_i = \mathcal{O}(i^{-A})$ and $b_i = \mathcal{O}(i^{-B})$. Then, this optimal bound holds only for $2 \leq q < q(A,B)$ where $q(A,B) = 2[((B-1)\wedge A)/2]$.

   Furthermore, the uniform convergence is ensured in each cases by Theorems 2 and 3 but with sub-optimal convergence rates.

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 $0 = a_0 \leq a_1 \cdots \leq a_n = 1$ such as $F \in \mathcal{C}_1$ and $|F'(x)| > 0$ on $]a_{i-1}, a_i[$ for each $i = 1, \ldots, n$.

- (expansivity) Let I_n be the set on which $(F^n)'$ is defined. There exists $a > 0$ and $s > 1$ such as $\inf_{x \in I_n} |(F^n)'| > as^n$.
- (Topological mixing) For any nonempty open sets U, V, there exists $n_0 \geq 1$ such as $F^{-n}(U) \cap V \neq \varnothing$ 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 [BG00] 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 $\mathcal{F}$.

Let us consider one such **Markov chain**. We then control the coefficient $\tilde{\phi}$ of the associated **Markov chain**: $\tilde{\phi}(r) = O(e^{-ar})$ where $a > 0$ (see [DP04]). Theorems 1 and 2 give us the $\mathbb{L}^q$ rate $n^{-\frac{\rho}{2\rho+1}}$, the uniform $\mathbb{L}^q$ rate and almost sure rate $\left(\frac{\log^4 n}{n}\right)^{\frac{\rho}{2\rho+1}}$ of the approximation of the density of μ_0. This density estimator also satisfies a Central Limit Theorem is given in [PR01].

3.3 Extensions.

Let us recall that the stationarity is not required here so that 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.

We have seen that results in geometric cases are better than in Riemannian cases. Choosing $h_n(i)$ equals to $i2^n$ is then a way to transform the Riemannian weak dependent process $(x_n)_{n \in \mathbb{Z}}$ into a geometric one. Then, if we consider the linear estimator $\tilde{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_{m_n}(x, X_{i,n})$, we have constructed an estimator with at worst a logarithmic loss.

We may consider other sequences of functions $(h_n(\cdot))_{n \in \mathbb{Z}}$ in order to increase the rate of the decay of the weak dependence coefficients. In the geometric cases [H1] or [H1'] for $(x_n)_{n \in \mathbb{Z}}$, with dependent coefficients $= \mathcal{O}\left(e^{-ar^b}\right)$, $(X_{i,n})$ with $h_n(i) = i^c$ has dependent coefficients $= \mathcal{O}\left(e^{-ar^{cb}}\right)$. When c is large, we approach the independent case where weak dependence coefficients vanish. The estimator $\tilde{f}_n$ has a uniform convergence rate $= \mathcal{O}\left(\left(\log^{2\frac{cb+1}{cb}}(n)/n\right)^{\rho/(d+2\rho)}\right)$ better than the one of $\hat{f}_n$. Unfortunately, the minimax rate is never obtained, even asymptotically ($c \to +\infty$), and we have to observe the process $(x_n)_{n \in \mathbb{Z}}$ during a long period in order to apply those methods.

4 Proofs.

We observe the process $(X_n)_{n\in\mathbb{Z}}$ with $X_i \in \mathbb{R}^d$. The sequence m_n is called the window parameter. We assume that $1/m_n + m_n/n \to 0$ as $n \to \infty$. The asymptotic behavior of density estimators (1) is based on the decomposition:

$$\hat{f}_n(x) - f(x) = \underbrace{\hat{f}_n(x) - \mathbb{E}\left(\hat{f}_n(x)\right)}_{FL_n(x)=\text{fluctuation}} + \underbrace{\mathbb{E}\left(\hat{f}_n(x)\right) - f(x)}_{R_n(x)=\text{bias}} . \tag{11}$$

In the following section § 4.1, we carefully choose the kernels K_{m_n} such that assumptions (a), (b), (c) and (d) of subsection § 2.2 are satisfied. We will see in that framework that we can also ensure the classical rate $m_n^{-\rho/d}$ for the bias $R_n(x)$. This work on bias is totally disconnected from the dependence properties of the observations and only depends on the technique of estimation. In subsection § 4.2, we derive from three lemmas the decay rate for the fluctuation in those weak dependence context. Finally, in subsection § 4.3, we choose the optimal window parameter m_n^*, the one that equalize bias and variance.

