



Adaptative density estimation with dependent observations

Irène Gannaz, Olivier Wintenberger

► To cite this version:

Irène Gannaz, Olivier Wintenberger. Adaptative density estimation with dependent observations. 2005. hal-00012077v1

HAL Id: hal-00012077

<https://hal.science/hal-00012077v1>

Preprint submitted on 14 Oct 2005 (v1), last revised 15 Oct 2008 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Adaptative density estimation with dependent observations

Gannaz Irène¹, Wintenberger Olivier²

14th October 2005

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 MATISSE-SAMOS, Université Paris I Panthéon Sorbonne, 90 rue de Tolbiac, 75 634 Paris Cedex 13, France. E-mail: olivier.wintenberger@univ-paris1.fr

Abstract

Assume that $(X_n)_{n \in \mathbb{Z}}$ is a real valued stationary time series admitting a common density f . To estimate f in an independent and identically distributed setting, Donoho, Johnstone, Kerkycharian & Picard (1996) proposed a quasi-minimax method based on thresholding wavelets. The aim of the present work is to extend this methodology to the dependent case. For this purpose, we introduce the new Φ -weak dependence based on a probability inequality, which includes a large spectrum of classical weak dependence cases. Actually, we establish a link between this condition and the $\tilde{\phi}$ -dependence of Dedecker & Prieur (2004) and the η -weak dependence Φ condition introduced by Doukhan & Louhichi (1999). The estimator we propose adapts the threshold to the dependence of the observations. We obtain near minimax convergence rates for L^p losses, $p \geq 1$. We thus apply this method on simulations of non stationary but geometrically ergodic cases like dynamical systems and Markovian fields on the line.

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

1 Introduction

Density estimation is a classical topic in nonparametric theory. Let $(X_n)_{n \in \mathbb{Z}}$ be a real valued time series with a common density f . The purpose is to estimate this probability density function with an optimal rate of convergence, without assumption on f . Among the classical tools in density estimation (see Devroye (1985)), we are interested in wavelets estimation. We will not develop the theory of wavelets in this article, and we refer to Daubechies (1992) and Meyer (1992) for general theory on wavelets, or to Härdle et al. (1998) and Antoniadis (1997) for their application in statistics; Concerning their use in density estimation, Vannucci (1998) propose a brief survey. The use of wavelets in density estimation was introduced by Doukhan & León (1990), soon followed by Kerkyacharian & Picard (1992) and Walter (1992). The first results established concerned linear wavelet estimators. Among others, we can cite Walter (1994) who studied continuous densities estimation, Kerkyacharian & Picard (1993) who considered densities f belonging to a Besov space, or Antoniadis & Carmona (2001)... Yet, relaxing linearity allows to obtain better rates of convergence. Consequently, Donoho et al. (1996) introduced nonlinearity within the estimator by the means of a threshold on wavelet coefficients. Many kinds of thresholding exist, such as Hall & Patil's (1995) or those presented in Antoniadis & Fan (2001), but we restrict ourselves to hard-thresholding.

The purpose of the present work is to consider the density estimation of weakly dependent observations. Dealing with times series provide a huge amount of applications. Several ways of modeling the weak dependence have already been worked out. One of the most popular is the notion of mixing, see Doukhan (1994) for bibliography. Adaptative wavelet estimation in such a setting has been explored by Zanten & Zareba (1999). However, mixing condition presents lots of restrictions. For example, Andrews (1984) exhibits the simple counter-example of an auto-regressive process which does not satisfy a mixing condition. Doukhan and Louhichi (1999) introduced several new weak dependence conditions in order to extend dependence notion to such cases. In particular, we will consider the causal $\tilde{\phi}$ -dependence condition as in Dedecker & Prieur (2004) and the non causal η -condition of Doukhan & Louhichi (1999) (see Section 3 for the definition of those conditions). In linear estimation case, the causal $\tilde{\phi}$ -dependence condition leads to better rates than the non causal η -condition (see Doukhan & louhichi (1999), Prieur (2001), Ango Nze, Buhlman & Doukhan (2002) and Ragache & Wintenberger (2005)). In this paper, we will introduce non linearity, by generalizing to both cases the result of Donoho, Johnstone, Kerkyacharian & Picard (1996), which deals with independent observations.

Let us recall the result of Donoho et al. (1996). Assume that the common density f has a compact support and belongs to a Besov Space $B_{\pi,r}^s$. We do not know a priori the regularity of f , and we consider the adaptative thresholding wavelets estimator \hat{f}_n of Donoho et al. (1996) (see Section 2.2 for more details). When the observations are i.i.d., the minimax theory of density estimation proves that \hat{f}_n has the optimal convergence \mathbb{L}^p -rates up to a logarithmic term:

Theorem 1 (Donoho, Johnstone, Kerkyacharian & Picard (1996)). *Suppose the observations X_1, \dots, X_n are independent and identically distributed (i.i.d. for short). If*

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

where N is the regularity of the wavelet, then

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

where $\|f\|_p^p = \int f(x)^p dx$, $\epsilon = s\pi - (p - \pi)/2$ and α is the minimax rate:

$$\alpha = \begin{cases} s/(1 + 2s) & \epsilon \geq 0 \\ (s - 1/\pi + 1/p)/(1 + 2s - 2/\pi) & \epsilon \leq 0. \end{cases} \quad (1)$$

Our method of estimation is based on Donoho et al.'s procedure, but we consider threshold levels which integrates the dependence of the observations. In the case of causal weak dependence, we achieve the same rate than in a i.i.d. setting whereas for non causal dependent observations a logarithmic loss appears. We thus have proposed an adaptative near-minimax density estimator for dependent series. This result is innovative because, to our knowledge, no similar estimation exists for weakly dependent data.

The paper is structured as follows. In the next section, we will present the context of estimation, defining the dependence notion used and the estimator studied. For this purpose, we will first introduce a new Φ -weak dependence condition which deals with exponential probability inequalities similar to Bernstein's. Next, we will define our hard-threshold wavelet estimator. The last part of the Section 2 gives hypotheses and results. In Section 3, we will survey different applications of the main result and apply the estimator in simulations to understand its adaptability to the dependence of the observations. To this end, we study first independent setting, next stationary and non stationary (but geometrically ergodic) causal cases like dynamical systems and finally non causal observations through the example of Markovian fields on the line. The proof of our main Theorem is given in the last Section.

2 Main results

In this section, we introduce a new weak dependence condition. The objective is to cover various notions of weak-dependence in a unique definition, which deals with functionals of dependent sequences. We then recall the nonlinear wavelet estimator considered in Donoho et al. (1996) and explain how it will be modified in order to take into account the dependence of the observations. We finally establish the main result, showing the rate of convergence of this estimator is quasi-minimax.

2.1 Φ -dependence

The notion introduced here defines dependence through a functional probability inequality similar to Bernstein's. It includes various cases of dependence, and gives a control not only on probabilities but also on moments.

Definition 1 (Φ -dependence). Let $(X_i)_{i \in \mathbb{N}}$ be a stationary real valued sequence. Assume that Φ is an increasing function such that

$$\Phi(0) = 0, \quad \lim_{x \rightarrow +\infty} \Phi(x) = +\infty \quad \text{and} \quad \int_0^{+\infty} \exp(-\Phi^{-1}(u)) du < +\infty,$$

where Φ^{-1} is the general inverse of Φ .

The process $(X_i)_{i \in \mathbb{N}}$ is said to be Φ -weakly dependent if, for all integer n ,

$$\sigma_n^2 = \text{Var}(X_1 + \cdots + X_n) \leq Cn,$$

with C a constant, and for all Ψ compactly supported Lipschitz function such that:

$$\mathbb{E}[\Psi(X_0)] = 0 \quad \text{and} \quad \|\Psi\|_\infty \leq \sqrt{n},$$

the dependence on $(X_i)_{i \in \mathbb{N}}$ ensures that we have a moment inequality of the form:

$$\mathbb{P}\left(\left|\sum_{i=1}^n \Psi(X_i)\right| \geq \delta\right) \leq \exp(-\Phi^{-1}(\delta/\sqrt{n})), \quad (2)$$

for all $\delta \geq \sqrt{n}$.

Φ -dependence covers independence and many notions of geometric weak dependence. We will develop in this article some examples in order to show that Φ -dependence is less constraining than mixing. We show in Section 3 that weakly dependent sequences like in Doukhan (1988) and Doukhan (2002) are also Φ -dependent when their coefficients $\tilde{\phi}$ and η are geometrically decreasing.

The last assumption on Φ gives us the rate of all the absolute moments of the sum:

$$\begin{aligned} \mathbb{E}\left|\sum_{i=1}^n \Psi(X_i)\right|^q &= \int_0^{+\infty} \mathbb{P}\left(\left|\sum_{i=1}^n \Psi(X_i)\right| \geq t^{1/q}\right) dt \\ &\leq \int_0^{+\infty} 2 \exp\left(-\Phi^{-1}(t^{1/q}/\sqrt{n})\right) dt \\ &\leq 2\sqrt{n}^q \int_0^{+\infty} \exp(-\Phi^{-1}(u)) du \end{aligned}$$

changing $t = \sqrt{n}^q u$.

Consequently, we obtain

$$\mathbb{E}\left|\sum_{i=1}^n \Psi(X_i)\right|^q \leq n^{-q/2} \quad (3)$$

Actually, Φ -weak dependence also implies a control on the moments of the sum $\sum_{i=1}^n \Psi(X_i)$.

2.2 Construction of the estimation

Let $\bar{\phi}$ generate a multiresolution analysis $\{\bar{\phi}_{j,k}(x) = 2^{j/2} \bar{\phi}(2^j x - k), j \in \mathbb{N}, k \in \mathbb{Z}\}$ on \mathbb{R} and let $\bar{\psi}$ be the wavelet associated with. We consider ϕ and ψ respectively the scale and the wavelet functions built by periodization of $\bar{\phi}$ and $\bar{\psi}$ on $[0; 1]$ (see Meyer (1992) or

Daubechies (1992)). Furthermore, we will assume in the following that the father wavelet ϕ and the mother wavelet ψ are such that:

$$\begin{aligned} \forall k = 0 \dots N \quad \int \phi(x) x^k dx &= \delta_{0,k}, & \int \psi(x) x^k dx &= 0, \\ \text{and} \quad \int |\phi(x) x^{N+1}| dx &< \infty, & \int |\psi(x) x^{N+1}| dx &< \infty, \end{aligned}$$

for a given integer N . Finally, we suppose ϕ and ψ are such that the functions $\sum_k |\phi(\cdot - k)|$ and $\sum_k |\psi(\cdot - k)|$ belong to \mathbf{L}^p space. Daubechies (1992) proposed wavelet bases verifying such conditions, with a number N of vanishing moments as great as wanted. Thus we are assured of the existence of such wavelets and we will consider Daubechies' bases for our simulations.

We introduce the Besov space $B_{\pi,r}^s([0;1])$ on $[0;1]$, for any $s > 0$, $1 \leq \pi, r \leq \infty$. Let j_0 be an arbitrary integer. Any function $f \in \mathcal{B}_{\pi,r}^s([0;1])$ can be decomposed as

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

where $\alpha_{j,k} = \int_0^1 f \phi_{j,k}$ and $\beta_{j,k} = \int_0^1 f \psi_{j,k}$. Moreover, the quantities

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

define a norm on the space $\mathcal{B}_{\pi,r}^s([0;1])$ (see Härdle et al. (1998) for more details on wavelet theory and Besov spaces). We will note $\mathcal{B}_{\pi,r}^s$ the space $\mathcal{B}_{\pi,r}^s([0;1])$ throughout.

A natural estimation of f is then obtained by:

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

where j_0 and j_1 are functions of n and $\hat{\alpha}_{j,k}$ and $\hat{\beta}_{j,k}$ are estimators of the coefficients $\alpha_{j,k}$ and $\beta_{j,k}$. We shall take in the following:

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

for every $(j,k) \in \{j \in \mathbb{N}, k = 0 \dots 2^j - 1\}$. In order to attain the minimax rate of convergence, the optimal choice of j_0 and j_1 depends on the regularity s of the function, as it is detailed in Donoho et al. (1996). Since s is actually unknown, we will adopt a thresholding procedure for estimating f in an adaptative way (see Donoho et al. (1996)).

For any $\lambda > 0$, define γ_λ a hard-threshold function; $\gamma_\lambda(\beta) = \beta \mathbb{1}_{|\beta| > \lambda}$. We will consider the following estimator:

$$\hat{f}_n = \sum_{k=0}^{2_0^j-1} \hat{\alpha}_{j_0,k} \phi_{j_0,k} + \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \gamma_{\lambda_j}(\hat{\beta}_{j,k}) \psi_{j,k}.$$

where j_0 and j_1 and $(\lambda_j)_{j=j_0 \dots j_1}$ will be defined later.

The idea is that the wavelet coefficients' behavior is related to the regularity of the function. The level of thresholding needs to be chosen such as the resulting bias is insignificant, whereas the noise due to the estimation of the coefficients is reduced. Actually, the threshold realizes a balance between bias and variance of the estimation.

The threshold levels let to be chosen in view of this consideration. In case of independent observations Donoho et al. choose the threshold $\lambda_j = K \sqrt{\frac{j}{n}}$ where K is a constant (see Donoho et al. (1996)). Furthermore, they take

$$\begin{aligned} 2^{j_0} &= n^{1/(1+N)} \\ 2^{j_1} &= n/\log(n). \end{aligned}$$

Note that Delyon & Juditsky propose in 1993 different choices for both the sequence j_1 and the threshold λ_j . In our case, we will need to introduce a specific j_1 and specific thresholds in order to take into account the dependence of the observations, because this will interfere on the properties of the coefficients estimators $\hat{\alpha}_{j,k}$ and $\hat{\beta}_{j,k}$.

2.3 Main result

The main result is the following:

Theorem 2. *Let $(X_i)_{i \in \mathbb{Z}}$ be a Φ -weakly dependent time series admitting a common density f with a compact support and belonging to a Besov Space $\mathcal{B}_{\pi,r}^s$. Assume the indexes satisfy the inequalities:*

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

Then we have:

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

with the following choice of j_0 and j_1 :

$$2^{j_0} = n^{1/(1+N)} \tag{4}$$

$$2^{j_1} = \frac{n}{\Phi^2(\log(n))}. \tag{5}$$

and level dependent thresholds:

$$\lambda_j = K \frac{\Phi(j)}{\sqrt{n}} \tag{6}$$

where K is a positive constant and Φ is given by the definition of Φ -weak dependence.

Note that the conditions on s , π , and r are not very restrictive. First, the condition $s \leq N/2$ is easily verified. Actually, we recalled previously that Daubechies (1992) proposed wavelet bases without limitation on the choice of vanishing moments N . We will consider in simulations Daubechies' bases with a fixed number N of vanishing moments, taking N sufficiently large in order to cover each case. Next, the condition $1/\pi < s$ assesses the density f is continuous, which does not seem too constraining. Observe that the assumptions on p are particularly weak: the result holds even with $p \leq 2$.

One important restriction of this result is the hypothesis of a compactly supported density. In order to extend this result to more general functions, one needs Rosenthal's inequality. In the case of independent observations, in 2004, Juditsky & Lambert-Lacroix prove this estimation is quasi-minimax even if the support is not compact. However, dependence prevents us from having such an inequality, and thus we are not able to prove the theorem on the whole real line.

The main idea of this result is to take into account the dependence of the observations through the threshold. Apart this particular choice of the threshold, the estimator is actually the same than in Donoho et al. (1996). The proof of this theorem, presented in the last section is also similar to Donoho et al.'s. It is based on probability and moments inequalities, provided here by the Φ -dependence condition.

The special case $\Phi(x) = C\sqrt{x}$ (with C a constant) corresponds to the independent setting. In this case, the rate obtained is the same than in Donoho et al. (1996) and we are under

the same hypotheses on f (that take in account a large class of density functions). In the cases of $\Phi(x) = x^a$, the estimator is also quasi-minimax, meaning that it achieves the minimax rate of convergence up to a logarithmic term (see Donoho et al. (1996) for a detailed proof of the minimax rate in this problem).

This estimator is adaptative in the sense it does not need any prior information on the regularity of f and nevertheless nearly attains the minimax rate, which depends on this regularity. However, the procedure is not adaptative with the dependence, in the sense it needs the knowledge of the function Φ . This function characterizes the degree of dependence of the observations. Actually, no tool exists to estimate it. In Section 3, we consider $\Phi(x) = x^\gamma$ and we propose to choose the parameter γ by cross-validation. As it will be seen in Section 3, such a procedure can eventually provide a way to determine the kind of dependence followed by the observations. Effectively, the optimal value of γ depends on the kind of dependence followed by the series.

3 Applications and simulations

This section give examples of Φ -dependent sequences. First we study the independent setup, then two different cases of dependent variables. For each case, we have simulated observations in order to show the behavior of the MISE of the estimation.

The simulations are implemented in Matlab using the package Wavelab, available on <http://www-stat.stanford.edu/software/wavelab/index.html>. We are aware of the approximations caused by the implementation of the estimation. Computers in general do not consider continue data but discrete data. This is in particular the case for the storage of wavelets in Wavelab; we are not able to consider really $\psi_{j,k}(X_i)$ as we would want to. The values of wavelet functions in Wavelab are available on the dyadic grids of the interval. So, instead of computing $\psi_{j,k}(X_i)$ we approximate it by $\psi_{j,k}(2^i/n)$ where 2^i is the power of 2 closer to nX_i . This approximation of course introduces a bias in estimation. We have not studied in detail this bias, because the objective here was to illustrate the behavior of the estimator in different cases, and to study the impact of the threshold on the quality of estimation. All the results presented are based on $n = 2^{10}$ observations.

3.1 Independence

At first, we study the case of independent and identically distributed observations. After showing Theorem 2 holds, we will illustrate the estimation procedure with an example.

In order to prove Theorem 2 can be applied for an i.i.d. setup, we must verify Φ -dependence. We have the following result:

Proposition 1. *Every i.i.d. setting is Φ -dependent with $\Phi(x) = C\sqrt{x}$ where C is a constant.*

Proof. We consider i.i.d. observations $(X_i)_{i \in \mathbb{Z}}$. Let Ψ be such that $\mathbb{E}\Psi(X_i) = 0$ and $\|\Psi(X_i)\|_\infty \leq \sqrt{n}$. Denoting $\sigma_n^2 = \text{Var}(\Psi(X_1) + \dots + \Psi(X_n))$, Bernstein's inequality

gives:

$$\mathbb{P} \left(\sum_{i=1}^n \Psi(X_i) \geq \delta \right) \leq \exp \left(- \frac{\delta^2}{2(\sigma_n^2 + \delta\sqrt{n}/3)} \right). \quad (7)$$

By independence, we have $\sigma_n^2 = n \mathbb{E} [\Psi^2(X_1)]$. Then, noting that

$$\mathbb{E} \Psi^2(X_1) \leq \|\Psi\|_\infty \mathbb{E} |\Psi(X_1)| \leq 1,$$

inequality (7) leads to:

$$\mathbb{P} \left(\sum_{i=1}^n \Psi(X_i) \geq \delta \right) \leq \exp \left(- \frac{\delta^2}{2(n + \delta\sqrt{n}/3)} \right). \quad (8)$$

Finally, using the assumption $\delta \leq \sqrt{n}$, we find the bound $\exp(-1/2(\delta/\sqrt{n})^2)$ for the left hand side of (8). Using the symmetry, we recognize Φ -dependence holds with $\Phi^{-1}(t) = 1/2t^2$ or equivalently $\Phi(t) = \sqrt{2t}$. \square

Consequently, the convergence rate in Theorem 2 is achieved, and is similar to Donoho et al. (1996)'s.

Simulations.

Seeing that the Theorem holds for i.i.d. setups, we simulate an example in order to visualize the behavior of the estimation with respect to the level of threshold. We choose to consider $n = 2^{10}$ i.i.d. variables with the cumulative distribution function $F(x) = 2/\pi \sin^{-1}(\sqrt{x})$, defined on the compact set $[0; 1]$. The density we wish to estimate is: $f(x) = 1/(\pi\sqrt{x(1-x)})$. In order to study the role of the threshold in the quality of estimation and, more precisely, how much the choice of the function Φ in the threshold is determinant, we compare the qualities of estimation obtained when we consider $\Phi(x) = x^\gamma$. The theory in Donoho et al. (1996) tells us $\gamma = 1/2$ should be appropriate, as we see on the plot of the MISE of the estimation with respect to γ (see Figure 1).

Remark: In the Figure 1 we display the behavior of the MISE with respect to the parameter γ . We fix the constant in the threshold as $K = 4$ in order to obtain that the minimum is for $\gamma = 1/2$. This threshold constant has a large impact on the behavior of the MISE. When it increases, the curve of the evolution of the MISE is shifted to the right, and then the minimum is no more attained at $\gamma = 1/2$.

3.2 Causal weak dependence

In this section, we consider a weakly dependent real valued sequence $(X_i)_{i \in \mathbb{Z}}$. We suppose moreover that the dependence is causal, meaning an observation X_i is independent from the future $\{X_j, j > i\}$, but depends on the past. We will more precisely consider the case of $\tilde{\phi}$ -dependence, introduced by Dedecker & Prieur (2005), and then apply it to study some dynamical systems.

To this end, we introduce some extra notation: if $h : \mathbb{R}^u \rightarrow \mathbb{R}$ for some $u \in \mathbb{N}^*$, we set

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

A Lipschitz function h is a function such that $\text{Lip}(h) < \infty$.

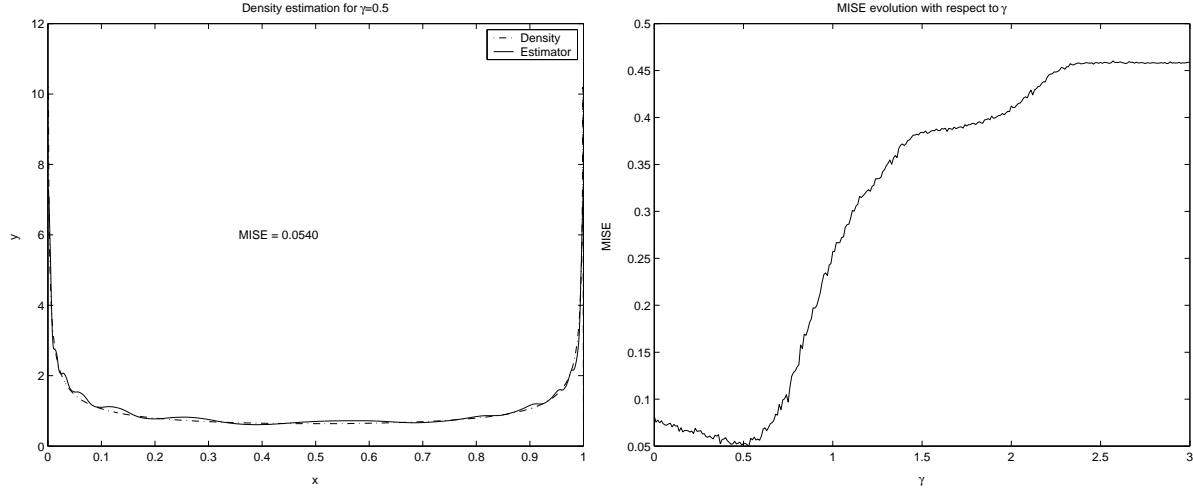


Figure 1: Independent case

3.2.1 $\tilde{\phi}$ -dependence.

Let us first recall the definition of $\tilde{\phi}$ -dependence, which is a causal one:

Definition 2 (Dedecker & 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:

$$\tilde{\phi}(\mathcal{M}, X) = \sup \{ \|\mathbb{E}(g(X)|\mathcal{M}) - \mathbb{E}(g(X))\|_{\infty}, g \in \Lambda_1 \},$$

where $\Lambda_1 = \{h : \mathbb{R}^d \mapsto \mathbb{R}, \text{Lip}(h) < 1\}$. The sequence of coefficients $(\tilde{\phi}_k(r))_{r \in \mathbb{N}}$ is then defined by

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

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

Proposition 2. Let $(X_i)_{i \in \mathbb{Z}}$ be a $\tilde{\phi}$ -weakly dependent sequence.

Suppose the dependence coefficients $(\tilde{\phi}_k(r))_{r \in \mathbb{N}}$ decrease sub-geometrically, i.e. for every r sufficiently large, there exist $a > 0$ and $b > 0$ such that $\tilde{\phi}_r \leq \exp(-ar^b)$ up to a constant. Then (X_i) is Φ -dependent with $\Phi(x) = C\sqrt{x}$, C positive constant.

Proof. We are able to control the covariance of $\tilde{\phi}$ -dependent processes. If $(X_i)_{i \in \mathbb{Z}}$ be a $\tilde{\phi}$ -weakly dependent sequence, for any Lipschitz function k and integrable function h :

$$|\text{Cov}(h(X_{i_1}, \dots, X_{i_u}), k(X_{i_{u+1}}, \dots, X_{i_{u+v}}))| \leq v \mathbb{E}[|h(X_{i_1}, \dots, X_{i_u})|] \text{Lip}(k) \tilde{\phi}(r). \quad (9)$$

Furthermore, Dedecker & Prieur show in Dedecker & Prieur (2005) that the coefficient $\tilde{\phi}(r)$ reaches the least upper bound in (9).

In order to find a probability inequality giving the Φ -dependence under the condition of $\tilde{\phi}$ -dependence, we use the following Bernstein's inequality from Doukhan & Neumann (2005) for weakly dependent random variables.

Theorem 3 (Doukhan & Neumann (2005)). Suppose that X_1, \dots, X_n are real-valued random variables with $E[X_i] = 0$ and $P(|X_i| \leq M) = 1$, for all $i = 1, \dots, n$ 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, \dots, s_u) and all v -tuples (t_1, \dots, t_v) with $1 \leq s_1 \leq \dots \leq s_u \leq t_1 \leq \dots \leq t_v \leq n$ the following inequality is fulfilled:

$$|Cov(X_{s_1} \cdots X_{s_u}, X_{t_1} \cdots X_{t_v})| \leq K^2 M^{u+v-2} v \rho(t_1 - s_u), \quad (10)$$

and

$$\sum_{s=0}^{\infty} (s+1)^{k-2} \rho(s) \leq (k!)^\mu. \quad (11)$$

Then

$$P\left(\sum_{i=1}^n X_i \geq t\right) \leq \exp\left(-\frac{t^2/2}{\sigma_n^2 + B_n^{1/(\mu+2)} t^{(2\mu+3)/(\mu+2)}}\right),$$

where $B_n = 2(K \vee M) \left(\frac{2^{4+\mu} n K^2}{\sigma_n^2} \vee 1\right)$.

We apply this inequality to $(\Psi(X_i))_{i \in \mathbb{Z}}$. We thus need to check that the assumptions of the Theorem 3 hold.

In (9) the bound of the left hand side becomes:

$$v \mathbb{E} [|\Psi(X_{i_1}), \dots, \Psi(X_{i_u})|] \text{Lip}(\Psi) \|\Psi\|_\infty^{v-1} \tilde{\phi}(r) \leq \text{Lip}(\Psi) \|\Psi\|_\infty^{u+v-3} v \|\Psi\|_\infty \mathbb{E} |\Psi(X_0)| \tilde{\phi}(r).$$

Using $\|\Psi\|_\infty \mathbb{E} [|\Psi(X_0)|] \leq 1$, we can take $M \leq \sqrt{n}$ and $K^2 = \text{Lip}(\Psi)$. This is due to the fact that $\Psi \in \mathbb{L}^2(X_1)$ and that this dependence is a causal one.

We now study (11). We consider that the weak dependence coefficients are sub-geometrical, with a and b defined in Proposition 2. Quote that

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

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

$$\sum_{r=0}^{n-1} (r+1)^{q-2} \exp(-ar^b) \leq \frac{1}{ba^{\frac{q-1}{b}}} \int_0^\infty u^{\frac{q-1}{b}-1} \exp(-u) du = \frac{1}{ba^{\frac{q-1}{b}}} \Gamma\left(\frac{q-1}{b}\right),$$

and, finally, Jensen's inequality implies

$$\sum_{r=0}^{n-1} (r+1)^{q-2} \exp(-ar^b) \leq \frac{1}{ba^{\frac{q-1}{b}}} \Gamma^{\frac{1}{b}}(q-b) \leq \frac{1}{ba^{\frac{q-1}{b}}} (q-1)!^{\frac{1}{b}}.$$

The last inequality follows from Stirling's formula which entails that for any constant $C > 0$, and any $\epsilon > 0$ there exists $D_\epsilon > 0$ such that $C^k \leq D_\epsilon k!^\epsilon$. Thus, under this rate of dependence, assumptions of Theorem 3 hold with $\mu = \frac{1}{b}$. Moreover, Doukhan & Wintenberger (2005) have shown that $\sigma_n^2 \sim n$. So Theorem 3 can be applied, and it is easy to check that there exists a new constant C such that the parameter B_n is smaller than $C\sqrt{n}$. As a consequence, sub-geometrical $\tilde{\phi}$ -weak dependence implies Φ -dependence with $\Phi(x) = C\sqrt{x}$. \square

In the sub-geometrical $\tilde{\phi}$ -weak dependence setting, we achieve the same result than in the independence case. It means that the threshold chosen in estimation and the rate given by Theorem 2 are the same for i.i.d. and $\tilde{\phi}$ -weakly dependent observations.

3.2.2 Dynamical systems.

The notion of $\tilde{\phi}$ -weak dependence is specially well adapted for studying dynamical systems. Let us first recall the following definition:

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

$$\forall i \in \mathbb{N}, X_i := T^i(X_0) \quad (12)$$

where $T : I \rightarrow I$ for I a compact set of \mathbb{R} . T is a transformation admitting an invariant probability measure μ_0 , that we assume to be Lebesgue dominated, and X_0 is a random variable distributed according to μ_0 .

We restrict ourselves to a one-dimensional dynamical system class called *expanding map*:

Example.

$(X_i = T^i(X_0))_{i \in \mathbb{N}}$ is an expanding map (or equivalently T is a Lasota-Yorke function) if it satisfies the following three criteria:

- (*Regularity*) There exists a grid $0 = a_0 \leq a_1 \leq \dots \leq a_n = 1$ such as $T \in \mathcal{C}_1$ and $|T'(x)| > 0$ on $]a_{i-1}, a_i[$ for each $i = 1, \dots, n$.
- (*Expansivity*) Let I_n be the set on which $(T^n)'$ is defined. There exists $a > 0$ and $s > 1$ such that $\inf_{x \in I_n} \{|(T^n)'(x)|\} > as^n$.
- (*Topological mixing*) For any nonempty open sets U, V , there exists $n_0 \geq 1$ such that $T^{-n}(U) \cap V \neq \emptyset$ for all $n \geq n_0$.

This class has remarkable properties. As Viana proves in 1997, those dynamical systems are geometrically ergodic and they are $\tilde{\phi}$ -weakly dependent, with $\tilde{\phi}_r = \mathcal{O}(\exp(-ar))$ with $a > 0$. It follows from Theorem 3 that such dynamical systems are Φ -weakly dependent with $\Phi(x) = \sqrt{x}$. The corresponding parameter γ is equal to $1/2$.

Simulations.

Examples of expanding maps are given in Barbour (2000) and in Dedecker & Prieur (2005). One of the simplest one is when the transformation associated to (X_i) is $T(x) = 4x(1-x)$ (simulations and a Central Limit Theorem are given in Prieur (2001)). The invariant measure in this case is the distribution F defined in Section 3.1. Consequently the density we wish to estimate is again $f(x) = 1/(\pi\sqrt{x(1-x)})$. To insure the conditions of the Theorem 2, we first simulate 2^{10} stationary observations, such that $X_0 \rightsquigarrow F$ and $X_i = T^i(X_0)$. The MISE of the estimator for $\gamma = 1/2$ appears to be quite as good as for the i.i.d. case, and the optimal value for γ is 0.5 (see Figure 2).

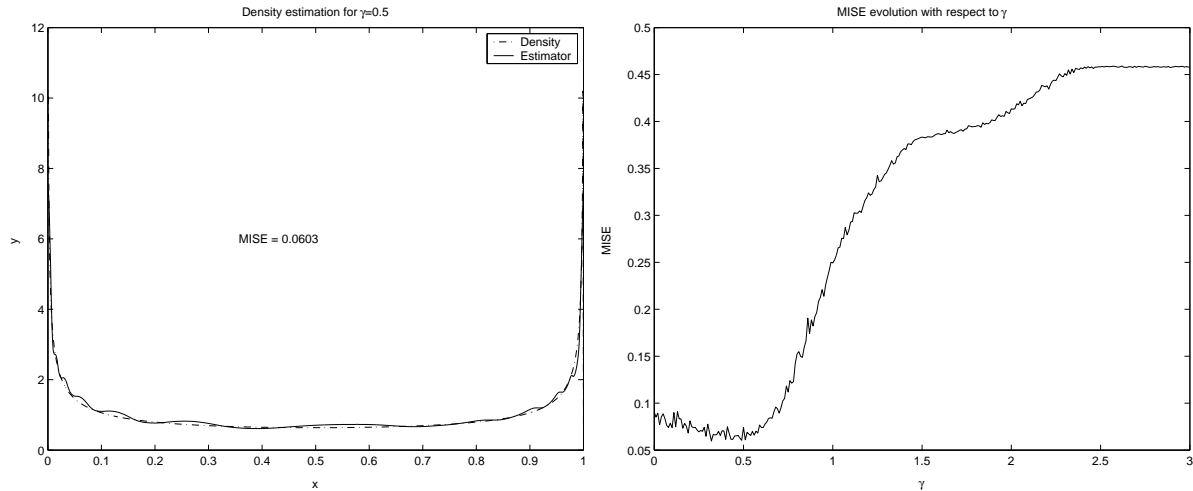


Figure 2: Causal and stationary case.

Unfortunately, the stationary case is irrelevant when the stable distribution is unknown, which is the case of interest for density estimation. This is why we simulated a geometrically ergodic sequence. Taking $X_0 \rightsquigarrow \mathcal{U}([0; 1])$, we simulated 2^{11} observations and retained only the 2^{10} last terms. The previous results remain true (see Figure 3) because the error, introduced by the lack of stationarity, has a geometric rate (i.e. it exists $c > 0$ such that the error is $= \mathcal{O}(\exp(-cn))$) and is then negligible compared to the error of estimation of the stationary case.

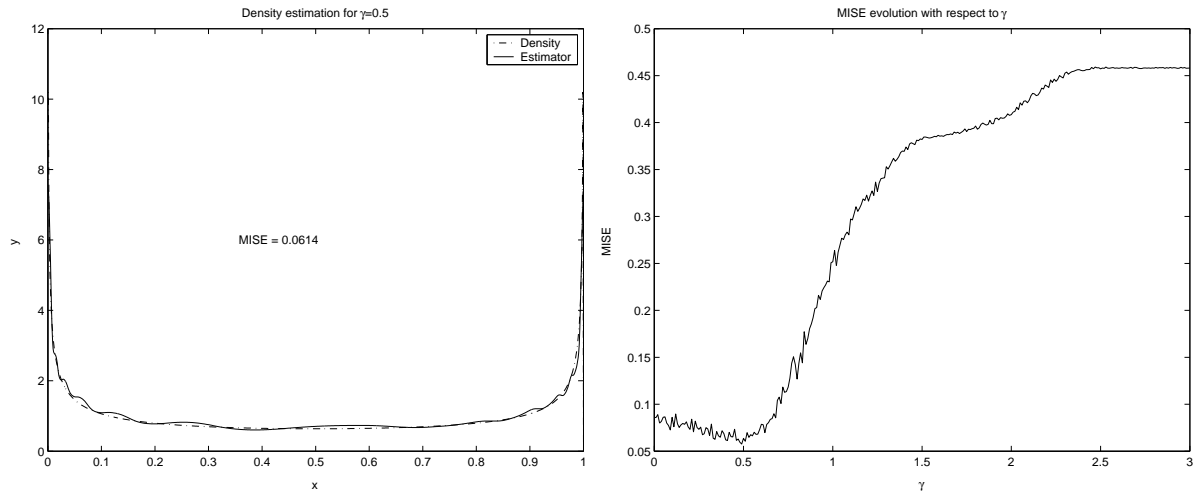


Figure 3: Causal and non stationary case.

Remark. Let T be a Lasota-Yorke transformation having an invariant density. An issue is often to determine this density. We propose to estimate it by simulating the corresponding ergodic dynamical system and then computing the threshold density estimator with $\gamma = 1/2$.

3.3 Non causal weak dependence: η -dependence

We are now interested by another kind of dependence which is the η -dependence, introduced by Doukhan (2002). A particularity of this notion of dependence is that it includes non-causal dependence; previously, we restricted ourselves to series depending uniquely on the past, whereas this definition allows observations to depend on the past and on the future. As previously, we will apply this definition to a concrete case, here a two-sided linear sequences.

Definition 4 (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 $(\eta_r)_{r \in \mathbb{N}}$ verifying $\eta_r \rightarrow 0$ when $r \rightarrow \infty$ and such that:*

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

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

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

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. Dedecker & Doukhan (2003) and Doukhan & Louhichi (1999) also proposed the θ -condition, which offers a similar definition in the case of causal dependence. But we did not improve our results in this case for geometric decays.

As previously, we check Φ -weak dependence of such sequences holds, and try to determine Φ . In this perspective, let us recall two lemmas of Ragache & Wintenberger (2005):

Lemma 1 (Ragache & Wintenberger (2005)). *Let X_n be a sub-geometrically η -weakly dependent sequence, i.e. $\eta_r = \mathcal{O}(e^{-ar^b})$ with $a > 0$ and $b > 0$. If $f_{j,k}$ the joints densities of (X_j, X_k) exist and are uniformly bounded on $j \neq k$, then for a sufficient n , X_n satisfies the following moment inequality, for all even integer q :*

$$\mathbb{E} \left[\left| \sum_{i=1}^n \Psi(X_i) \right| \right] \leq n^{q/2} \Phi(q), \quad \text{with } \Phi(x) = Cx^{1+1/b}.$$

Here C is a constant and Ψ is a function such as in the definition of Φ -weak dependence.

Lemma 2 (Ragache & Wintenberger (2005)). *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 $\lim_{x \rightarrow +\infty} \Phi(x) = +\infty$. Then:

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

Assume $(X_i)_{i \in \mathbb{Z}}$ is a $\tilde{\phi}$ -weakly dependent sequence and let Ψ be such that $\mathbb{E}[\Psi(X_0)] = 0$ and $\|\Psi\|_\infty \leq \sqrt{n}$. Taking $V_n = \sum_{i=1}^n \Psi(X_i)/\sqrt{n}$, we deduce from Lemmas 1 and 2 that sub-geometrical η -dependence implies Φ -weak dependence with $\Phi(x) = Cx^{1+1/b}$. We obtain the following proposition:

Proposition 3. *Let $(X_i)_{i \in \mathbb{Z}}$ be a eta-weakly dependent sequence.*

Suppose the dependence coefficients $(\eta_r)_{r \in \mathbb{N}}$ decrease sub-geometrically, i.e. for every r sufficiently large, there exist $a > 0$ and $b > 0$ such that $\tilde{\phi}_r \leq \exp(-ar^b)$ up to a constant. Then Φ -dependence holds with $\Phi(x) = Cx^{1+1/b}$, C positive constant.

Note that the value of the exponent $\gamma = 1 + 1/b$ to choose is here always superior to 1, whereas in the cases of i.i.d. or non causal dependence, we found that $\gamma = 1/2$ was a convenient choice. It is to say causal dependence differs from the last cases in the way it needs a different threshold in estimation. The rate obtain by Theorem 2 also differs of a logarithmic term from i.i.d. and $\tilde{\phi}$ -weakly dependent settings. We recalled previously this noncausal dependence does not achieve the same rates of causal dependence in linear estimation. This result tends to confirm that in a causal dependence setting such as η -weak dependence we do not attain the same rates. Nevertheless, the convergence given by Theorem 2 is still near minimax.

3.3.1 Two sided linear sequences.

We apply Theorem 2 to the case of two sided linear sequences $(X_i)_i$ of the form:

$$X_i = \sum_{j \in \mathbb{Z}} a_j \xi_{i-j},$$

where the series $(a_j)_{j \in \mathbb{Z}}$ is square integrable and $(\xi_j)_{j \in \mathbb{Z}}$ are i.i.d. variables. It follows from Doukhan & Lang (2002) that such a sequence is η -weakly dependent with:

$$\eta_r = \sqrt{\sum_{|j| > [r/2]} a_j^2}. \quad (13)$$

For the simulation, we focus ourselves on the special case of linear Markovian fields defined by:

$$X_t = (X_{t-1} + X_{t+1})/4 + \xi_t. \quad (14)$$

Then $X_t = \sum_{j \in \mathbb{Z}} a_j \xi_{t-j}$ with $a_j = 2/\sqrt{3} * (2 - \sqrt{3})^{|j|}$ is a stationary solution of (14) in \mathbb{L}^1 . Furthermore, X_i is geometrically η -weakly dependent by equation (13); this means there exists a such that $\eta_r = \mathcal{O}(\exp(-ar))$. In this example, we have $b = 1$. We will consider that ξ_0 is distributed as $\mathcal{N}(0, 1)$. The resulting common distribution of X_i is $\mathcal{N}(0, 8\sqrt{3}/9)$. Note that Theorem 2 assumes the density to have a compact support. We will thus consider for each simulation the restriction of f on the interval $[\min_i(X_i); \max_i(X_i)]$.

Under this framework, perfect simulation is not available. We will approach the true distribution of $(X_1, \dots, X_{2^{10}})$ applying the Gibbs algorithm. The idea is to simulate a vector Markov Chain $(X_1^{(t)}, \dots, X_{2^{11}}^{(t)})_{t \in \mathbb{N}^*}$ by setting:

$$\forall j \in \{1 \dots 2^{11}\}, \quad X_j^{(0)} \rightsquigarrow \mathcal{U}([0; 1])$$

$$X_j^{(t)} = (X_{j-1}^{(t)} + X_{j+1}^{(t-1)})/4 + \xi_t, \text{ where } \xi_t \rightsquigarrow \mathcal{N}(0, 1) \text{ and } (\xi_s)_{s \in \mathbb{N}} \text{ are i.i.d.}$$

Conditionally to the initial values $(X_1^{(0)}, X_{2^{11}}^{(0)})$, this chain is geometrically ergodic and its invariant distribution is the true one of $(X_1, \dots, X_{2^{11}})$. When simulating via Gibbs algorithm, it remains a dependence on the initial values. This dependence is geometrically decreasing with the number of iteration of Gibbs sampler (see Guyon (1995)). In order to relax it, one choose the observations obtained after a large number of iterations. We therefore take as observations $(X_{2^9+1}^{(1000)}, \dots, X_{2^9+2^{10}}^{(1000)})$. The error coming from the non-stationarity is negligible (see previous section) and the dependence with the initial value is also negligible compared to the error of estimation. Consequently, the convergence of Theorem 2 holds.

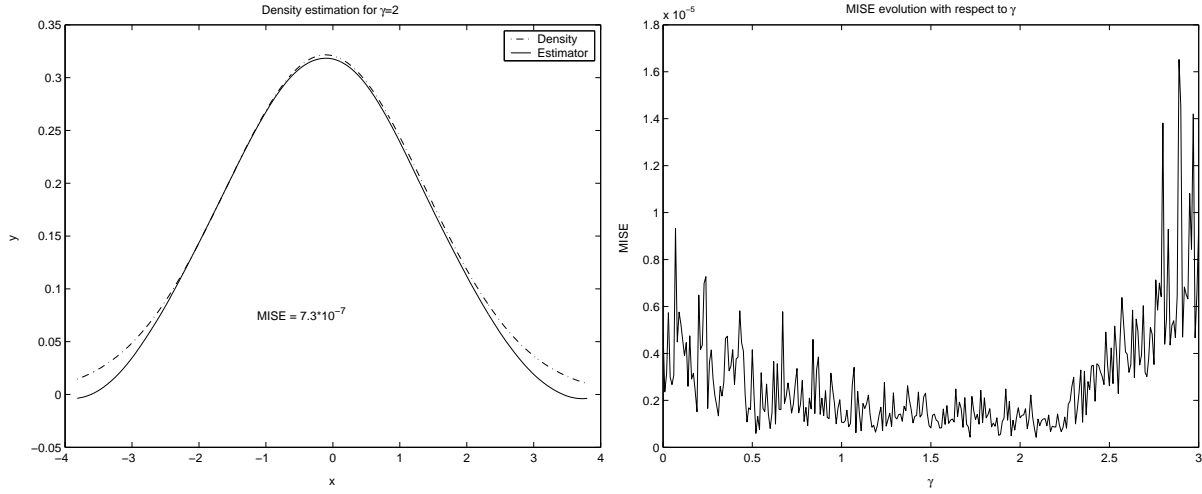


Figure 4: Non causal and non stationary case.

In Figure 4, the MISE of the estimation is of order smaller than before, due to the regularity of the density of the gaussian densities. Furthermore, the best estimation is achieved here for $\gamma = 2$, corresponding to $b = 1$. The result of Theorem 2 is then validated by simulation, even if we do not have hope to reach minimax results. The study of the optimal exponent γ told us that $\Phi(x) = x^{1+1/b}$ seems to be the smallest function we can put in the probability inequality of the definition of Φ -weak dependence, when dealing with η -weakly dependent sequences. We then provide a numeric way to determine the shape of the coefficient Φ for any dependent time series.

The purpose of the present work was to increase the scope of the hard-threshold wavelet density estimator to dependent settings. In the independent and identically distributed case, Donoho, Johnstone, Kerkycharian & Picard (1996) proved this nonlinear estimation achieves quasi-minimax rates. In order to extend this result, we first introduced a new functional dependence notion, which covers the independence setup and weak dependence conditions such as Dedecker & Prieur's (2004) $\tilde{\phi}$ -dependence or Doukhan & Louhichi's (1999) η -dependence. Actually, we proved that an appropriate modification of the threshold in estimation could take into account the dependence of the observations. With this

new threshold, the wavelet density estimator is also quasi-minimax for dependent frameworks.

The application of this result to independent, causal dependent and finally non causal dependent observations illustrates the large amount of cases covered. We study with simulations the behavior of the estimation with respect to the level of threshold and used cross-validation to choose it. The results obtained were satisfactory and this approach can provide a way for getting information on the kind of dependence of the data.

4 Proof of the Theorem 2

Let $(X_1 \dots X_n)$ be n observations of a Φ -weakly dependent process (see Definition 1), with a common density f . As it was told before, we restrict ourselves to the case of compact supported distributions. We consider 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 $\mathcal{B}_{\pi,r}^s$ and that the assumptions on the indexes given in Theorem 2 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 wavelets spaces. (See Härdle et al. (1998)).

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 $\lambda_j = K \frac{\Phi(j)}{\sqrt{n}}$. The functions j_0 and j_1 are given by $j_0 = \log_2(n^{1/(1+N)})$ and $j_1 = \log_2(n/\Phi^2(\log(n)))$.

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

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

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.

4.1 Some technical tools

The assumption of Φ -weak dependence is useful in the proof of the Theorem 2 because it provides us moments inequalities for the estimation of the scale and wavelet coefficients. We have, for all $j \in \mathbb{N}$ and $k \in \mathbb{Z}$:

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

Furthermore, we restrict ourselves to the cases where $j \leq j_1$. 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} . 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 $k \in \mathbb{Z}$. The definition of Φ -weak dependence leads through the inequality (3) to the moments inequalities:

$$\left. \begin{aligned} \mathbb{E}|\hat{\alpha}_{j,k} - \alpha_{j,k}|^p \\ \mathbb{E}|\hat{\beta}_{j,k} - \beta_{j,k}|^p \end{aligned} \right\} \preceq n^{-p/2}, \quad (15)$$

where \preceq signifies “up to a constant that does not depends of n , j and k ”.

4.2 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), or by applying Theorem 9.3 of Härdle et al. (1998), we obtain

$$T_3 \preceq \left(2^{-j_1 s'}\right)^p.$$

Recall that we have taken $2^{-j_1} = \frac{\Phi^2(\log(n))}{n}$. Moreover, $p \geq \pi$ implies $s' \leq \alpha$ and as a consequence T_3 has the right rate of convergence.

4.3 Bias of scale estimation T_1

Paralleling the proof of Theorem 5 of Donoho et al. (1996), we can apply the result of Meyer (1992) for the scaling function ϕ satisfying a concentration assumption. It gives

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

Thanks to the inequality (15), the rate of convergence of T_1 is bounded by:

$$T_1 \preceq 2^{j_0 p/2} n^{-p/2}.$$

Note that the choice (4) of j_0 implies that the order of the bound is $(2^{j_0}/n)^{p/2} = n^{-pN/(2+2N)}$. We conclude T_1 is negligible thanks to the hypothesis $N \geq 2s$.

4.4 Details term T_2

We apply the same result from Meyer (1992) 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:

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

Then $T_2 \preceq T_{21} + T_{22} + T_{23} + T_{24}$ 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.

4.4.1 Term T_{23}

If we are in the set $\hat{B}_j \cap B_j^-$ or in 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:

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

For the special choice of the threshold $\lambda_j = K \frac{\Phi(j)}{\sqrt{n}}$, assumption of Φ -weak dependence leads to the following bound:

$$\mathbb{P}(|\hat{\beta}_{j,k} - \beta_{j,k}| > \lambda_j/2) \preceq e^{-\Phi^{-1}(K\Phi(j)/2e)}. \quad (16)$$

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

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

Added to the fact we control $\sum_k |\beta_{j,k}|^p 2^{jp(s'+1/2-1/p)}$ because assumptions in Theorem 2 implies that f belongs to a Besov space $\mathcal{B}_{p,\infty}^{s'}$, we obtain:

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

It follows that $T_{23} \preceq 2^{-j_0(s'p+\nu)} \preceq n^{-(s'p+\nu)/(1+N)}$.

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.

4.4.2 Term T_{21}

For 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} . Added to Cauchy-Schwarz inequality, we obtain:

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

Using (16), we have:

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

Inequality (15) now leads to

$$T_{21} \preceq n^{-p/2} \sum_{j_0}^{j_1} 2^{jp/2-\nu/2}.$$

As for T_{23} , we may choose the constant K large enough to get $\nu \geq p$. Then, we can write:

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

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.

4.4.3 Term T_{24}

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

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

As $\pi - p < 0$, we have the inequality

$$\mathbb{1}_{\hat{B}_j^C \cap B_j^{+C}} \leq \mathbb{1}_{|\beta_{j,k}| \leq 2\lambda_j} \leq \left| \frac{\beta_{j,k}}{2\lambda_j} \right|^{(\pi-p)}. \quad (17)$$

We thus can easily bound T_{24} by:

$$T_{24} \preceq \sum_{j=j_0}^{j_1} 2^{j(p/2-1)} \lambda_j^{p-\pi} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^\pi.$$

Replacing λ_j by its value, we have:

$$T_{24} \preceq \left(\frac{\Phi(j_1)^2}{n} \right)^{(p-\pi)/2} \sum_{j=j_0}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^\pi. \quad (18)$$

Then, using the control of $\sum_k |\beta_{j,k}|^\pi 2^{j\pi(s+1/2-1/\pi)}$ given by the inclusion $\mathcal{B}_{\pi,r}^s \subset \mathcal{B}_{\pi,\infty}^s$, we obtain as for T_{22} :

$$T_{24} \preceq \sum_{j=j_0}^{j_1} 2^{jp/2} \lambda_j^{p-\pi} 2^{-j(s+1/2)\pi} \|f\|_{s,\pi,\infty}^\pi \preceq \left(\frac{\Phi(j_1)^2}{n} \right)^{(p-\pi)/2} \sum_{j=j_0}^{j_1} 2^{-j\epsilon}. \quad (19)$$

Let us define $v_n := n^{-(p-\pi)/2} \sum_{j=j_0}^{j_1} 2^{-j\epsilon}$.

We now must distinguish on the values of ϵ .

- First, if $\epsilon > 0$, v_n is bounded by $n^{-(p-\pi)/2} 2^{-j_0\epsilon}$. So, it gives with j_0 as in (4): $v_n \leq n^{-\epsilon/(1+N)-(p-\pi)/2}$. With the hypothesis $N \geq 2s$ and thanks to the equality $\epsilon/(1+2s) + (p-\pi)/2 = \alpha p$ when $\epsilon \geq 0$, we thus obtain that the rate of v_n is at worse $n^{-\alpha p}$.
- If $\epsilon < 0$, the bound becomes $n^{-(p-\pi)/2} 2^{-j_1\epsilon}$. Thanks to the choice of j_1 as in (5), this bound is smaller than $(\Phi^2(\log(n)))^\epsilon n^{-\epsilon-(p-\pi)/2}$. Like above, we can notice that $s' \leq \alpha$ and that if $\epsilon \leq 0$, we have the equality $\epsilon\alpha/s' + (p-\pi)/2 = \alpha p$. Then the exponent of n is greater than $-\alpha p$, and so $v_n \leq (\Phi^2(\log(n)))^\epsilon n^{-\alpha p}$.

The case $\epsilon = 0$ needs to be treated separately. Indeed, if we consider equation (19), we obtain $v_n \leq j_1 n^{-\alpha p} \leq \log(n) n^{-\alpha p}$ and the resulting bound is not satisfactory.

First, when $\epsilon = 0$, inequality (18) becomes

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

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

$$T_{24} \preceq \left(\frac{\Phi(j_1)^2}{n} \right)^{(p-\pi)/2} \sum_{j=j_0}^{j_1} t_j^\pi,$$

and Hölder inequality implies:

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

with $1/r + 1/r' = 1/\pi$. As f belongs to $\mathcal{B}_{\pi,r}^s$, which is included in $\mathcal{B}_{\pi,\infty}^s$, it comes

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

Thus,

$$T_{24} \preceq \left(\frac{\Phi(j_1)^2}{n} \right)^{(p-\pi)/2} j_1^{(1-\pi/r)}.$$

To conclude,

$$T_{24} \leq \Phi(j_1)^{p-\pi} n^{-\alpha p} \begin{cases} 1 & \text{if } \epsilon > 0, \\ (\Phi^2(\log(n)))^\epsilon & \text{if } \epsilon < 0, \\ (\log n)^{(1-\pi/r)_+} & \text{if } \epsilon = 0. \end{cases} \quad (20)$$

Moreover, as Φ is increasing, $\Phi(j_1) \leq \Phi(\log(n))$. Finally, we obtain the announced rate, by noting that $(p - \pi)/2 \leq \alpha p$ for $\epsilon \geq 0$ and $(p - \pi)/2 + \epsilon \leq \alpha p$ for $\epsilon < 0$.

4.4.4 Term T_{22}

Using the same method as in (17), we note that

$$\mathbb{1}_{\hat{B}_j \cap B_j^{-C}} \leq \mathbb{1}_{|\beta_{j,k}| > \lambda_j/2} \leq \left| \frac{2\beta_{j,k}}{\lambda_j} \right|^\pi. \quad (21)$$

Under the Φ -weak dependence, the stationarity of X_n and using Jensen's inequality as in Subsection 4.1, we bound $\mathbb{E}|\hat{\alpha}_{j,k} - \alpha_{j,k}|^p$ by $n^{-p/2}$ uniformly in j and k . We then obtain:

$$T_{22} \preceq n^{-p/2} \sum_{j=j_0}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} \left| \frac{\beta_{j,k}}{\lambda_j} \right|^\pi.$$

Replacing λ_j by its value, this can be written:

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

At this step, we clearly see that the bound of (19) for T_{24} is larger than the one for T_{22} of (22) as soon as $\Phi(j_1) \geq 1$. Effectively, we have:

$$T_{22} \preceq \Phi(j_1)^{-p} \left(\frac{\Phi(j_1)}{\Phi(j_0)} \right)^\pi \frac{\Phi(j_1)^{p-\pi}}{n^{(p-\pi)/2}} \sum_{j=j_0}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^\pi,$$

with $\Phi(j_1) \geq \Phi(j_0)$ because Φ is increasing. As j_1 and Φ have infinite limits, $\Phi(j_1) \geq 1$ is verified for n sufficiently large, and thus T_{22} achieves the right rate. \square

References

- [1] D. Andrews. Non strong mixing autoregressive processes. *Journal of Applied Probability*, 21:930–934, 1984.
- [2] P. Ango Nze, P. Buhlmann, and P. Doukhan. Weak dependence beyond mixing and asymptotics for nonparametric regression. *The Annals of Statistics*, 30(2):397–430, 2002.
- [3] A. Antoniadis. Wavelets in statistics: a review. Technical report, Université Joseph Fourier, Grenoble, 1997.
- [4] A. Antoniadis and R. Carmona. Multiresolution analyses and wavelets for density estimation. Technical report, University of California at Irvine, 2001.
- [5] A. Antoniadis and J. Fan. Regularization of wavelet approximations. *Journal of the American Statistical Association*, 96(455):939–967, 2001.
- [6] A.D. Barbour, R.M. Gerrard, and G. Reinert. Iterates of expanding maps. *Probability Theory and Related Fields*, 116:151–180, 2000.
- [7] I. Daubechies. *Ten lectures on wavelets*, volume 61. SIAM press, 1992.
- [8] J. Dedecker and C. Prieur. New dependence coefficients: Examples and applications to statistics. *Probability Theory and Related Fields*, (132):203–235, 2005.
- [9] P. Dedecker, J. Doukhan. A new covariance inequality and applications. *Stochastic Processes and their Applications*, 106-1:63–80, 2003.
- [10] B. Delyon and A. Juditsky. Wavelet estimators, global error measures: Revisited. Technical report, IRISA-INRIA, 1993.
- [11] L. Devroye. *Nonparametric density estimation*. Wiley, 1985.
- [12] D. Donoho, I. Johnstone, G. Kerkycharian, and D. Picard. Density estimation by wavelet thresholding. *The Annals of Statistics*, 24(2):508–539, 1996.
- [13] P. Doukhan. Formes de Toeplitz associées à une analyse multi-échelle. *note au CRAS*, 306(1):663–666, 1988.
- [14] P. