Adaptive density estimation under dependence

Gannaz Irène[1], Wintenberger Olivier[2]

9th October 2006

1. Laboratoire de Modélisation et Calcul, Université Joseph Fourier, 38041 Grenoble Cedex 9, France. E-mail: irene.gannaz@imag.fr

2. Statistique Appliquée et MOdélisation Stochastique SAMOS-MATISSE Centre d'Economie de la Sorbonne UMR CNRS 8174 et Université Paris 1 Panthéon-Sorbonne, 90 rue de Tolbiac, 75 634 Paris Cedex 13, France. E-mail: olivier.wintenberger@univ-paris1.fr

Abstract

Assume that $(X_t)_{t\in\mathbb{Z}}$ is a real valued time series admitting a common marginal density f with respect to Lebesgue measure. Donoho *et al.* (1996) propose a near-minimax method based on thresholding wavelets to estimate f on a compact set in an independent and identically distributed setting. The aim of the present work is to extend this methodology to different weakly dependent cases. Bernstein's type inequalities are proved to be sufficient to extend near-minimax results. Assumptions are detailed for dynamical systems and under the η-weak dependence condition from Doukhan & Louhichi (1999). The threshold levels in our estimator integrates the dependence structure of the sequence $(X_t)_{t\in\mathbb{Z}}$ through one parameter γ. The near minimaxity is obtained for $\mathbb{L}^p$-convergence rates ($p \geq 1$). An estimator of γ is obtained by a cross-validation procedure. The procedure is illustrated via a simulation study of some dynamical systems and non Markovian η-weakly dependent sequences.

Keywords. Adaptive estimation, Asymptotic minimax, Hard thresholding, Nonparametric density estimation, Wavelets, Weak dependence.

AMS Subject Classification. 62G07, 60G10, 60G99, 60E15.

Introduction

Let $(X_t)_{t\in\mathbb{Z}}$ be a real valued time series admitting a common marginal density f that is compactly supported. The general purpose of this paper is to estimate f by a wavelet estimator $\hat{f}_n$ constructed from n observations $(X_1, \ldots, X_n)$. We refer to Vannucci (1998) [27] for a survey of the use of wavelet bases in density estimation. Wavelets are interesting because they are localised both in time and frequency. Donoho *et al.* (1996) [8] showed that projection-like linear estimators are not optimal in a minimax approach, i.e. that their rates of convergence are slower than the minimax one. Introducing nonlinearity by thresholding wavelet coefficients allow them to obtain quasi-optimal results. The corresponding estimators are called near-minimax because their rates differ from the minimax one only up to a logarithm term. The present work extends wavelet density estimation from the independent and identically distributed (iid for short) framework to cases where dependence between variables occurs. Many kinds of thresholding exist, such as for example Hall and Patil's (1995) [16] or those presented in Antoniadis and Fan (2001) [2], but we restrict ourselves to hard-thresholding.

Several ways of quantifying dependence have already been worked out. One of the most popular is the notion of mixing. For β-mixing, Tribouley and Viennet (1998) [26] have proved the optimality of an estimator $\hat{f}_n$ for which the Mean Integrated Square Error (MISE for short) is the minimax one. Comte and Merlevède (2002) [5] obtain a near-minimax result for more general α-mixing models, paying the gain of generality by a logarithm loss in the convergence rate. However, the class of α and β-mixing models is quite restrictive. Andrews (1984) [1] exhibits the simple non-mixing process

$$X_t = \frac{1}{2}\left(X_{t-1} + \xi_t\right), \text{ where } (\xi_t)_{t\in\mathbb{Z}} \text{ iid with law } Bern(p). \tag{0.1}$$

New coefficients have been recently introduced to study such non mixing models. Via a time reversion, the Markov chain (0.1) is distributed as the dynamical system $(Y_t = T^t Y_0)_{t\in\mathbb{Z}}$, where Y_0 follow the uniform distribution on $[0,1]$ and $T(x) = 2x \bmod 1$ (see Barbour *et al.* (2000) [3] for more details). As noted by Dedecker and Prieur (2005) [7], such dynamical systems (and associated Markov chains) are dependent but not mixing. The above cited authors introduce the notion of $\tilde{\phi}$-weak dependence to quantify dependence of such processes. Another (eventually) non-mixing class of models is the one of non-causal Bernoulli shifts –$X_t = H(\xi_{t-j}, j \in \mathbb{Z}))_{t\in\mathbb{Z}}$ where $(\xi_t)_{t\in\mathbb{Z}}$ is an iid process–. These models belong to the class of η-weak dependent processes (see Doukhan and Louhichi, 1999, [11]).

The estimation scheme is based on Donoho *et al.* (1996)'s procedure, in [8], and it is adaptative with respect to the regularity of f. The hard-threshold levels of the estimator $\hat{f}_n$ integrate the dependence of the observations via a parameter γ. If the weak dependence context is known, the estimator is near minimax: in the case of dynamical systems the same rate as in iid setting is achieved thanks to inequalities obtained by Collet *et al.* (2002) [4] and for non-causal η-weakly dependent observations another logarithmic loss appears. If the weak dependence context is unknown, a cross validation procedure gives an estimator $\hat{\gamma}_n$. The resulting estimator $\hat{f}_n^{\hat{\gamma}_n}$ is adaptative with respect to the dependence and it is available for a large weak dependence range in an unified approach. We believe that this is a real improvement on existing results because no restrictive mixing assumptions on the observations are required.

The paper is structured as follows. The context of estimation is presented in the next section. Examples of different weakly dependent sequences of interest are given in Section 2. Section 3 is devoted to the main results where relevent probability inequalities and near minimax rates are obtained. In Section 4 some numerical applications are discussed and near minimax density estimators are given by a new cross validation procedure. The proofs are relegated in the last Section.

1 Estimation framework

Without loss of generality f is considered supported by $[0,1]$. Let us recall some useful facts about wavelet estimation (see Hardle, Kerkyacharian, Picard and Tsybakov (1998) [17] for more details on wavelet estimation). For all $p \geq 1$, $\mathbb{L}^p$ denotes in the sequel the space of all functions f supported by $[0,1]$ such that $\|f\|_p^p = \int |f(x)|^p dx < \infty$.

Definition 1. *An orthogonal multiresolution analysis of $\mathbb{L}^2$ is an increasing sequence $(V_j)_{j\in\mathbb{N}}$ of closed subsets of $\mathbb{L}^2$ satisfying*

(i) $\bigcap_{j\in\mathbb{N}} V_j = \{0\}$ *and* $\overline{\bigcup_{j\in\mathbb{N}} V_j} = \mathbb{L}^2$,

(ii) $\forall\ f \in \mathbf{L}^2$, $\forall\ j \in \mathbb{N}$, $f \in V_j$ *if and only if* $x \mapsto f(2^{-j}x)$ *belongs to* V_{j+1},

(iii) *There exists a function ϕ, called father wavelet, such that $\{x \mapsto \phi(x-k)\}_{k\in\mathbb{Z}}$ forms an orthonormal base of V_0.*

If W_j denotes the orthogonal supplement of V_j in V_{j+1}, i.e $V_{j+1} = V_j \oplus W_j$, then there exists a function ψ –called mother wavelet– such that $\{x \mapsto \psi(x-k)\}_{k\in\mathbb{Z}}$ forms an orthonormal basis of W_0. At every resolution level $j \geq 0$ the families $\{\phi_{j,k} : x \to 2^{j/2}\phi(2^j x - k)\}_{k\in\mathbb{Z}}$ and $\{\psi_{j,k} : x \mapsto 2^{j/2}\psi(2^j x - k)\}_{k\in\mathbb{Z}}$ are orthonormal bases of respectively V_j and W_j. Assume that ϕ is supported by $[0,1]$ and has a zero-moments property of order $N \in \mathbb{N}^*$, i.e.:

$$\forall k = 0 \dots N, \quad \int \phi(x)x^k dx = \delta_{0,k}, \quad \int \left|\phi(x)x^{N+1}\right| dx < \infty, \tag{1.1}$$

$$\text{and} \qquad x \mapsto \sum_k |\phi(x-k)| \in \mathbb{L}^2, \tag{1.2}$$

(here $\delta_{0,k} = 1$ if $k = 0$ and else $\delta_{0,k} = 0$). The above assumptions imply that the associated mother wavelet ψ also satisfies

$$\forall k = 0 \dots N, \quad \int \psi(x)x^k dx = 0, \quad \int \left|\psi(x)x^{N+1}\right| dx < \infty.$$

Wavelet bases on $[0,1]$ proposed by Daubechies (1992) [6] are considered with a sufficient number ($N \geq 4$) of vanishing moments to be Lipschitz functions. Recall that a Lipschitz function $h : \mathbb{R}^u \to \mathbb{R}$ for some $u \in \mathbb{N}^*$ is a function such that $\mathrm{Lip}\,(h) < \infty$ with

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

Note that wavelets ϕ and ψ are bounded as Lipschitz functions supported by $[0,1]$.

For any fixed integer j_0, an arbitrary function $f \in \mathbb{L}^2$ can be decomposed as

$$f = \sum_{k=0}^{2^{j_0}-1} \alpha_{j_0,k}\phi_{j_0,k} + \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} \beta_{j_0,k}\psi_{j_0,k},$$

where $\alpha_{j,k} = \int_0^1 f(x)\phi_{j,k}(x)dx$, $\beta_{j,k} = \int_0^1 f(x)\psi_{j,k}(x)dx$. The projection-type estimator is

$$\tilde{f}_n = \sum_{k=0}^{2^{j_0}-1} \hat{\alpha}_{j_0,k}\phi_{j_0,k} + \sum_{j=j_0}^{j_1(n)}\sum_{k=0}^{2^j-1} \hat{\beta}_{j,k}\psi_{j,k},$$

where $\hat{\alpha}_{j,k} = 1/n\sum_{i=1}^n \phi_{j,k}(X_i)$ and $\hat{\beta}_{j,k} = 1/n\sum_{i=1}^n \psi_{j,k}(X_i)$ are the empirical estimators of the coefficients $\alpha_{j,k}$ and $\beta_{j,k}$. Such density estimators can be write on the form: $\tilde{f}_n(x) = \sum_{i=1}^n K_n(x,X_i)$, with $(K_n)_{n\geq 1}$ sequence of kernels defined by $K_n(x,y) = \sum_{k=0}^{2^{j_0}-1}\phi_{j_0,k}(x)\phi_{j_0,k}(y) + \sum_{j=j_0}^{j_1(n)}\sum_{k=0}^{2^j-1}\psi_{j,k}(x)\psi_{j,k}(y)$. The kernels K_n are linear (they can be written as an integral with respect to the empirical distribution) and thus the estimator $\tilde{f}_n$ is linear.

Let (s,π,r) be a triplet such that $s>0$, $1\leq \pi, r\leq\infty$. The expression

$$\|f\|_{s,\pi,r} = |\alpha_{0,0}| + \left(\sum_{j\in\mathbb{N}} 2^{j(s+1/2-1/\pi)r}\left(\sum_{k=0}^{2^j-1}|\beta_{j,k}|^\pi\right)^{r/\pi}\right)^{1/r},$$

defines a semi-norm on $\mathbb{L}^2$ (see Härdle *et al.* (1998) [17] for more details). Besov spaces and Besov balls are respectively denoted

$$\mathcal{B}^s_{\pi,r} := \{f\in\mathbb{L}^2 \text{ such that } \|f\|_{s,\pi,r}<+\infty\},\quad \mathcal{B}^s_{\pi,r}(M) := \left\{f\in\mathcal{B}^s_{\pi,r},\ \|f\|_{s,\pi,r}\leq M\right\}.$$

Note that Besov spaces do not depend on ϕ and ψ. For $f\in\mathcal{B}^s_{\pi,r}(M)$, the $\mathbb{L}^p$-mean error of an estimator f_n is defined by $\mathbb{E}\|f_n - f\|_p^p$. The associated minimax rate verify

$$\inf_{f\in\mathcal{B}^s_{\pi,r}(C)}\ \sup_{f_n \text{ estimator of } f} \mathbb{E}\|f_n-f\|_p^p = \mathcal{O}\left(n^{-\alpha}\right),$$

with

$$\alpha = \begin{cases} s/(1+2s) & \epsilon\geq 0,\\ (s-1/\pi+1/p)/(1+2s-2/\pi) & \epsilon\leq 0,\end{cases} \qquad \text{where } \epsilon = s\pi-(p-\pi)/2. \tag{1.3}$$

Linear estimators, including $\tilde{f}_n$, do not achieve such rates for $f\in\mathcal{B}^s_{\pi,r}(M)$ with $\pi\leq p$. In order to bypass this drawback, Donoho *et al.* (1996) introduce in [8] non-linear estimators $\hat{f}_n$ via thresholding. Let $T_\lambda(\beta) = \beta\mathbb{1}_{|\beta|>\lambda}$ be the hard-threshold function of level $\lambda>0$; given n they consider integers j_0, j_1 and parameters $(\lambda_j)_{j=j_0\cdots j_1}$ and define

$$\hat{f}_n = \sum_{k=0}^{2^{j_0}-1}\hat{\alpha}_{j_0,k}\phi_{j_0,k} + \sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1} T_{\lambda_j}(\hat{\beta}_{j,k})\psi_{j,k},$$

In order to achieve a minimax rate –up to a logarithmic term– the paramters defining $\hat{f}_n$ have to be chosen appropriately. Donoho *et al.* (1996) obtain in the iid framework the following result:

Theorem 1 (Donoho *et al.*, 1996). *Suppose that f belongs to a Besov ball $\mathcal{B}^s_{\pi,r}(M)$ with*

$$1/\pi < s \le N/2,\ 1 \le \pi \le p,\ 1 \le r \le \infty,$$

where N is the regularity of the wavelet. If $(X_t)_{t\in\mathbb{Z}}$ is an iid sequence, there exists a constant $C_0(N,p,s,\pi,M)$ such that

$$\mathbb{E}\|\hat{f}_n - f\|_p^p \le C_0 \begin{cases} \left(\dfrac{\log n}{n}\right)^{p\alpha}, & \text{if } \epsilon \ne 0 \\ \left(\dfrac{\log n}{n}\right)^{p\alpha} (\log n)^{(p/2-\pi/r)_+}, & \text{if } \epsilon = 0, \end{cases}$$

where the minimax rate α and ϵ are given in (1.3) and

$$2^{j_0} \simeq n^{1/(1+N)}, \tag{1.4}$$

$$2^{j_1} \simeq n/\log n, \tag{1.5}$$

$$\lambda_j = K\sqrt{j/n}, \text{ for a suitable constant } K > 0. \tag{1.6}$$

In a regression framework, Donoho and Johnstone (1995) [9] fix the constant $K = \sqrt{2}$ for practical implementation. Hereafter, we will adopt the same choice.

2 Dependent models

Models of this section satisfy the conditions of the theorems stated in Section 3.

2.1 Expanding maps

Dynamical systems $(X_t)_{t\ge 0}$ are defined through a function $T : [0,1] \to [0,1]$ by

$$X_i = T^i(X_0), \qquad \forall i \in \mathbb{N} \tag{2.1}$$

where X_0 is distributed as a Lebesgue dominated measure μ on $[0,1]$ and T^i denotes $\underbrace{T \circ T \circ \cdots \circ T}_{i \text{ terms}}$.

Expanding maps $(X_t = T^t(X_0))_{t\in\mathbb{N}}$ (or equivalently Lasota-Yorke functions T) are dynamical systems such that

- *(Regularity)* The function T is differentiable, with a continuous derivate T' and there exists a grid $0 = a_0 \le a_1 \cdots \le a_k = 1$ such that $|T'(x)| > 0$ on $]a_{i-1}, a_i[$ for each $i = 1, \ldots, k$.
- *(Expansivity)* For any integer i, let I_i be the set on which the first derivate of T^i, $(T^i)'$, is defined. There exists $a > 0$ and $s > 1$ such that $\inf_{x\in I_i}\{|(T^i)'(x)|\} > as^i$.
- *(Topological mixing)* For any nonempty open sets U, V, there exists $i_0 \ge 1$ such that $T^{-i}(U) \cap V \ne \varnothing$ for all $i \ge i_0$.

The class of dynamical systems has remarkable properties (see Viana, 1997, [28]). Its members admit an invariant measure μ_0 with a density $f \in BV_1$ where BV_1 is the set of bounded variation functions h defined on $[0,1]$ such that

$$\sup_{n\in\mathbb{N}} \sup_{a_0=0<a_1<\cdots<a_n=1} \sum_{i=1}^{n} |h(a_i) - h(a_{i-1})| = \|h\|_{BV} < +\infty.$$

Note that $\mathcal{B}^1_{1,1} \subset BV_1 \subset \mathcal{B}^1_{1,\infty}$ (see e.g. Donoho *et al.* (1996), [8]). Expanding maps are also geometrically ergodic in mean, i.e. there exists constants $\alpha, C > 0$ with $\alpha < 1$ such that $\mathbb{E}|X_t - \tilde{X}_t| \le C\alpha^t$ for all $t \ge 0$, where $(\tilde{X}_t)_{t\ge 0}$ is the stationary expanding map $(\tilde{X}_t = T^t(\tilde{X}_0))_{t\in\mathbb{N}}$ obtained with $\tilde{X}_0$ following a distribution μ_0. These models are not mixing (see Dedecker and Prieur, 2005, [7]).

2.2 η-weakly dependence

Doukhan and Louhchi have introduced this notion in 1999 in [11].

Definition 2 (Doukhan and Louhichi, 1999). *The stationary process $(X_t)_{t\in\mathbb{Z}}$ is η-weakly dependent if there exists a sequence of non-negative real numbers $(\eta_r)_{r\in\mathbb{N}}$ satisfying $\eta_r \to 0$ when $r \to \infty$ and such that:*

$$\left|Cov\left(h\left(X_{i_1},\dots X_{i_u}\right),\, k\left(X_{i_{u+1}},\dots,X_{i_{u+v}}\right)\right)\right| \;\le\; \left(u\,\mathrm{Lip}(h) + v\,\mathrm{Lip}(k)\right)\eta_r$$

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

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

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. This notion of dependence includes general models which may be non-mixing. We will consider a subgeometric decay, meaning that:

$$\text{There exist } a, b, C > 0 \text{ such that } \eta_r \le C\exp(-ar^b), \tag{2.2}$$

and we will also assume that the joint densities $f_{j,k}$ of (X_j, X_k) exist and are uniformly bounded for $j \ne k$.

2.3 Bernoulli shifts

Let $H : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ be a measurable function. If the sequence $(\xi_t)_{t\in\mathbb{Z}}$ is iid on $\mathbb{R}$, a Bernoulli shift with input process $(\xi_t)_{t\in\mathbb{Z}}$ is defined as

$$X_t = H\left((\xi_{t-i})_{i\in\mathbb{Z}}\right), \qquad t \in \mathbb{Z}.$$

Such Bernoulli shifts are $\eta-$weakly dependent (see Doukhan and Louhichi, 1999, [11]) with $\eta_r \le 2\delta_{[r/2]}$ if

$$\mathbb{E}\left|H\left(\xi_j, j\in\mathbb{Z}\right) - H\left(\xi_j \mathbb{1}_{|j|\le r}, j\in\mathbb{Z}\right)\right| \le \delta_r. \tag{2.3}$$

Different values of b in equation (2.2) arise naturally for specific functions H.

2.3.1 Infinite moving average

The most simple case of infinitely dependent Bernoulli shift is the infinite moving average process

$$X_t = \sum_{i \in \mathbb{Z}} \alpha_i \xi_{t-i}. \tag{2.4}$$

Doukhan and Lang (2002) [10] prove they are η-weakly dependent with

$$\eta_r = \sqrt{\sum_{|j|>[r/2]} a_j^2}. \tag{2.5}$$

If $(a_j)_{j \neq 0}$ satisfies $a_j \leq K\alpha^{|j|}$ for $j \neq 0$, $K > 0$ and $0 < \alpha < 1$ then equation (2.2) holds with $b = 1$.

2.3.2 LARCH(∞) inputs

A vast literature is devoted to the study of conditionally heteroscedastic models. A simple equation in terms of a vector valued process allows a unified treatment of those models, see [13]. Let $(\xi_t)_{t \in \mathbb{Z}}$ be an iid centered real valued sequence and $a, a_j, j \in \mathbb{N}^*$ be real numbers. LARCH(∞) models are solutions of the recurrence equation

$$X_t = \xi_t \left(a + \sum_{j=1}^{\infty} a_j X_{t-j} \right). \tag{2.6}$$

Assume that $\Lambda = \mathbb{E}|\xi_0| \sum_{j \geq 1} |a_j| < 1$ then one (essentially unique) stationary solution of eqn. (2.6) is given by

$$X_t = \xi_t \left(a + \sum_{k=1}^{\infty} \sum_{j_1, \ldots, j_k \geq 1} a_{j_1} \xi_{t-j_1} a_{j_2} \cdots a_{j_k} \xi_{t-j_1-\cdots-j_k} a \right). \tag{2.7}$$

The solution (2.7) of equation (2.6) is geometrically $\eta-$weakly dependent with $b = 1/2$ if there exists $0 < \alpha < 1$ and $K > 0$ such that $a_j \leq K\alpha^{j-1}$ for all $j > 0$ and $\mathbb{E}|\xi_0| < 1 - \alpha$.

2.3.3 Non-Causal LARCH(∞) inputs

The previous approach extends for the case of Non-Causal LARCH(∞) inputs

$$X_t = \xi_t \left(a + \sum_{j \neq 0} a_j X_{t-j} \right).$$

Doukhan, Teyssière and Winant (2005) prove in [13] the same results of existence of a stationary solution as for the previous causal case (only replace summation for $j > 0$ by summation for $j \neq 0$). This solution satisfies equation (2.2) with $b = 1/2$ if there exists $K, \alpha > 0$ and $\alpha < 1$ such that $a_j \leq K\alpha^{|j|}$ for all $j \neq 0$.

3 Main results

In all this section, let $(X_t)_{t\in\mathbb{Z}}$ be a stationary real valued sequence, ϕ a father wavelet in $\mathbb{L}^2$ satisfying (1.1) for $N \geq 4$ and the condition (1.2), and $\{\psi_{j,k}, j \in \mathbb{N}, k \in \mathbb{Z}\}$ the wavelet functions associated with. A probability inequality is given in an unified way for the different cases of dependence introduced in Section 2. Theorem 3 gives near-minimax estimators for weakly dependent observations.

3.1 Probability inequalities

Under weakly dependent conditions on $(X_t)_{t\in\mathbb{Z}}$, we may find that for all $(j,k) \in \mathbb{N}^2$, $n \in \mathbb{N}^*$ and $q > 0$ there exists a constant $C > 0$ such that

$$\mathbb{E}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\psi_{j,k}(X_i) - \mathbb{E}\psi_{j,k}(X_i))\right|^q \leq C. \tag{3.1}$$

Here, the constant C increases with the order of the moment q. In iid framework, this inequality easily follows from Rosenthal's one. Under some additional assumptions, C could be bounded by $q^{q/2}$, proportional to the moment of order q of a Gaussian. For expanding maps, the moment inequality (3.1) also holds (see Dedecker and Prieur, 2005, [7]). For η-weakly dependent sequences, Ragache and Wintenberger (2006) prove in [25] such inequality under condition (2.2). They also bound the constant $C < Kq^{q+q/b}$ where K is a constant that does not depends on q. The rate $q^{q/2}$ is not obtained because covariance terms appears in the development of the moments.

These bounds of q-th moments are linked to probability inequalities by the following useful lemma:

Lemma 1 (Ragache and Wintenberger, 2006). *If the variables $\{V_n\}_{n\in\mathbb{Z}}$ satisfy, for all $k \in \mathbb{N}^*$*

$$\|V_n\|_{2k} \leq \Phi(2k)$$

where Φ is an increasing function with $\Phi(0) = 0$ and $\lim_{x\to+\infty}\Phi(x) = +\infty$, then:

$$\mathbb{P}(|V_n| > \delta) \leq e^2 \exp\left(-\Phi^{-1}(\delta/e)\right).$$

Taking $V_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{j,k}(X_i) - \mathbb{E}\psi_{j,k}(X_i)$ and $\Phi(x) = Cx^{1+1/b}$ leads directly to the first assertion of the following Theorem:

Theorem 2. *For all $(j,k) \in \mathbb{N}^2$ and $n \in \mathbb{N}^*$ there exists constants $B, C > 0$ and $\gamma > 1/2$ such that*

$$\mathbb{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\psi_{j,k}(X_i) - \mathbb{E}\psi_{j,k}(X_i))\right| \geq \delta\right) \leq B\exp\left(-C\delta^{1/\gamma}\right), \tag{3.2}$$

for all $\delta \geq 0$ such that there exists $K' > 0$ with $\delta^{2l}\|\psi_{j,k}\|_\infty^2 \leq K'n$ for some constant $l \geq 0$. More precisely

1. *$\gamma = 1 + 1/b$ and $l = 0$, if coefficients η_r of $(X_t)_{t\in\mathbb{Z}}$ satisfy (2.2) and if for all $j \neq k$ the joints densities $f_{j,k}$ of (X_j, X_k) exist and are uniformly bounded.*

2. $\gamma = 0.5$ *and* $l = 1$*, if* $(X_t)_{t\in\mathbb{Z}}$ *is an iid process.*

3. $\gamma = 0.5$ *and* $l = 5$*, if* $(X_t)_{t\in\mathbb{Z}}$ *is a stationary expanding map.*

Assertions 2 and 3 of Theorem 2 follow directly from the following Bernstein's type inequality: for all $\lambda > 0$

$$P\left(\left|\sum_{i=1}^{n}\psi_{j,k}(X_i) - \mathbb{E}\psi_{j,k}(X_i)\right| \geq \lambda\right) \leq B\exp\left(-C\frac{\lambda^2}{\sigma_n^2 + \lambda^{\frac{2l}{l+1}}\|\psi_{j,k}\|_\infty^{\frac{2}{l+1}}}\right), \tag{3.3}$$

where $B, C > 0$ and $\sigma_n^2 = \mathrm{Var}\sum_{i=1}^{n}\psi_{j,k}(X_i)$. From equation (3.1) in the case $q = 2$, we know that σ_n^2 has the rate n. Fixing $\delta = \lambda/\sqrt{n}$ and under $\delta^{2l}\|\psi_{j,k}\|_\infty^2 \leq K'n$, $\lambda^{2l/(l+1)}\|\psi_{j,k}\|_\infty^{2/(l+1)} \leq K''n$ with $K'' > 0$, the result of Theorem 2 with $\gamma = 0.5$ follows. Equation (3.3) with $B = 2$ and $C = 1$ is the classical Bernstein's one in the iid case, see for instance Petrov (1995), [23]. The result is new for expanding maps and its proof is given in the last section.

3.2 Near-minimaxity of the estimation scheme

This result extends the one of Donoho *et al.* (1996), [8], to dependent settings.

Theorem 3. *Suppose that* f *belongs to a Besov ball* $\mathcal{B}^s_{\pi,r}(M)$ *with*

$$1/\pi < s \leq N/2,\ 1 \leq \pi \leq p,\ 1 \leq r \leq \infty,$$

where $N \geq 4$ *is the regularity of the wavelet. Suppose that there exists a couple* (γ, l) *and that conditions of dependence on* $(X_t)_{t\in\mathbb{Z}}$ *hold such that the inequalities (3.1) and (3.2) are satisfied. Then there exists a constant* $C_0(N, p, s, \pi, M)$ *such that*

$$\mathbb{E}[\|\hat{f}_n - f\|_p^p] \leq C_0 \begin{cases} \left(\dfrac{\log^{2\gamma} n}{n}\right)^{p\alpha} & \text{if } \epsilon \neq 0 \\ \left(\dfrac{\log^{2\gamma} n}{n}\right)^{p\alpha}(\log n)^{(1-\pi/r)_+} & \text{if } \epsilon = 0 \end{cases}$$

where the minimax rate α *and* ϵ *are given in (1.3) and*

$$2^{j_0} \simeq n^{1/(1+N)}, \tag{3.4}$$

$$2^{j_1} \simeq n/\log^{2\gamma l} n, \tag{3.5}$$

$$\lambda_j = Kj^\gamma/\sqrt{n}, \text{ for a well chosen constant } K > 0. \tag{3.6}$$

The above result takes into account the dependence of the observations through the threshold. The dependence of the observations presently changes the behavior of the moments of wavelet coefficients, determined by Theorem 2. The idea of adapting the threshold level to the behavior of the moments of the wavelet coefficients estimators is not new but it is usually developped in ill-posed inverse problems. As it is put in evidence in the Theorem 5.1 of Kerkyachrian and Picard (2000), [20], the quality of the estimator relies on an adapted threshold level, that is given by moments and probability inequalities on the estimators of the wavelet coefficients. For example, Johnstone *et al.* (1998) apply this principle in an ill-posed inverse problems in [19]. Here the idea is exactly the same, but the moment inequalities obtained relie on the dependence of the setting rather than on the inversion of the problem.

The result for $\gamma = 0.5$ is the same as in Donoho *et al.* (1996), [8]. As noted previously, such values of γ arise in iid context and for dynamical systems. Consequently, Donoho *et al.* (1996) result is extended to dynamical systems. It is also extended to expanding maps, where the assertion 1. of Theorem 2 holds, but the rate obtained differs from a logarithmic term.

The parameter γ determines the convergence rates through Theorem 3. It also calibrates the degree of dependence of the observations through Theorem 2. The optimal γ_0 corresponds to the smallest error of estimation. This value could change with the dependence of the observations. Theorem 2 gives us a theoretical value γ in inequality (3.2) (see section 3.1) that leads to near minimax estimators. Such probability inequality exhibits a bound $\gamma_0 \leq \gamma$ but not necessarily the optimal γ_0.

4 Numerical results

In the next section, we propose to estimate numerically the optimal parameter γ_0 by an estimator $\hat{\gamma}_n$ obtained by a cross-validation procedure.

4.1 Cross-validation procedure

According to Theorem 3, near-minimax estimators $\hat{f}_n^\gamma$ with hard-threshold of the form $\lambda_j = Kj^\gamma/\sqrt{n}$ (with $K, \gamma > 0$) are considered in the sequel. Let us fix $K = \sqrt{2}$, $2^{j_0} = n^{1/(1+N)}$ and $2^{j_1} = n/\log n$ as in Donoho and Johnstone (1995), [9], in order that $\hat{f}_n^\gamma$ depends only on γ. The MISE, corresponding to the $\mathbb{L}^p$-mean error for $p = 2$, is:

$$MISE(\hat{f}_n^\gamma) = \mathbb{E} \int \left(\hat{f}_n^\gamma(x) - f(x)\right)^2 dx = \mathbb{E}L(\hat{f}_n^\gamma).$$

Note that

$$L(\hat{f}_n^\gamma) = \underbrace{\int \left(\hat{f}_n^\gamma(x)\right)^2 dx - 2\int \hat{f}_n^\gamma(x) f(x) dx}_{J(\hat{f}_n^\gamma)} + \underbrace{\int f^2 dx}_{constant}.$$

Minimizing the MISE is then equivalent to minimize $\mathbb{E}J(\hat{f}_n^\gamma)$. Following Hart and Vieu (1990), [18], we define a leave-out procedure. Let b_n be positive integers and $\mathcal{X}_{-i}$, $i = 1 \ldots n$, be the associated sub-sampling:

$$\mathcal{X}_{-i} = \{X_j,\ 1 \leq j \leq n,\ |i - j| \geq b_n\}.$$

We define the "leave-out b cross validation function" by:

$$CV_b(\gamma) = \int \left(\hat{f}_n^\gamma(x)\right)^2 dx - 2n^{-1} \sum_{i=1}^{n} \hat{f}_{-i}^\gamma(X_i),$$

where $\hat{f}_{-i}^\gamma$ is the hard threshold estimator based on the sub-sampling $\mathcal{X}_{-i}$.

The idea is to break the dependence by considering blocks of size $2 * b_n + 1 \to_{n\to\infty} \infty$ around the observation X_i in order to obtain $n^{-1}\sum_{i=1}^n \hat{f}_{-i}^\gamma(X_i) \sim_{n\to\infty} \mathbb{E}\hat{f}_{-i}^\gamma(X) \sim_{n\to\infty} \int \hat{f}_n^\gamma(x)f(x)dx$. Lahiri (2003) shows in [21] the efficiency of such sub-sampling methods in a weakly dependent context. Hall *et al.* (1995) prove in [15] that for kernel estimators the bandwith chosen by minimizing the corresponding "leave-out b cross validation function" is a good estimator of the optimal bandwith. Theoretical results are not developed here but we conjecture that such result remains true for two reasons:

- all cases considered here are (sub-)geometrically weakly dependent and restrictions on decays of coefficient ensure asymptotic results as law of large numbers,
- as shown by Theorem 3, our parameter γ plays a fundamental role on the convergence rates as much as the classical bandwith does.

According to remark 2.1 of Hart and Vieu (1990), [18], we fix $2 * b_n \sim n^{1/3}$. Note that the orthonormality of wavelets basis gives

$$CV_b(\gamma) = \sum_{k=0}^{2^{j_0}-1} \hat{\alpha}_{j_0,k}^2 + \sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1} T^2_{\sqrt{2}j^\gamma/\sqrt{n}}(\hat{\beta}_{j,k}) - 2n^{-1}\sum_{i=1}^n \hat{f}_{-i}^\gamma(X_i).$$

Our proposed procedure consists in choosing $\hat{\gamma}_n = \arg\min_{\gamma=0,0.05,\ldots,1.95,2} CV_b(\gamma)$.

The implementation has been done in Matlab using the package Wavelab, heely available on the net on [29]. In order to implement numerically the procedure, we have used some approximations; indeed, computation is based on an equispaced grid, while the data considered is not equispaced. In order to reduce the resulting approximation error, we consider a finre grid than the number n of observations. Here $I = 4n$ in order to increase the precision without araising too much calculus time. Actually, we will consider in implementation values of the form $\psi_{j,k}(l_i/I)$, wether than $\psi_{j,k}(X_i)$, with l_i the integer part of X_iI. The bias caused by implementation is not studied here.

4.2 The estimator $\hat{\gamma}_n$

This section illustrates and evaluates our estimation procedure on simulated examples.

Consider first the case of an iid sample $(X_1, \ldots, X_n)$ generated according to the cumulative distribution function $F(x) = 2/\pi \arcsin(\sqrt{x})$ for $x \in [0,1]$. The criteria is minimized for $\gamma = 0.5$ in Figure 1. Thus, the cross validation procedure selects $\hat{\gamma}_n$ near 0.5 for $n = 2^{10}$ (see Figure 2). For a sufficiently large number of observations this estimator $\hat{\gamma}_n$ obtained numerically has value concording with $\gamma = 0.5$ chosen by Donoho *et al.* (1996) in [8].

Let us now consider the case of an expanding map $T(x) = 4x(1-x)$. The invariant measure has the same distribution F as in the previous iid case (see Prieur, 2001, [24]). We then first simulate $\tilde{X}_1, \ldots, \tilde{X}_n$ such that $\tilde{X}_1 \rightsquigarrow F$ and then $\tilde{X}_i = T^{i-1}(\tilde{X}_1)$ for $2 \le i \le n$. As for the iid case, the cross validation criteria in Figure 1(a) chooses $\hat{\gamma}_n = 0.5$. Figure 2 corroborates this value: for high values of n, the distribution of $\hat{\gamma}_n$ obtained by croos-validation is centered on the value 0.5.

The assumption of stationarity could be irrelevant in many cases. Simulating the first observation $\tilde{X}_1 \rightsquigarrow F$ to estimate the (unknown) distribution F of the invariant measure of a given transformation T is impossible. Thus we simulate $X_i = T^i(X_0)$ for $2 \leq i \leq 2n$ with $X_0 \rightsquigarrow \mathcal{U}([0,1])$ and retain only the n last terms. The process $(X_t)_{t\geq 1}$ is geometrically ergodic in mean and the error with respect to the stationary case is negligeable regarding with the error of the density estimation. Results remain valid and the procedure acts as in previous cases of stationary and iid cases, see Figures 1(b) and 2.

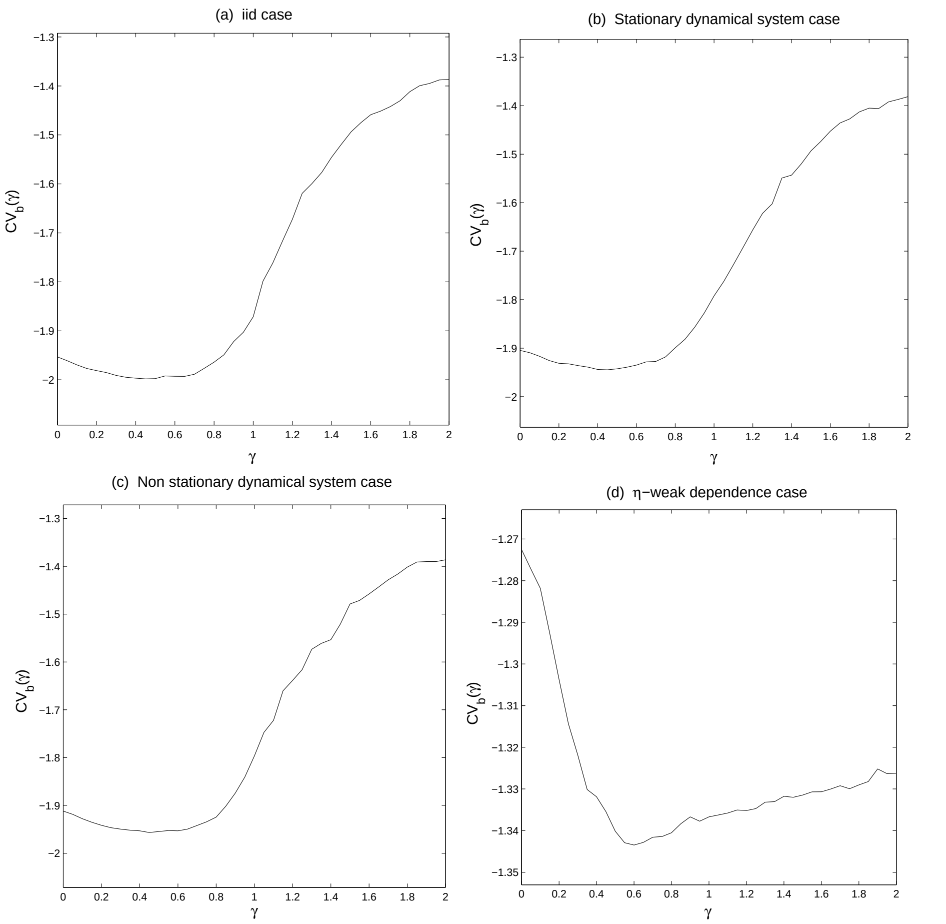

Figure 1: Cross validation criterion's evolution with respect to γ for $n = 2^{10}$ observations. The curves represent the evolution of the mean of the criteria calculated on 100 simulations with respect to $\gamma = 0, 0.05 \ldots 1.95, 2$. Four cases were considered: in (a) the observations are iid, in (b) we simulate stationary dynamical system, in (c) a non stationary dynamical system and in (d) we consider a η-weak dependent case.

Study now a non-Markovian extension of Andrews' example (0.1)

$$X_t = 2(X_{t-1} + X_{t+1})/5 + 5\xi_t/21 \text{ where } (\xi_t)_{t\in\mathbb{Z}} \text{ are iid with the common law } Bern(p). \tag{4.1}$$

A stationary solution of this AR(2) equation is the non-causal process $(X_t)_{t\in\mathbb{Z}}$

$$X_t = \sum_{j\in\mathbb{Z}} a_j \xi_{t-j},$$

where $a_j = 1/3 * (1/2)^{|j|}$. This solution belongs to $[0,1]$ and its density is the one of $(U + U' + \xi_0)/3$ where U and U' are independent variables following $\mathcal{U}([0,1])$. As in Andrews' example such processes are not mixing. However, using equation (2.5) there exists $a, C > 0$ such that $\eta_r \leq C\exp(-ar)$. Assumption (2.2) is satisfied with $b = 1$.

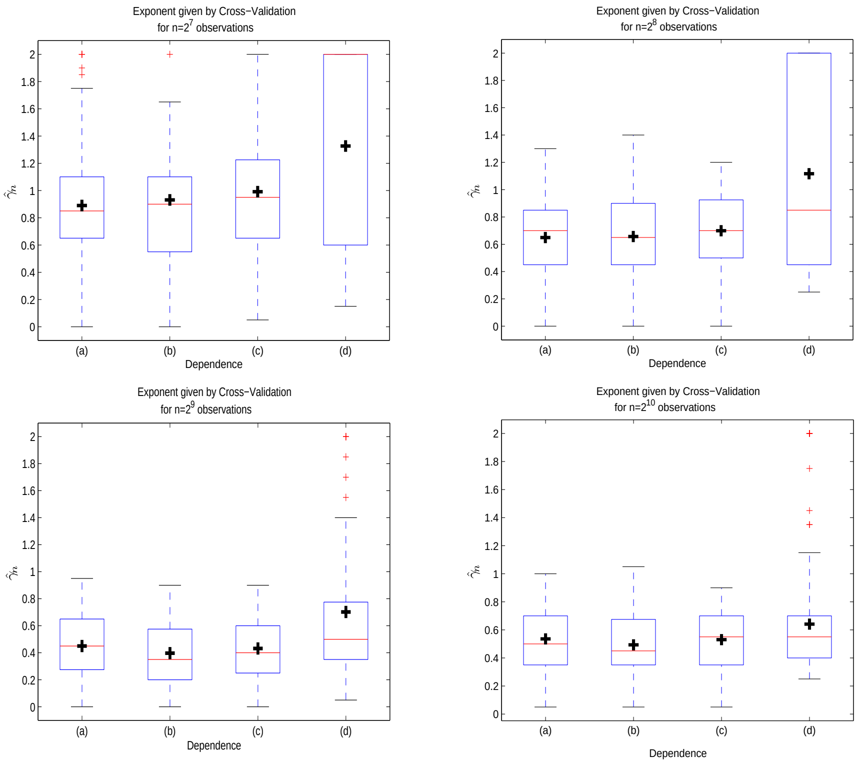

Figure 2: Empirical distribution of $\hat{\gamma}_n$. The figures represent the boxplots (the black cross is the mean) of the empirical ditribution of $\hat{\gamma}_n$ obtained on 100 simulations for different initial sample sizes: $n = 2^7,\ 2^8,\ 2^9,\ 2^{10}$. For each value of n, we abut the boxplots of differents cases; From left to right : (a) iid case, (b) stationary dynamical system case, (c) non-stationary dynamical system case and (d) the η-weak dependent case.

Perfect simulation is not available in this framework and Gibbs algorithm is appropriate for approching the true distribution of $(X_1, \ldots, X_n)$. A vectorial Markov Chain $(X_1^{(t)}, \ldots, X_{n+2}^{(t)})_{t\in\mathbb{N}^*}$ is simulated by:

$$\text{forall } j \in \{1\ldots 2n\}, \quad \text{the distribution of } X_j^{(0)} \text{ is uniform on } [0,1],$$
$$X_j^{(t)} = 2(X_{j-1}^{(t)} + X_{j+1}^{(t-1)})/5 + 5 * \xi_t/21.$$

This chain is uniformly geometrically ergodic and its invariant distribution is the true one of $(X_1, \ldots, X_{n+2})$. The dependence on the initial values is then geometrically decreasing with the number of iteration of the Gibbs sampler (see Guyon, 1995, [14]). One chooses to take as observations $(X_2^{(n)}, \dots, X_{n+1}^{(n)})$ in order to reduce edge effects. The error coming from the simulation compared with the error of estimation is negligable. The shape of the criteria in Figure 1 is completely different in this non Markovian context. The cross-validation procedure leads to larger choices of $\hat{\gamma}_n$ than before (see Figure 2).

This numerical procedure has a practical interest especially when dealing with density estimation for time series. In practice, when the independence of the observations is not acquired, the cross-validation estimator $\hat{f}_n^{\hat{\gamma}_n}$ seems more adapted than the classical one with $\gamma = 0.5$.

4.3 Numerical study of the convergence rates

This section does not deal directly with our cross validation procedure. We study numerically the rates of convergence in different contexts by Monte Carlo simulations. The true densities of the simulated sequences are known. The $\mathbb{L}^2$-error rates is approximated by the Riemannian sum:

$$2^{-5} \sum_{i=0}^{2^5-1} \hat{f}_n^{\gamma}(2^{-6} + i2^{-5}) - f(2^{-6} + i2^{-5}).$$

With the parameter γ fixed at 0.5, we estimate the MISE on 100 iterations using this formula for the different cases of dependence. Figure 3 represents the evolution of the values of MISE obtained with respect to the sample-size n in the cases of iid series, stationary and non stationary expanding maps. We can observe that the evolution is the same for the differents cases considered in this figure. Added to the fact that the value of γ given by Theorem 2 is the same and that the evolutions of the criterion in Figure 2.(a) to (c) are very similar, compared with the η-weak dependent case, this means that in our example, these kinds of causal dependence do not modify the behaviour of the classical estimator from the iid case. In particular, we can note that the framework of non stationnary expanding maps is very similar to the iid case for our estimation procedure.

As already mentioned by Tribouley and Viennet (1998) in [26], a safe strategy to avoid large errors dealing with dependent data is to increase the threshold. A way to enlarge the threshold is to choose a larger γ. The special shape of the criteria shows that $\gamma = 0.5$ is a critical value and larger ones seems much more stable in the non causal context. In figure 4 are ploted estimations on 100 iterations of the MISE for three choices of $\gamma = 0.5, 0.75, 2$ and for various numbers of observations $n = 2^6, 2^7, \dots, 2^{14}, 2^{15}$. The MISE is clearly the smallest for $\gamma = 0.75$.

Theorically Theorem 2 tolds us inequality (3.2) holds with $\gamma = 2$ (see Section 3.1 for more details). Actually, this Theorem only gives an upper bound of γ veryfing inequality (3.2) and smaller choices of γ can be possible. Choosing a lower γ such that inequality (3.2) holds allows to reduce the convergence rate via Theorem 3. But as we do not have a lower bound, we do not have access to the optimal choice of such γ. Figure 4 confirms that $\gamma = 2$ is probably not the optimal choice of γ in the studied example because we observe that the MISE of $\hat{f}_n^2$ is always greater than the one obtained with $\hat{f}_n^{0.75}$ even for large values of

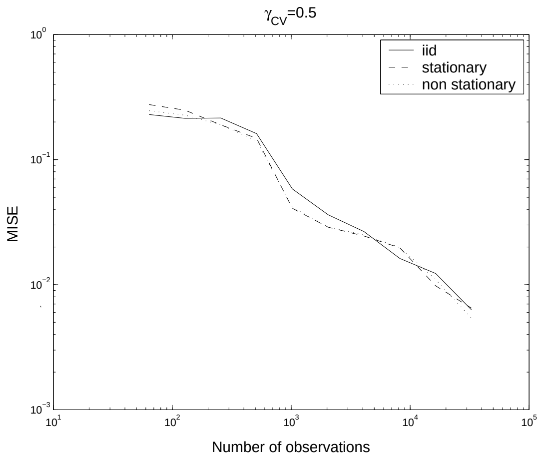

Figure 3: Evolution of the MISE with respect to the sample size. The figure represents the estimation of the MISE obtained by 100 simulations for $n = 2^6, 2^7, \dots, 2^{14}, 2^{15}$ in a log-log scale. Three cases were considered : iid observations (solid), stationary expanding maps (dash) and non stationary expanding maps (dots)

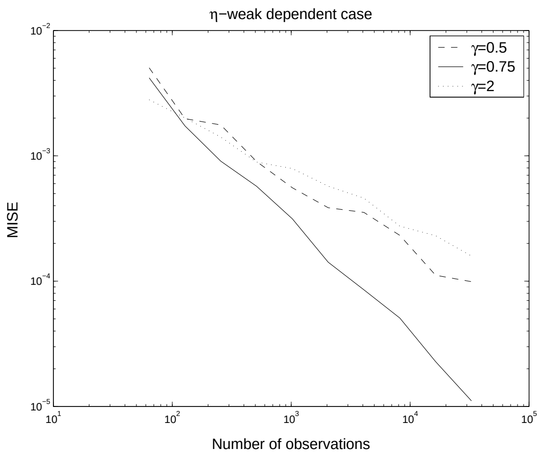

Figure 4: Evolution of the MISE with respect to the sample size for η-weak dependent data. The figure represents the estimation of the MISE obtained by 100 simulations for $n = 2^6, 2^7, \dots, 2^{14}, 2^{15}$ in a log-log scale. Three values of γ were considered : $\gamma = 0.5$ (dash), $\gamma = 0.75$ (solid) and $\gamma = 2$ (dots)

n. The fact $\hat{f}_n^{0.5}$ also have less satisfying results than $\hat{f}_n^{0.75}$ means a priori that inequality (3.2) is not available with $\gamma = 0.5$ in this context.

This study put in evidence the limitation of our theorical study: Theorem 2 does not gives the optimal choice of γ but only a possible choice. Using a cross-validation procedure allows us to bypass this drawback, giving the value of γ the most adapted to the observations and leading to a better estimation than an arbitrary choice of the threshold.

Conclusion

Probability inequalities give a γ that controls the logarithmic loss in the convergence rate. It leads to near minimax estimators $\hat{f}_n^{\gamma}$. This value of γ is not necessarly the optimal one, i.e. the one that minimizes the error of the estimator. The proposed estimator $\hat{f}_n^{\hat{\gamma}_n}$, where $\hat{\gamma}_n$ is obtained by a cross-validation procedure, seems a better density estimator when dealing with time series. For iid sequences or expanding maps the estimator $\hat{\gamma}_n$ converges to 0.5 according to Theorem 3. Even for non stationary expanding maps the previous result remains valid. In the η-weak dependence case, larger values $\gamma > 0.5$ are preferable, like the ones given by the cross validation procedure. Both theorical results and implementation on simulated examples tolds us that the behavior of the hard-threshold density estimator is similar for expanding maps (stationary or not) than for the iid case, while non causal η-weak dependence needs to calibrate differently the threshold. In our opinion, the cross-validation procedure presented gives a satisfactory unified approach of these different cases.

5 Proofs

In this section, proofs of Theorem 3 and inequality (3.3) are collected.

5.1 Proof of the Theorem 3

We restrict ourselves to the case of compactly supported distributions. We consider that f is defined on $[0, 1]$ without loss of generality. We assume furthermore that the function f we wish to estimate belongs to a Besov Ball $\mathcal{B}^s_{\pi,r}(M)$ and that the assumptions on the indexes given in Theorem 3 hold. The density f can be written as follows:

$$f = \underbrace{\sum_{k=0}^{2^j_0-1} \alpha_{j_0,k}\phi_{j_0,k}}_{E_{j_0}f} + \underbrace{\sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1} \beta_{j,k}\psi_{j,k}}_{D_{j_0,j_1}f} + \underbrace{\sum_{j=j_1+1}^{\infty}\sum_{k=0}^{2^j-1} \beta_{j,k}\psi_{j,k}}_{D_{j_1,\infty}f}\,.$$

This decomposition is linked with the properties of the Besov Balls: the first term $E_{j_0}f$ represents the projection of f over the space generated by scale functions of order j_0 whereas the part $D_{j_0,j_1}f + D_{j_1,\infty}f$ are the "details", it is to say the projection over the wavelet spaces. (See Härdle *et al.* (1998), [17]).

Let us recall below the form of the estimator:

$$\hat{f}_n = \underbrace{\sum_{k=0}^{2^j_0-1} \hat{\alpha}_{j_0,k}\phi_{j_0,k}}_{\hat{E}_{j_0}f} + \underbrace{\sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1} \gamma_{\lambda_j}(\hat{\beta}_{j,k})\psi_{j,k}}_{\hat{D}_{j_0,j_1}f},$$

where γ_{λ_j} is the hard-thresholding function with the level dependent threshold $\lambda_j = K\frac{j^{\gamma}}{\sqrt{n}}$. The functions j_0 and j_1 are given by $j_0 = \log_2(n^{1/(1+N)})$ and $j_1 = \log_2(n/\log^{2\gamma l} n)$.

Thanks to Minkowski's inequality, we can decompose the risk of $\hat{f}_n$ in three parts:

$$\mathbb{E}[\|\hat{f}_n - f\|_p^p] \leq 3^{p-1}\left(\underbrace{\mathbb{E}[\|\hat{E}_{j0}f - E_{j0}f\|_p^p]}_{T_1} + \underbrace{\mathbb{E}[\|\hat{D}_{j_0,j_1}f - D_{j_0,j_1}f\|_p^p]}_{T_2} + \underbrace{\|D_{j_1,\infty}f\|_p^p}_{T_3}\right).$$

We may now study the convergence rate of each of these terms. We will not consider them in their order of appearance but of difficulty.

5.2 Some technical tools

We suppose the series (X_i) satisfy the conditions (3.1) and (3.2) with γ a positive constant. Those assumptions provide us moments inequalities for the estimation of the scale and wavelet coefficients. We have, for all $j \in \mathbb{N}$ and $k \in \mathbb{Z}$:

$$\begin{cases} \hat{\alpha}_{j,k} - \alpha_{j,k} &= \frac{1}{n}\sum_{i=1}^{n}(\phi_{j,k}(X_i) - \mathbb{E}\phi_{j,k}(X_i)), \\ \hat{\beta}_{j,k} - \beta_{j,k} &= \frac{1}{n}\sum_{i=1}^{n}(\psi_{j,k}(X_i) - \mathbb{E}\psi_{j,k}(X_i)). \end{cases}$$

When we restrict ourselves to the cases where $j \leq j_1$, we can also control more generally the terms: $\mathbb{E}|\hat{\alpha}_{j,k} - \alpha_{j,k}|^p$ and $\mathbb{E}|\hat{\beta}_{j,k} - \beta_{j,k}|^p$ for a fixed real $p > 0$ and uniformly for all $j \leq j_1$ and $k \in \mathbb{Z}$. Then $\|\phi_{j,k}\|_\infty$ and $\|\psi_{j,k}\|_\infty$ are bounded, up to a constant, by $2^{j_1/2}$, which is always smaller than $\sqrt{n}$. Through the inequality (3.1) it leads there exists $C > 0$ such that

$$\mathbb{E}|\hat{\beta}_{j,k} - \beta_{j,k}|^p \leq Cn^{-p/2}. \tag{5.1}$$

The functions ϕ satisfy also the condition (1.1) and the same inequalities hold for each $\alpha_{j,k}$.

5.3 Approximation error T_3

This term is the bias introduced by the fact that in reality we do not estimate f but only its projection over a space of scale functions of order j_1. The fact that we observe dependent data does not affect this term because it is deterministic, so we can apply the usual bounds. As in the proof of Theorem 5 of Donoho et al. (1996), [8], or by applying Theorem 9.3 of Härdle *et al.* (1998), [17], there exists a constant $C > 0$ such that

$$T_3 \leq C\left(2^{-j_1 s'}\right)^p \text{ where } s' = s - 1/\pi + 1/p.$$

The index s' is coming from the Sobolev inclusion $\mathcal{B}^s_{\pi,r} \subset \mathcal{B}^{s'}_{p,r}$. Recall that we have taken $2^{-j_1} = \log^{2\gamma l} n/n$. Moreover, $p \geq \pi$ and $s > 1/\pi$ implies $s' > \alpha$:

- If $\epsilon \leq 0$, $\alpha = s'/(1 + 2(s - 1/\pi))$. Then $\alpha < s'$ because $s > 1/\pi$ by hypothesis.
- If $\epsilon \geq 0$, $\alpha - s' = s/(1 + 2s) - s' = -(2s^2 + (1 + 2s)(1/\pi - 1/p))/(1 + 2s)$. As $s > 0$, we conclude noting that $s' > \alpha$.

As a consequence T_3 has always a smaller rate than the convergence one, i.e. $T_3 \leq C\left(\log^{2\gamma l} n/n\right)^{s'p} \leq C'\left(\log^{2\gamma} n/n\right)^{\alpha p}$ for a well chosen constant $C' > 0$.

5.4 Bias of scale estimation T_1

Paralleling the proof of Theorem 5 of Donoho *et al.* (1996),, [8], we can apply the result of Meyer (1992), [22], for the scaling function ϕ satisfying a concentration assumption. It gives, for a suitable $C > 0$

$$T_1 = \mathbb{E} \left\| \sum_{k=0}^{2_0^j - 1} (\hat{\alpha}_{j_0,k} - \alpha_{j_0,k}) \phi_{j_0,k} \right\|_p^p \leq C\, 2^{j_0(p/2-1)} \sum_{k=0}^{2^{j_0}-1} \mathbb{E} |\hat{\alpha}_{j_0,k} - \alpha_{j_0,k}|^p.$$

Thanks to the inequality (5.1), the rate of convergence of T_1 is bounded by

$$T_1 \leq C\, 2^{j_0 p/2} n^{-p/2}.$$

Note that the choice (3.4) of j_0 implies that the order of the bound is $(2^{j_0}/n)^{p/2} = n^{-pN/(2+2N)}$. We conclude that T_1 is negligible thanks to the hypothesis $N \geq 2s$.

5.5 Details term T_2

We apply the same result from Meyer (1992), [22], on the wavelet basis $\{\psi_{j,k}\}$:

$$T_2 = \mathbb{E} \left\| \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j - 1} (\gamma_\lambda(\hat{\beta}_{j,k}) - \beta_{j,k}) \psi_{j,k} \right\|_p^p \sim \sum_{j=j_0}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} \mathbb{E} |\gamma_\lambda(\hat{\beta}_{j,k}) - \beta_{j,k}|^p.$$

In order to prove that this term achieves the desired rate of convergence, we need to decompose it again. We will need in particular to distinguish whether the estimation of the coefficients are thresholded or not. We therefore introduce the following sets:

$$\begin{aligned} \hat{B}_j &:= \left\{ k = 0 \dots 2^j - 1,\quad |\hat{\beta}_{j,k}| > \lambda_j \right\}, \\ B_j^+ &:= \left\{ k = 0 \dots 2^j - 1,\quad |\beta_{j,k}| > 2\lambda_j \right\}, \\ B_j^- &:= \left\{ k = 0 \dots 2^j - 1,\quad |\beta_{j,k}| < \lambda_j/2 \right\}. \end{aligned}$$

Then $T_2 \leq T_{21} + T_{22} + T_{23} + T_{24}$ with:

$$\begin{aligned} T_{21} &= \sum_{j_0}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} \mathbb{E} \left[|\hat{\beta}_{j,k} - \beta_{j,k}|^p \mathbb{1}_{\hat{B}_j \cap B_j^-} \right], \\ T_{22} &= \sum_{j_0}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} \mathbb{E} \left[|\hat{\beta}_{j,k} - \beta_{j,k}|^p \mathbb{1}_{\hat{B}_j \cap B_j^{-c}} \right], \\ T_{23} &= \sum_{j_0}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \mathbb{P}(\hat{B}_j^c \cap B_j^+), \\ T_{24} &= \sum_{j_0}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \mathbb{P}(\hat{B}_j^c \cap B_j^{+c}). \end{aligned}$$

where the exponent c denotes the complementary.

5.5.1 Term T_{23}

Within the set $\hat{B}_j \cap B_j^-$ or the set $\hat{B}_j^C \cap B_j^+$, we can notice that $|\hat{\beta}_{j,k} - \beta_{j,k}|$ is lower bounded by $\lambda_j/2$. We then have for a constant $C > 0$:

$$T_{23} \leq C \sum_{j_0}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \mathbb{P}(|\hat{\beta}_{j,k} - \beta_{j,k}| > \lambda_j/2).$$

We can have an upper bound of $\mathbb{P}(|\hat{\beta}_{j,k} - \beta_{j,k}| > \lambda_j/2)$ thanks to Theorem 2, choosing $\delta = \sqrt{n}\lambda_j/2 = Kj^\gamma/2$. With $2^{j_1} = n(\log n)^{-2\gamma}$, for a sufficiently large n:

$$\delta^{2l}\|\psi_{j,k}\|_\infty^2 \leq K' j^{\gamma 2l} 2^{j_1} \leq K' \log^{\gamma 2l} n * n / \log^{2\gamma l} n \leq K' n.$$

We thus obtain that $\delta^{2l}\|\psi_{j,k}\|_\infty^2 \leq K'n$ and consequently Theorem 2 can be applied with this choice of δ. It leads to the following bound

$$\mathbb{P}(|\hat{\beta}_{j,k} - \beta_{j,k}| > \lambda_j/2) \leq e^{-CK^{1/\gamma} j}. \tag{5.2}$$

Thus, there exists an increasing function of K denoted $\nu(K)$ such that

$$\mathbb{P}(|\hat{\beta}_{j,k} - \beta_{j,k}| > \lambda_j/2) \leq 2^{-j\nu}.$$

Added to the fact that we control $\sum_k |\beta_{j,k}|^p 2^{jp(s'+1/2-1/p)}$ since the assumptions in Theorem 3 imply that f belongs to a Besov space $\mathcal{B}_{p,\infty}^{s'}$ with $s' = s - 1/\pi + 1/p$, we obtain for $C > 0$,

$$T_{23} \leq C \sum_{j_0}^{j_1} 2^{-j(s'p+\nu)}.$$

It follows that T_{23} has the rate $2^{-j_0(s'p+\nu)} \leq n^{-(s'p+\nu)/(1+N)}$.
Taking K sufficiently large, we can control as accurately as we want the parameter ν and then T_{23} becomes asymptotically negligible compared to the other terms.

5.5.2 Term T_{21}

We first introduce $\mathbb{P}(|\hat{\beta}_{j,k} - \beta_{j,k}| > \lambda_j/2)$ as we did it for the term T_{23}. By the Cauchy-Shwarz inequality there exists $C > 0$

$$T_{21} \leq C \sum_{j_0}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} \left[\mathbb{E}|\hat{\beta}_{j,k} - \beta_{j,k}|^{2p}\right]^{1/2} \mathbb{P}(|\hat{\beta}_{j,k} - \beta_{j,k}| > \lambda_j/2)^{1/2}.$$

We achieve the following rate for T_{21} using (5.2)

$$\sum_{j_0}^{j_1} 2^{j(p/2-1-\nu/2)} \sum_{k=0}^{2^j-1} \left[\mathbb{E}|\hat{\beta}_{j,k} - \beta_{j,k}|^{2p}\right]^{1/2}.$$

Inequality (5.1) leads to the rate $n^{-p/2} \sum_{j_0}^{j_1} 2^{jp/2-\nu/2}$. As for T_{23}, we may choose the constant K large enough to get $\nu \geq p$. Then, we can write

$$T_{21} \leq C\, n^{-p/2} 2^{j_1(p-\nu)/2}.$$

Here again, we can choose the constant K sufficiently large, in order to achieve a sufficiently large ν such that T_{21} becomes negligible compared to the other terms.

5.5.3 Term T_{24}

This term corresponds to the leading one, meaning it is the one which determines the convergence rate. As

$$\mathbb{1}_{\hat{B}_j^C \cap B_j^{+C}} \leq \mathbb{1}_{|\beta_{j,k}| \leq 2\lambda_j},$$

there exists $C > 0$ such that

$$T_{24} \leq C \sum_{j=j_0}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \mathbb{1}_{\{|\beta_{j,k}| \leq 2\lambda_j\}}. \tag{5.3}$$

We now must distinguish on the values of ϵ.

- If $\epsilon > 0$, we introduce j_{0+} as the largest integer such that

$$2^{j_{0+}} \leq \left(\frac{n}{\log^{2\gamma} n}\right)^{\frac{1}{1+2s}}. \tag{5.4}$$

  The hypothesis $s \leq N/2$ implies $j_{0+} \geq j_0$. We decompose the inequality (5.3) as follows

$$T_{24} \leq \underbrace{\sum_{j=j_0}^{j_{0+}} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \mathbb{1}_{\{|\beta_{j,k}| \leq 2\lambda_j\}}}_{T_{241+}} + \underbrace{\sum_{j=j_{0+}}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \mathbb{1}_{\{|\beta_{j,k}| \leq 2\lambda_j\}}}_{T_{242+}}.$$

  We study separately the behaviours of T_{241+} and T_{242+}.

  Term T_{241+}. This term satisfy the following inequality $T_{241+} \leq \sum_{j=j_0}^{j_{0+}} 2^{j(p/2-1)} 2^j (2\lambda_j)^p$. With λ_j given by (3.6), we have $T_{241+} \leq n^{-p/2} 2^{j_{0+}p/2} (j_{0+})^{\gamma p}$. The choice of j_{0+} in (5.4) gives $T_{241+} \leq C(\log^{2\gamma} n/n)^{\alpha p}$, with C positive constant.

  Term T_{242+}. As $\pi - p < 0$, we have the inequality

$$\mathbb{1}_{|\beta_{j,k}| \leq 2\lambda_j} \leq \left|\frac{\beta_{j,k}}{2\lambda_j}\right|^{(\pi-p)}. \tag{5.5}$$

  We thus can easily bound T_{242+} by

$$\sum_{j=j_{0+}}^{j_1} 2^{j(p/2-1)} \lambda_j^{p-\pi} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^\pi. \tag{5.6}$$

  Replacing λ_j by its value,

$$T_{242+} \leq C \left(\frac{j_1^{2\gamma}}{n}\right)^{(p-\pi)/2} \sum_{j=j_{0+}}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^\pi. \tag{5.7}$$

  Using the control of $\sum_k |\beta_{j,k}|^\pi 2^{j\pi(s+1/2-1/\pi)}$ given by the inclusion $\mathcal{B}^s_{\pi,r} \subset \mathcal{B}^s_{\pi\infty}$

$$T_{242+} \leq C \sum_{j=j_{0+}}^{j_1} 2^{jp/2} \lambda_j^{p-\pi} 2^{-j(s+1/2)\pi} \|f\|^\pi_{s,\pi,\infty} \leq C \left(\frac{j_1^{2\gamma}}{n}\right)^{(p-\pi)/2} \sum_{j=j_0}^{j_1} 2^{-j\epsilon}. \tag{5.8}$$

Let us define $v_n^+ := n^{-(p-\pi)/2} \sum_{j=j_{0+}}^{j_1} 2^{-j\epsilon}$.
As $\epsilon > 0$, v_n^+ is bounded by $n^{-(p-\pi)/2} 2^{-j_{0+}\epsilon}$. So, it gives with j_{0+} as in (5.4): $v_n^+ = (\log n)^{2\gamma\epsilon/(1+2s)} n^{-\epsilon/(1+2s)-(p-\pi)/2}$. The equality $\epsilon/(1+2s)+(p-\pi)/2 = \alpha p$ if $\epsilon \geq 0$ implies obtain that v_n^+ is equal to $(\log^{-2\gamma}(n)n)^{-\alpha p}(\log^{\gamma(p-\pi)}(n)$.
Finally, we have

$$T_{242+} \leq C\,(\log^{2\gamma} n/n)^{\alpha p},$$

which is the rate we are looking for.

To conclude, when $\epsilon > 0$, we obtain

$$T_{24} \leq C\,(\log^{2\gamma} n/n)^{\alpha p}.$$

- If $\epsilon < 0$, we introduce j_{1-} as the largest integer satisfying

$$2^{j_{1-}} \leq \left(\frac{n}{\log^{2\gamma} n}\right)^{\frac{\alpha}{s'}}. \tag{5.9}$$

When $\epsilon < 0$, $\alpha < s'$ and so $j_{1-} \leq j_1$ for a sufficient n. We decompose the inequality (5.3) as follows

$$T_{24} \leq \underbrace{\sum_{j=j_0}^{j_{1-}} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \mathbb{1}_{\{|\beta_{j,k}|\leq 2\lambda_j\}}}_{T_{241-}} + \underbrace{\sum_{j=j_{1-}}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \mathbb{1}_{\{|\beta_{j,k}|\leq 2\lambda_j\}}}_{T_{242-}}.$$

We consider separately T_{241-} and T_{242-}.

Term T_{241-}. Following the scheme used for T_{242+}, the inequality (5.8) becomes

$$T_{241-} \leq C \sum_{j=j_0}^{j_{1-}} 2^{jp/2} \lambda_j^{p-\pi} 2^{-j(s+1/2)\pi} \|f\|_{s,\pi,\infty}^{\pi} \leq C \left(\frac{j_1^{2\gamma}}{n}\right)^{(p-\pi)/2} \sum_{j=j_0}^{j_{1-}} 2^{-j\epsilon}. \tag{5.10}$$

Let us define $v_n^- = n^{-(p-\pi)/2} \sum_{j=j_0}^{j_{1-}} 2^{-j\epsilon}$.
As $\epsilon < 0$, v_n^- is bounded by $n^{-(p-\pi)/2} 2^{-j_{1-}\epsilon}$. Thanks to the choice of j_{1-} this bound is equal to $(\log^{2\gamma l} n)^{\epsilon} n^{-\epsilon\alpha/s'-(p-\pi)/2}$. Like above, we can notice that if $\epsilon \leq 0$, we have the equality $\epsilon\alpha/s' + (p-\pi)/2 = \alpha p$. Then $v_n^- = (\log(n)^{2\gamma})^{\epsilon\alpha/s'} n^{-\alpha p} = (\log(n)^{2\gamma})^{\alpha p-(p-\pi)/2} n^{-\alpha p}$. Together with (5.10), we obtain

$$T_{241-} \leq C\,(\log^{2\gamma} n/n)^{\alpha p}.$$

Term T_{242-}. If h is defined as

$$h = \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \left(\beta_{j,k} \mathbb{1}_{|\beta_{jk}|\leq 2\lambda_j}\right) \psi_{j,k},$$

then h is a fonction belonging to the Besov ball $\mathcal{B}_{\pi,r}^s$. Then we can write T_{242-} as $T_{242-} = \mathbb{E}\left[\left\|h - \hat{h}\right\|_p^p\right]$ with

$$\hat{h} = \sum_{j=j_0}^{j_{1-}} \sum_{k=0}^{2^j-1} \left(\beta_{j,k} \mathbb{1}_{|\beta_{jk}|\leq 2\lambda_j}\right) \psi_{j,k}.$$

This remark allows us to apply the same bound as for the term T_3:

$$T_{242-} \leq C\,(2^{-j_1 - s'})^p.$$

The choice of j_{1-} in (5.9) gives the exact bound $(\log^{2\gamma} n/n)^{\alpha p}$.

Combining the bounds of T_{241-} and T_{242-}, we have:

$$T_{24} \leq (\log^{2\gamma} n/n)^{\alpha p},$$

when $\epsilon < 0$.

- We finally consider the case $\epsilon = 0$. First, if $\epsilon = 0$, inequality (5.7) becomes

$$T_{24} \leq C \left(\frac{j_1^{2\gamma}}{n}\right)^{(p-\pi)/2} \sum_{j=j_0}^{j_1} 2^{j(s+1/2-1/\pi)\pi} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^\pi.$$

  - Noting that $\mathcal{B}^s_{\pi,r}$ is included in $\mathcal{B}^s_{\pi,r'}$ for all $r' \geq r$, we first can control the term T_{24} by $\|f\|^\pi_{s,\pi,\pi}$ if $\pi \geq r$.
  - When $\pi < r$, then we can apply Hölder inequality to obtain $\|f\|_{s,\pi,r}$. Let for every integer j

$$t_j = 2^{j(s+1/2-1/\pi)\pi} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^\pi.$$

  With this notation and using Hölder inequality:

$$T_{24} \leq C \left(\frac{j_1^{2\gamma}}{n}\right)^{(p-\pi)/2} \sum_{j=j_0}^{j_1} t_j^\pi \leq C \left(\frac{j_1^{2\gamma}}{n}\right)^{(p-\pi)/2} \left(\sum_{j=j_0}^{j_1} t_j^r\right)^{\pi/r} \left(\sum_{j=j_0}^{j_1} t_j^{r'}\right)^{\pi/r'},$$

  with $1/r + 1/r' = 1/\pi$. As f belongs to $\mathcal{B}^s_{\pi,r}$ which is included in $\mathcal{B}^s_{\pi,\infty}$, we have:

$$T_{24} \leq C \left(\frac{j_1^{2\gamma}}{n}\right)^{(p-\pi)/2} \|f\|^\pi_{s,\pi,r} \left((j_1 - j_0)\|f\|^{r'}_{s,\pi,\infty}\right)^{\pi/r'} \leq C \left(\frac{j_1^{2\gamma}}{n}\right)^{(p-\pi)/2} j_1^{(1-\pi/r)}.$$

To conclude, for a suitable constant $C > 0$

$$T_{24} \leq C \left(\frac{\log^{-2\gamma} n}{n}\right)^{-\alpha p} \begin{cases} 1 & \text{if } \epsilon \neq 0, \\ (\log n)^{(1-\pi/r)_+} & \text{if } \epsilon = 0. \end{cases} \tag{5.11}$$

Actually, this rate is the one given in Theorem 3.

5.5.4 Term T_{22}

The scheme of the proof of the convergence of this term is very similar to the term T_{24}.

- If $\epsilon > 0$, we introduce j_{0+} like in (5.4) and decompose T_{22} as follows:

$$T_{22} = \underbrace{\sum_{j=j_0}^{j_{0+}} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \mathbb{1}_{\{|\beta_{j,k}| \geq 2\lambda_j\}}}_{T_{221+}} + \underbrace{\sum_{j=j_{0+}}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \mathbb{1}_{\{|\beta_{j,k}| \geq 2\lambda_j\}}}_{T_{222+}}.$$

Term T_{221+}**.** As told in Subsection 5.2, we bound $\mathbb{E}|\hat{\beta}_{j,k}-\beta_{j,k}|^p$ by $n^{-p/2}$ uniformly in j and k. We can deduce

$$T_{221+} \leq Cn^{-p/2}\sum_{j_0}^{j_{0+}} 2^{jp/2} \leq Cn^{-p/2}2^{j_{0+}p/2}.$$

Replacing j_{0+} by its value, we see that this rate is smaller than $(\log^{2\gamma} n/n)^{\alpha p}$.

Term T_{222+}**.** Using the same method as in (5.5), we note that

$$\mathbb{1}_{\{|\beta_{j,k}|\geq \lambda_j/2\}} \leq \left|\frac{2\beta_{j,k}}{\lambda_j}\right|^{\pi}. \tag{5.12}$$

Recall the inequality : $\mathbb{E}|\hat{\beta}_{j,k}-\beta_{j,k}|^p \leq n^{-p/2}$. There exists $C>0$ such that

$$T_{222+} \leq C\, n^{-p/2}\sum_{j=j_0}^{j_1} 2^{j(p/2-1)}\sum_{k=0}^{2^j-1}\left|\frac{\beta_{j,k}}{\lambda_j}\right|^{\pi}.$$

Replacing λ_j by its value the rate becomes

$$\frac{j_0^{-\gamma\pi}}{n^{(p-\pi)/2}}\sum_{j=j_{0+}}^{j_1} 2^{j(p/2-1)}\sum_{k=0}^{2^j-1}|\beta_{j,k}|^{\pi}\,. \tag{5.13}$$

At this step, we recognize a bound of the same form than the one obtained in inequality (5.8) for T_{242+}. Actually, for $C>0$

$$T_{222+} \leq C\, j_{0+}^{-\gamma\pi}\left(n^{-(p-\pi)/2}\sum_{j=j_{0+}}^{j_1} 2^{j(p/2-1)}\sum_{k=0}^{2^j-1}|\beta_{j,k}|^{\pi}\right).$$

Considering the proof for the term T_{242+} this leads to

$$T_{222+} \leq C\, j_{0+}{}^{-\gamma\pi}(\log^{2\gamma} n/n)^{\alpha p},$$

which converges faster than $(\log^{2\gamma} n/n)^{\alpha p}$.

- When $\epsilon<0$, we decompose as follows:

$$T_{24} \leq \underbrace{\sum_{j=j_0}^{j_{1-}} 2^{j(p/2-1)}\sum_{k=0}^{2^j-1}|\beta_{j,k}|^p\mathbb{1}_{\{|\beta_{j,k}|\geq 2\lambda_j\}}}_{T_{241-}} + \underbrace{\sum_{j=j_{1-}}^{j_1} 2^{j(p/2-1)}\sum_{k=0}^{2^j-1}|\beta_{j,k}|^p\mathbb{1}_{\{|\beta_{j,k}|\geq 2\lambda_j\}}}_{T_{242-}},$$

with j_{1-} defined in (5.9). We then consider the terms separately.

Term T_{221-}**.** Exactly like for T_{222+}, we can prove the inequality

$$T_{222+} \leq C\, j_0^{-\gamma\pi}\left(n^{-(p-\pi)/2}\sum_{j=j_0}^{j_{1-}} 2^{j(p/2-1)}\sum_{k=0}^{2^j-1}|\beta_{j,k}|^{\pi}\right),$$

Applying the same developments than for T_{241-} we then obtain $T_{221-} \leq j_0^{-\gamma\pi}(\log^{2\gamma} n/n)^{\alpha p}$ which proves that T_{221-} converges to 0 with a better rate than the one wanted.

Term T_{222-}. Exactly like for the term T_{242-}, this term can be seen as the L^p-risk of an estimator and it can be bounded like T_3. We obtain the same rate than for T_{242-}.

- If $\epsilon = 0$, inequality (5.12) leads to an inequality of the form (5.13)

$$T_{22} \leq j_0^{-\gamma\pi} n^{-(p-\pi)/2} \sum_{j_0}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p.$$

When $\epsilon = 0$, we have $p/2 - 1 = s + 1/2 - 1/\pi$. Consequently, using the Sobolev inclusion $\mathcal{B}^s_{\pi,r} \subset \mathcal{B}^s_{\pi,\infty}$, we have:

$$T_{22} \leq j_1 j_0^{-\gamma\pi} n^{-(p-\pi)/2} \|f\|_{s,\pi,\infty}.$$

We can notice $(p-\pi)/2 = \alpha p$ because of the condition $\epsilon = 0$ and thus

$$T_{22} \leq C \log(n)^{2\gamma\alpha p - (\gamma(\pi+2\alpha p)-1)} n^{-\alpha p}.$$

It is then sufficient to prove $(\gamma(\pi - 2\alpha p) - 1) \geq 0$. Replacing αp by $(p-\pi)/2$, we have $\gamma(\pi - 2\alpha p) - 1 = \gamma p - 1$. As $p/2 - 1 = s + 1/2 - 1/\pi > 0$, we have $p > 2$, and γ is always greater or equal to $1/2$. Consequently, $\gamma p - 1$ is always positive, which achieves the proof. $\square$

5.6 Proof of inequality (3.3)

We only consider the case of expanding maps and $l = 5$. The φ-dependence was introduced by Dedecker and Prieur (2005) in [7]. It seems well-adapted to study Lasota-Yorke functions.

Definition 3 (Dedecker and Prieur (2005)). *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $\mathcal{M}$ a σ-algebra of $\mathcal{A}$. For any random variable $X \in \mathbb{R}^d$ we define:*

$$\varphi(\mathcal{M}, X) = \sup\left\{\|\mathbb{E}(g(X)|\mathcal{M}) - \mathbb{E}(g(X))\|_\infty,\ g \in BV_1\right\},$$

where, in our case, the coefficients $(\varphi_r)_{r\in\mathbb{N}}$ are defined by

$$\varphi_r = \sup_{i+r\leq s} \left\{\varphi(\sigma(\{X_j/j \geq i\}), X_s)\right\}.$$

The process is φ-dependent if φ_r tends to 0 as r tends to infinity.

Dedecker and Prieur (2005) prove in [7] that expanding maps are geometrically φ-weakly dependent, with $\varphi_r = \mathcal{O}(\exp(-ar))$ with $a > 0$. In order to find our probability inequality we use the following Bernstein's inequality from Doukhan and Neumann (2006), [12], for weakly dependent random variables.

Theorem 4 (Doukhan and Neumann, 2006). *Suppose that $Y_1, \ldots, Y_n$ are real-valued random variables with $E[Y_i] = 0$ and $P(|Y_i| \leq M) = 1$, for all $i = 1, \ldots, n$ and some $M < \infty$. We assume that there exist constants $K < \infty$, $\mu \geq 0$ and a non increasing sequence of real coefficients $(\rho_n)_{n\in\mathbb{N}_0}$ such that, for all u-tuples $(s_1, \ldots, s_u)$ and all v-tuples $(t_1, \ldots, t_v)$ with $1 \leq s_1 \leq \cdots \leq s_u \leq t_1 \leq \cdots \leq t_v \leq n$ the following inequality is fulfilled:*

$$|Cov(Y_{s_1} \cdots Y_{s_u}, Y_{t_1} \cdots Y_{t_v})| \leq u\, K^2\, M^{u+v-2}\, \rho_{t_1-s_u}, \tag{5.14}$$

and

$$\sum_{s=0}^{\infty}(s+1)^{k-2}\rho_s \;\le\; (k!)^{\mu}. \tag{5.15}$$

Then

$$P\left(\sum_{i=1}^{n} Y_i \;\ge\; \lambda\right) \;\le\; \exp\left(-\frac{\lambda^2/2}{\sigma_n^2 \;+\; B_n^{1/(\mu+2)}\lambda^{(2\mu+3)/(\mu+2)}}\right),$$

where $B_n \;=\; 2\,(K \vee M)\left(\frac{2^{4+\mu}\,n\,K^2}{\sigma_n^2} \;\vee\; 1\right)$ *and* $\sigma_n^2 \;=\; Var(\sum_{i=1}^{n} Y_i)$.

We apply this inequality to $Y_i = \psi_{j,k}(X_i) - \mathbb{E}\psi_{j,k}(X_i)$ for all $i \in \mathbb{Z}$, with $j \le j_1$ and $K \in \mathbb{Z}$. It is easy to check (see for instance Barbour *et al.* (2000), [3]) that $(Y_0, Y_1, \dots, Y_n)$ has the same distribution as $(Z_n, Z_{n-1}, \dots, Z_0)$ where $(Z_t)_{t\ge 0}$ is a stationary Markov Chain. Then, following Dedecker and Prieur (2005), [7], we can rewrite the covariance term in order to use the definition of the φ-dependence:

$$\begin{aligned}
|\mathrm{Cov}\,(Y_{s_1}\cdots Y_{s_u}, Y_{t_1}\cdots Y_{t_v})| \;&\le\; \mathbb{E}|(\mathbb{E}(Y_{s_1}\cdots Y_{s_u}) - \mathbb{E}(Y_{s_1}\cdots Y_{s_u}|Y_{t_1}\cdots Y_{t_v}))Y_{t_1}\cdots Y_{t_v}|,\\
&\le\; \|\mathbb{E}(Y_{s_1}\cdots Y_{s_u}) - \mathbb{E}(Y_{s_1}\cdots Y_{s_u}|Y_{t_1})\|_\infty \mathbb{E}|Y_{t_1}\cdots Y_{t_v}|,
\end{aligned}$$

Using once more the Markov property of $(Y_0, Y_1, \dots, Y_n)$ under reverse time, we have:

$$\begin{aligned}
&\|\mathbb{E}(Y_{s_1}\cdots Y_{s_u}) - \mathbb{E}(Y_{s_1}\cdots Y_{s_u}|Y_{t_1})\|_\infty\\
&\qquad = \|\mathbb{E}(\mathbb{E}(Y_{s_1}\cdots Y_{s_u}|Y_{s_u}) - \mathbb{E}(\mathbb{E}(Y_{s_1}\cdots Y_{s_u}|Y_{s_u})|Y_{t_1})\|_\infty,\\
&\qquad\qquad = \|\mathbb{E}(g(X_{s_u})) - \mathbb{E}(g(X_{s_u})|X_{t_1})\|_\infty,
\end{aligned}$$

with the following functions g_k:

$$g_k:\; x \mapsto \mathbb{E}((\psi_{j,k}(X_{s_1}) - \mathbb{E}(\psi_{j,k}(X_0)))\cdots(\psi_{j,k}(X_{s_u}) - \mathbb{E}(\psi_{j,k}(X_0)))|X_{s_u} = x).$$

In order to apply the definition of φ-dependence, it remains to compute $\|g_k\|_{BV}$. Reminding that $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$ is null outside $[2^{-j}k; 2^{-j}(k+1)]$ for all $0 \le k \le 2^j - 1$, then the same is true for g_k. Thus $h = \sum_{k=0}^{2^j-1} g_k$ is a function defined on $[0,1]$. Applying the result in higher dimension of Collet *et al.* (2002), [4], h has bounded variation. More precisely there exists $K > 0$ and $0 \le \sigma \le 1$ such that $\|h\|_{BV} \le K\sum_{i=0}^{u}\sigma^i L_{i+1}$. The coefficients $L_1, \dots, L_u$ satisfy for $(x_1, \dots, x_u, y_1, \dots, y_u) \in [0,1]^{2u}$:

$$|h(x_1, \dots, x_u) - h(y_1, \dots, y_u)| \le L_1|x_1 - y_1| + \cdots + L_u|x_u - y_u|.$$

A simple computation leads to:

$$\|h\|_{BV} \le K\mathrm{Lip}\,(g_k) \le Ku2^{ju/2+1}\|\psi\|_\infty^{u-1}\mathrm{Lip}\,\psi.$$

Recalling that the functions g_k for $k = 1, \dots, 2^j - 1$ have distinct domains, $\|h\|_{BV} = \sum_{k=0}^{2^j-1}\|g_k\|_{BV}$. Note that all g_k are functions of dynamical systems with same characteristics; their bounds on variation norms are equal to $\|g_k\|_{BV} \le u2^{ju/2}\|\psi\|_\infty^{u-1}\mathrm{Lip}\,\psi$. We thence obtain a bound on the covariance

$$\mathbb{E}\,|\psi_{j,k}(Y_{t_1}), \dots, \psi_{j,k}(Y_{t_v})|\;\|g_k\|_{BV}\,\varphi_{t_1-s_u} \;\le\; u2^{\frac{j}{2}u+v-1}\|\psi\|_\infty^{u+v-2}\mathrm{Lip}\,\psi\,\mathbb{E}\,|\psi_{j,k}(X_0)|\;\varphi_{t_1-s_u}.$$

It remains to control the order of $\mathbb{E}\,|\psi_{j,k}(X_0)| = 2^{j/2}\int|\psi(2^jx-k)|f(x)dx$. Classically, posing $u = 2^jx-k$, using the fact that f is uniformly bounded and that ψ is integrable, we obtain $\mathbb{E}\,|\psi_{j,k}(X_0)| \le C2^{-j/2}$ with C depending on $\int|\psi|$ and $\|f\|_\infty$. We apply Theorem 4 with $M = 2^{j/2}\|\psi\|_\infty$, K^2 as a well chosen constant depending on $\mathrm{Lip}\,(\psi)$, $\|\psi\|_\infty$, $\int|\psi|$ and $\|f\|_\infty$ and $\rho = \varphi$.
We now study (5.15) in order to determine μ. Previous results on expanding maps imply that $\phi_r \le \exp(-ar)$ with $a > 0$. Quote that

$$\sum_{r=0}^{n-1}(r+1)^{q-2}\exp(-ar) \quad \le \quad \int_0^\infty r^{q-2}\exp(-ar)dr.$$

Then, the change of variable $u = ar$ gives

$$\sum_{r=0}^{n-1}(r+1)^{q-2}\exp(-ar) \le \frac{1}{a^{q-1}}\int_0^\infty u^{q-2}\exp(-u)du = \frac{1}{a^{q-1}}\Gamma\,(q-1) \le \frac{C}{a^{q-1}}(q-1)!$$

The last inequality follows from Stirling's formula which entails that for any constant $A > 0$, and any $\epsilon > 0$ there exists $B_\epsilon > 0$ such that $A^k \le B_\epsilon k!^\epsilon$. Thus, under this rate of dependence, assumptions of Theorem 4 hold with $\mu = 1$. It is easy to check that the order of B_n is smaller than $2^{j/2} \propto \|\psi_{j,k}\|_\infty$ and Theorem 4 can be applied. $\square$

References

[1] Andrews, D. Non strong mixing autoregressive processes. *Journal of Applied Probability 21* (1984), 930–934.

[2] Antoniadis, A., and Fan, J. Regularization of wavelet approximations. *Journal of American Statistical Association 96*, 455 (2001), 939–967.

[3] Barbour, A., Gerrard, R., and Reinert, G. Iterates of expanding maps. *Probability Theory and Related Fields 116* (2000), 151–180.

[4] Collet, P., Martnez, S., and Schmitt, B. Exponential inequalities for dynamical measures of expanding maps of the interval. *Probability Theory and Related Fields 123* (2002), 301–322.

[5] Comte, F., and Merlevède, F. Adaptive estimation of the stationary density of discrete and continuous time mixing processes. *ESAIM P&S 6* (2002), 211–238.

[6] Daubechies, I. *Ten lectures on wavelets*, vol. 61. SIAM press, 1992.

[7] Dedecker, J., and Prieur, C. New dependence coefficients: Examples and applications to statistics. *Probability Theory and Related Fields 132* (2005), 203–235.

[8] Donoho, D., Johnstone, I., Kerkyacharian, G., and Picard, D. Density estimation by wavelet thresholding. *Annals of Statistics 24*, 2 (1996), 508–539.

[9] Donoho, D. L., and Johnstone, I. A. Adapting to unknown smoothness via wavelet shrinkage. *Journal of American Statistical Association 90*, 432 (1995), 1200–1224.

[10] Doukhan, P., and Lang, G. Rates of convergence in the weak invariance principle for the empirical repartition process of weakly dependent sequences. *Statistical Inference for Stochastic Processes 5* (2002), 199–228.

[11] Doukhan, P., and Louhichi, S. A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and their Applications 84* (1999), 313–342.

[12] Doukhan, P., and Neumann, M. A Bernstein type inequality for times series. unpublished, available at http://www.crest.fr/pageperso/doukhan/publications.htm, 2006.

[13] Doukhan, P., Teyssière, G., and P., W. Vector valued arch infinity processes. *Dependence in Probability and Statistics* (2006).

[14] Guyon, X. *Random fields on a network: modeling, statistics, and applications.* New York: Springer-Verlag, 1995.

[15] Hall, P., Lahiri, S. N., and K., T. Y. On bandwidth choice for density estimation with dependent data. *Annals of Statistics 23*, 6 (1995), 2241–2263.

[16] Hall, P., and Patil, P. Formulae for mean integrated squared error of nonlinear wavelet-based density estimators. *Annals of Statistics 23*, 3 (1995), 905–928.

[17] Hardle, W., Kerkyacharian, G., Picard, D., and Tsybakov, A. *Wavelets Approximation and Statistical Applications*, vol. 129. Springer Verlag, Lecture Notes in Statistics, 1998.

[18] Hart, J. D., and Vieu, P. Data-driven bandwidth choice for density estimation based on dependent data. *Annals of Statistics 18*, 2 (1990), 873–890.

[19] Johnstone, I., Kerkacharyan, G., Picard, D., and M., R. Wavelet deconvolution in a periodic setting. *Journal of Royal Society Society 3*, 66 (2004), 1–27.

[20] Kerkyacharian, G., and Picard, D. Thresholding algorithms, maxisets and well-concentrated bases. *Test 9* (2000), 283–344.

[21] Lahiri, S. N. *Resampling Methods for Dependent Data.* Springer Series in Statistics, 2003.

[22] Meyer, Y. *Wavelets and operators.* Cambridge University Press, 1992.

[23] Petrov, V. *Limit Theorems of Probability Theory.* Oxford University Press, Oxford (UK), 1995.

[24] Prieur, C. *Applications statistiques de suites faiblement dépendantes et de systèmes dynamiques.* PhD thesis, CREST, 2001.

[25] Ragache, N., and Wintenberger, O. Convergence rates for density estimators of weakly dependent time series. *Dependence in Probability and Statistics* (2006).

[26] Tribouley, K., and Viennet, G. l_p adaptive density estimation in a β mixing framework. *Annales de l'Instut Henri Poincaré (B) Probabilité et Statistiques 34*, 2 (1998), 179–208.

[27] Vannucci, M. Nonparametric density estimation using wavelets. Tech. rep., Texas A and M university, 1998.

[28] Viana, M. Stochastic dynamics of deterministic systems. available at http://w3.impa.br/∼viana, 1997.

[29] Wavelab. http://www-stat.stanford.edu/∼wavelab/.