4.1 Density estimates and bias.

Kernel density estimation. The kernel estimators associated to the window 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 [TS04]). Using only the equality of the marginal law, 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 on $D \subset \mathbb{R}^d$. Then, if $f \in \mathcal{C}_\rho$ for $\rho > 0$, one can always choose a kernel function K of order $\lceil\rho\rceil$, i.e. for all $j = j_1 + \cdots + j_d$ with $(j_1, \ldots, j_d) \in \mathbb{N}^d$:

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

Then the kernels $K_{m_n}(x,y) = m_n K\left(m_n^{1/d}(x-y)\right)$ satisfy (a), (b), (c) and (d) and we ensure that $R_n(x) = \mathcal{O}\left(m_n^{-\rho/d}\right)$.

Projection estimation. Let us restrict in this section to the case $d = 1$. Under the assumption that the family $\{1, x, x^2, \ldots\}$ belongs to $L^2(I,\mu)$ (where

I is a bounded interval of $\mathbb{R}$ and μ is a measure on I), the orthonormal polynomials family $\{P_0, P_1, P_2, \dots\}$ is a basis of $L^2(I,\mu)$. Then, the fact that I is compact and the Christoffel-Darboux formula and its corollary (see [SZ33]) ensures properties (a), (b) and (d) for the elements of the basis. We assume a ρ-regularity assumption on f denoted $\mathcal{C}'_\rho$ (a bit more restrictive than $\mathcal{C}_\rho$ introduced in [H3], see Theorem 6.23 p.218 in [DS01] for details). Then for any $f \in L^2(I,\mu) \cap \mathcal{C}'_\rho$ it always exists a function $\pi_{f,m_n} \in V_{m_n}$ verifying $\sup_{x\in I} |f(x) - \pi_{f,m_n}(x)| = \mathcal{O}(m_n^{-\rho})$ (the optimal rate). Now consider the projection $\pi_{m_n} f$ of f on the subspace V_{m_n} =Vect$\{P_0, P_1, \dots, P_{m_n}\}$. The classical decomposition holds $\pi_{m_n} f(x) = \sum_{j=0}^{m_n} \int_I P_j(s) f(s) d\mu(s) P_j(x)$. A projection estimator of the common density f of the real variables $\{X_i\}_{1\le i\le n}$, naturally arises:

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

Then $\mathbb{E}\hat{f}_n(x) = \pi_{m_n} f(x)$ is an approximation of $f(x)$ belonging to V_{m_n}. We easily check that properties (a), (b), (c) and (d) hold for the kernels K_{m_n}. 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 $\{a_{m,j};\, m \in \mathbb{N},\, 0 \le j \le m\}$ is a weight sequence satisfying $\sum_{j=0}^m a_{m,j} = 1$ and for all j: $a_{m,j} \to_{m\to\infty} 0$. If the sequence $\{a_{m,j}\}$ is such that K_m^a is a nonnegative kernel then $\|K_m^a\|_1 = \int_I K_m^a(x,s) d\mu(s) = 1$. Remarking that $\mathbb{E}\hat{f}_n(x) = \int_I K_{m_n}^a(x,s) f(s) d\mu(s) = K_{m_n}^a * f(x)$ is the functional operator $f \mapsto K_m^a * f(x)$ with uniform norm $\sup_{\|f\|_\infty = 1} \|K_m^a * f\|_\infty = \|K_m^a\|_1 = 1$. The associated linear estimator:

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

satisfies (a), (b), (c), (d) and its bias has the optimal rate:

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

Example 2 (Fejer's 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 $\sup_{x\in[-\pi;\pi]} |f(x) - \pi_m f(x)| = \mathcal{O}(m^{-1}\log m)$. The associated estimator writes:

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

We remark that $\mathbb{E}\hat{f}_n$ is the classical 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 $D_m(x) = \sum_{k=-m}^{m} e^{ikx} = \frac{\sin \frac{2m+1}{2} x}{\sin \frac{1}{2} x}$ is the Dirichlet's kernel. We derive from this kernel the nonnegative Fejer's kernel:

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

The kernels F_m is the weighted kernel $K_m^a(x,y)$ corresponding to the sequence $a_{m,j} = 1/m$. Then, if the common density f of $\{X_i\}_{1 \le i \le n}$ belongs in $\mathcal{C}_1'$ and is 2π-periodic, the associated estimator of the Fejer's kernels:

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

satisfies (a), (b), (c) and (d) and reaches the optimal rate m_n^{-1} for its bias.

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

Wavelets estimation. Wavelets estimation is an important case of projection estimation. For the sake of simplicity, we restrict to $d = 1$.

Definition 5 (Scaling function (see [DO88])). *A function $\phi \in L^2(\mathbb{R})$ is called scaling function if the family $\{\phi(x-k)\, ;\, k \in \mathbb{Z}\}$ is orthonormal.*

We traditionally choose the window parameter $m_n = 2^{j(n)}$ and we defined the $V_j \subset L^2(\mathbb{R})$ as $\mathrm{Vect}\{\phi_{j,k}, k \in \mathbb{Z}\}$, where $\phi_{j,k} = 2^{j/2}\phi(2^j(x-k))$. Under the assumption that ϕ is compactly supported, we define (the sum on k is in fact finite):

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

We check that the wavelets estimator could be seen as a linear one of the form of (1) where $K(x,y) = \sum_{k=-\infty}^{\infty} \phi(y-k)\phi(x-k)$ and $K_m(x,y) = mK(mx,my)$.

Under the assumption of concentration $\sum_{k\in\mathbb{Z}}\phi(x-k)=1$ for almost all x, we can write:

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

One considers now that the scaling function ϕ is a Lipschitz function with regularity $\lceil\rho\rceil$, i.e., $\int\phi(x)x^j dx = 0$ if $0<j<\lceil\rho\rceil$ and $\int\phi(x)x^{\lceil\rho\rceil}dx\neq 0$. Then the taylor expansion of f leads to determine the order of the bias $R_n(x)=\mathcal{O}(m_n^{-\rho})$. Furthermore, we easily check that the kernel K_m also satisfy properties (a), (b), (c) and (d).

4.2 Fluctuations.

In the previous section, we have clearly defined our estimator $\hat{f}_n(x)$ and then fixed the rate of decay for the bias. The error term depends now only on the fluctuation by the equation (11). In order to precise the decay of the window parameter, we define a new assumption:

> [H6]: $m_n = m(n)$ with $m\in RV(\delta)$ for $0<\delta<1$, where $RV(\delta)$ is the set of regularly varying function. By definition, a continuous function $u:\mathbb{R}^+\mapsto\mathbb{R}^+$ is regularly varying at the order δ if there exists $\beta\ge 0$ such that $u(x)=x^\delta/\log^\beta(x)$ for all $x>0$.

Lemmas.

We now present three lemmas useful to derive the rate of the fluctuation term:

Lemma 1 (Moment's inequalities). *For each even integer q, under the assumption [H4] or [H5] and:*

- *in the geometric cases we furthermore assume either [H1] or [H1'],*
- *in the Riemannian cases we furthermore assume [H6] and:*
  - *under η-dependence, [H2] with*

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

  - *under $\tilde{\phi}$-dependence, [H2'] with*

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

then we have at each point $x \in \mathbb{R}^d$:

$$\lim_{n\to\infty} \sup \left(\frac{n}{m_n}\right)^{q/2} \|FL_n(x)\|_q^q < +\infty \, .$$

Lemma 2 (Probability inequalities).

- *Geometric cases. Under the assumptions [H4] or [H5] and [H1] or [H1'] there exists a constant C such that:*

$$\mathbb{P}\left(|FL_n(x)| \geq \epsilon\sqrt{\frac{m_n}{n}}\right) \preceq \exp\left(-C\epsilon^{\frac{b}{b+1}}\right) \, .$$

- *Riemannian cases. Under the assumptions [H4] or [H5], [H6], with additionally*
  - *η-dependence, [H2] and* $a > \max\left(1 + 2\frac{\delta + 1/d}{1-\delta}, 2 + \frac{1}{d}\right)$,
  - *$\tilde{\phi}$-dependence, [H2'] and* $a > \max\left(1 + 2\frac{1/d}{1-\delta}, 1 + \frac{1}{d}\right)$,

  for $q_0 = 2\left\lceil \frac{a-1}{2} \right\rceil$, we have:

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

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

- *Geometric cases.* $\sup_{\|x\|\leq M} |FL_n(x)| =_{a.s.} \mathcal{O}\left(\sqrt{\frac{m_n}{n}} \log^{\frac{b+1}{b}} n\right)$.
- *Riemannian cases. With $q_0 = 2\left\lceil \frac{a-1}{2} \right\rceil$:*

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

Remarks.

- In lemma 1, we improve the moments 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 lemma 2, we give exponential inequalities similar to the Bernstein inequality available in the i.i.d. case:

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

  where C is a constant. Unfortunately, for dependent sequences, we do not achieve such good rates. Other probability inequalities in dependence framework are presented in [DP04] and [KN05].
- The lemma 3 gives the almost sure bounds for the fluctuation. It derives directly from the two precedent lemmas.

Proofs of the lemmas.

Proof of lemma 1. Let x be a fixed point in $\mathbb{R}^d$. We note $Z_i = u_n(X_i) - \mathbb{E}u_n(X_i)$ where $u_n(.) = K_{m_n}(., x)/\sqrt{m_n}$, then we have the identities:

$$\sum_{i=1}^{n} Z_i = \sum_{i=1}^{n} u_n(X_i) - \mathbb{E}u_n(X_i) = \frac{n}{\sqrt{m_n}}(\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)) = \frac{n}{\sqrt{m_n}} FL_n(x) \ . \tag{12}$$

We derive the rate of the fluctuation of the estimator $\hat{f}_n(x)$ applying the inequality (5) at the centered sequence $\{Z_i\}_{1\le i\le n}$ defined above. We then control the normalized fluctuation of (12) with the covariance terms $C_k(r)$ defined in equation (6). Firstly, we bound those covariance terms:

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

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

  By definition of Z_i:

  $$\mathbb{E}|Z_i|^k \le 2^k \mathbb{E}|u_n(X_i)|^k \le 2^k \|u_n\|_\infty^{k-1} \mathbb{E}|u_n(X_0)| \ . \tag{13}$$

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

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

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

      $$\begin{aligned} C_k(r) &\le \left(p2^p\|u_n\|_\infty^{p-1} + (k-p)2^{p-k}\|u_n\|_\infty^{p-k-1}\right)\text{Lip}\, u_n \eta_r \ , \\ &\le k2^k \|u_n\|_\infty^{k-1} \text{Lip}\, u_n \eta_r \ . \end{aligned} \tag{14}$$

    - *$\tilde{\phi}$-dependence:* We use the inequality (7). Using the bound

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

      we derive a bound for the covariance terms:

      $$C_k(r) \le k2^k \|u_n\|_\infty^{k-2} \mathbb{E}|u_n(X_0)| \text{Lip}\, u_n \tilde{\phi}(r) \ . \tag{15}$$

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

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

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

Firstly, with $k \geq 2$:

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

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

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

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

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

Then, collecting the last four inequations yields:

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

So:

$$A \le \|u_n\|_\infty^{k-2}\left((2^k - 1)(\mathbb{E}|u_n(X_0)|)^2 + \mathbb{E}|u_n(X_0)u_n(X_r)|\right) \ . \qquad (16)$$

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

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

Then:

$$B_p \times B_{k-p} \le 2^k\|u_n\|_\infty^{k-4}(\mathbb{E}|u_n(X_0)|)^4 \le 2^k\|u_n\|_\infty^{k-2}(\mathbb{E}|u_n(X_0)|)^2 \ . \qquad (17)$$

Follow another interesting bound for $r > 0$, because according to inequalities (16) and (17) we have:

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

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

$$C_k(r) \le 2^{k+1}\|u_n\|_\infty^{k-2}\gamma_n(r) \ . \qquad (18)$$

We now use the different values of the bounds in inequalities (13), (14), (15) and (18). If we define the sequence $(w_r)_{0\le r\le n-1}$ as:

- $w_0 = 1$,
- $w_r = \gamma_n(r) \wedge \|u_n\|_\infty \mathrm{Lip}\, u_n \eta_r \wedge \mathbb{E}|u_n(X_0)| \mathrm{Lip}\, u_n \tilde{\phi}(r)$,

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

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

We derive from this inequality and from (5):

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

The symbol $\preceq$ means $\le$ up to an universal constant. 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:

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

- In the case of [H5], the lemma 2.3 of [PR01] proves that $\mathbb{E}|u_n(X_0)u_n(X_r)| \le (\mathbb{E}|u_n(X_0)|)^2$ 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), (c) and (d) of section 2.2 ensures that $\|u_n\|_1^2 \preceq \dfrac{1}{m_n}$, $\|u_n\|_\infty \mathrm{Lip}\, u_n \preceq m_n^{1+1/d}$ and $\mathbb{E}|u_n(X_0)| \mathrm{Lip}\, u_n \preceq m_n^{1/d}$. We then have for $r \ge 1$:

$$w_r \preceq \frac{1}{m_n} \wedge m_n^{1+1/d}\eta_r \wedge m_n^{1/d}\tilde{\phi}_r \ . \tag{19}$$

In order to obtain the lemma 1, it remains to control the sums

$$\left(\frac{\|u_n\|_\infty}{\sqrt{n}}\right)^{k-2}\sum_{r=0}^{n-1}(r+1)^{k-2}w_r \ , \tag{20}$$

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

- **Geometric cases.**
  Under [H1] or [H1']: We remark that $a \wedge b \le a^\alpha b^{1-\alpha}$ for all $\alpha \in [0;1]$. Using (19), we obtain first that $w_r \preceq (\eta_r \wedge \tilde{\phi}_r)^\alpha m_n^{\alpha(1+1/d)-(1-\alpha)}$ for n

sufficiently large. Then for $0 < \alpha \le \frac{d}{2d+1}$ we bound w_r independently of m_n: $w_r \preceq (\eta_r \wedge \tilde{\phi}_r)^\alpha$. For all even integer $k \ge 2$ we derive from the form of $\eta_r \wedge \tilde{\phi}_r$ that (in the third inequality $u = ar^b$):

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

Using the Stirling formula, we can find a constant B such that, for the both 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 cases.**
  Under [H6] and [H2]: Let us recall that [H6] implies that $m_n \le n^\delta$ 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 $a > \max\left(k-1, \dfrac{\delta(k-1)(4+2/d)}{k-2+\delta(4-k)}\right)$ 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 $0 < \delta < 1$:

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

We derive from the two previous inequalities that there exists $\zeta_k \in]0,1[$ verifying $\dfrac{(k+2/d)\delta+2-k}{2(a-k+1)} < \zeta_k < \dfrac{(4-k)\delta+k-2}{2(k-1)}$.

For $k = q$ or $k = 2$, we now use the Tran's technique as in [AB02]. We divide the sum (20) in two parts in order to bound it by sequences tending to 0, due to the choice of ζ_k:

$$\left(\sqrt{\frac{m_n}{n}}\right)^{k-2} \sum_{r=0}^{[n^{\zeta_k}]-1} (r+1)^{k-2} w_r \preceq \left(\sqrt{\frac{m_n}{n}}\right)^{k-2} \frac{[n^{\zeta_k}]^{k-1}}{m_n} ,$$
$$\preceq n^{(2\zeta_k(k-1)-((4-k)\delta+k-2))/2} ,$$
$$= \mathcal{O}(1) ,$$
$$\left(\sqrt{\frac{m_n}{n}}\right)^{k-2} \sum_{r=[n^{\zeta_k}]}^{n-1} (r+1)^{k-2} w_r \leq \left(\sqrt{\frac{m_n}{n}}\right)^{k-2} m_n^{1+1/d} [n^{\zeta_k}]^{k-1-a} ,$$
$$\leq n^{(-2\zeta_k(a-k-1)+((k+2/d)\delta+2-k))/2} ,$$
$$= \mathcal{O}(1) .$$

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 (20) in that case.

The results of the lemma 1 directly follow from the equalities (12).□

Remarks. We have in fact proved the following sharper result. There exists a universally constant C such that:

$$\left(\frac{n}{m_n}\right)^{q/2} \|FL_n(x)\|_q^q \leq (Cq)^q \qquad \text{in the Riemaniann cases,}$$
$$\leq (Cq^{1+1/b}\sqrt{n})^q \qquad \text{in the geometric cases.} \tag{21}$$

Proof of lemma 2. The two cases of Riemannian or geometric decays of the weak dependence coefficients are considered separately.

- **Geometric decays.** 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 $k \in \mathbb{N}^*$*

$$\|V_n\|_{2k} \leq \phi(2k) , \tag{22}$$

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

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

Proof: By Markov at the order $2k$ and using the assumption (22):

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

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 general inverse of ϕ, we have:

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

We rewrite the inequality (21): $\left\|\sqrt{\frac{n}{m_n}} FL_n\right\|_{2k} \leq \phi(2k)$ with $\phi(x) = Cx^{\frac{b+1}{b}}$ for a convenient constant C. Applying the lemma 4 to $V_n = \sqrt{\frac{n}{m_n}} FL_n$ we obtain:

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

and we obtain the result of the lemma 2.

- **Riemannian decays.** 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 $q_0 = 2\left\lceil\frac{a-1}{2}\right\rceil$ is the even integer such that $a-1 < q_0 \leq a+1$. 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: $\lim_{n\to\infty}\sup\left(\frac{n}{m_n}\right)^{q_0/2} \|FL_n(x)\|_{q_0}^{q_0} < +\infty$. We apply Markov to obtain the result of lemma 2:

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

Proof of lemma 3. We follow here the 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 $K_m(., y)$ is supported on a compact of diameter proportional to $1/m$, we have, for all j:

$$\sup_{x\in B_j} |FL_n(x)| \leq |\hat{f}(x_j) - \mathbb{E}\hat{f}(x_j)| + C\frac{{m_n}^{1/d}}{\mu}(|\tilde{f}(x_j) - \mathbb{E}\tilde{f}(x_j)| + 2|\mathbb{E}\tilde{f}(x_j)|) \, , \tag{23}$$

with C a constant and $\tilde{f} = \frac{m_n}{n}\sum_{i=1}^n \tilde{K}_{m_n}(x, X_i)$ where $\tilde{K}_{m_n}$ is a kernel of type $\tilde{K}_{m_n}(x,y) = K_{m_n}(x_j, y)\mathbb{1}_{|x-x_j|\leq \frac{1}{m_n}}$ satisfying properties (a), (b), (c) and (d) of section 2.2. Then using (23) and with obvious short notation:

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

Using the fact that f is bounded, $\mathbb{E}\tilde{f}_n = \int \tilde{K}_{m_n}(s)f(x-hs)ds$ is bounded independently of n. We deduce that the last probability term of the sum tends to 0 with n.

With the choice $\mu = m_n^{1/d}$, we apply the lemma 2 on f and $\tilde{f}$. We have then the same rate (uniform in x) for both terms $\mathbb{P}\left(C\frac{m_n^{1/d}}{\mu}|\tilde{F}L_n(x)| > \epsilon\sqrt{\frac{m_n}{n}}\right)$ and $\mathbb{P}\left(|FL_n(x)| > \epsilon\sqrt{\frac{m_n}{n}}\right)$. Then replacing μ^d by m_n, we obtain uniform probability inequalities in both cases of geometric or Riemannian decays:

$$\mathbb{P}\left(\sup_{\|x\|\leq M}|FL_n(x)| \geq \epsilon_n\sqrt{\frac{m_n}{n}}\right) \preceq m_n \exp\left(-C\epsilon_n^{\frac{b}{b+1}}\right) \, , \tag{24}$$

$$\mathbb{P}\left(\sup_{\|x\|\leq M}|FL_n(x)| \geq \epsilon_n\sqrt{\frac{m_n}{n}}\right) \preceq m_n\epsilon_n^{-q_0} \, . \tag{25}$$

In the geometric case, we take ϵ_n with the form $G(\log n)^{\frac{b+1}{b}}$ such that the bound in the inequality (24) becomes $m_n n^{-GC}$. Reminding that $m_n \leq n$, the sequence $m_n n^{-GC}$ bounded by n^{1-GC} is integrable in n for a conveniently chosen constant G and the lemma of Borel-Cantelli leads to the result of the lemma 3.

In the Riemannian case, we take $\epsilon_n = (m_n n)^{\frac{1}{q_0}} \log n$ such that the bound in the inequality (25) becomes $n^{-1} \log^{-q_0} n$. Reminding that $q_0 \geq 2$, this sequence is integrable in n and the lemma of Borel-Cantelli leads to the result of the lemma 3.□

4.3 Proof of the theorems.

The rate of the bias is given in the section 4.1 and the rate of the fluctuation in the section 4.2. We now determine the optimal window parameter m_n in each cases.

Proof of Theorem 1. Lemma 1 bounds the normalized fluctuation. The result is optimal (i.e. we obtain the same normalization as in the independent case) when q is an even integer. We extend directly the results in the cases where q is real using Jensen's inequalities and the result of lemma 1 at the order $2(\lceil \frac{q}{2} \rceil + 1) \geq 2$:

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

With the control of the bias, we arise a natural bound for the $\mathbb{L}^q$-error of estimation:

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

We now have to equalize the two terms of the sum. The optimal window $m_n^* = n^{\frac{d}{2\rho+d}}$ is the usual one (i.e. the same as for the i.i.d. case). We then assume that [H6] holds with $\delta = \frac{d}{2\rho+d}$. We rewrite with this value of δ the conditions on the parameter a of the lemma 2 and then we obtain theorem 1.□

Proof of Theorem 2. With the probability inequality (24) of the proof of the lemma 3 (see section 4.2) and using $\mathbb{E}|Y|^q = \int_0^{+\infty} \mathbb{P}\left(|Y| \geq t^{1/q}\right) dt$, we are able to obtain the rate of the uniform bounds mean:

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

Lemma 3 gives exactly the same rate almost surely:

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

In both cases, the optimal window parameter is $m_n^* = \left(n/\log^{2\frac{b+1}{b}} n\right)^{\frac{d}{2\rho+d}}$. It leads to the rate of the error of estimation of Theorem 2.□

Proof of Theorem 3. For the mean of the uniform bounds, we use again a probability inequality (inequality (25) of the section 4.2) and the classical identity linking moment and probability (see the proof of Theorem 2). We obtain:

$$\mathbb{E} \sup_{\|x\|\leq M} |\hat{f}_n(x) - f(x)|^q = \mathcal{O}\left(\left(\sqrt{\frac{m_n}{n}} m_n^{1/q_0}\right)^q + m_n^{-q\rho/d}\right) ,$$

where $q_0 = 2\left\lceil \frac{a-1}{2} \right\rceil$. The optimal window parameter $m_n^* = n^{dq_0/(d(q_0+2)+2\rho q_0)}$ implies [H6] with $\delta = dq_0/(2\rho q_0 + d(q_0+2))$. For this value of δ, the conditions on a of lemma 2 are satisfied as soon as $a \geq 4$ and $\rho > 2d$.

The lemma 3 gives another rate for the fluctuation in the almost sure case. This leads to another optimal window $m_n^* = \left(n/\log^{2+4/(q_0-2)} n\right)^{\frac{d(q_0-2)}{2\rho q_0 + d(q_0+2)}}$. We then deduce the two different rates of Theorem 3 either we are in the almost sure or in the $\mathbb{L}^q$ framework.□

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.

[AB02] 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.

[BG00] A.D. Barbour, R.M. Gerrard and G. Reinert (2000), *Iterates of expanding maps*, Probability Theory and Related Fields, vol. 116, pp. 151-180.

[DE98] 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.

[DD04] 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.

[DO88] P. Doukhan (1988), *Formes de Toeplitz associées à une analyse multi-échelle*, note au CRAS, vol. 306, série 1, p. 663-666.

[DO90] 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.

[DO94] 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.

[DM05] 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.

[GK96] 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.

[HL01] 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.

[PR01] C. Prieur (2001), *Density Estimation For One-Dimensional Dynamical Systems*, ESAIM , Probability & Statististics, pp. 51-76.

[SZ33] G. Szegö (1933), *Orthogonal polynomials*, American Mathematical Society Colloquium Publication, vol. 23.

[TS04] A.B. Tsybakov (2004), *Introduction à l'estimation non-paramétrique*, Springer.

[VI97] G. Viennet (1997), *Inequalities for absolutely regular sequences : application to density estimation*, Probability Theory and Related Fields, vol. 107, pp. 467-492.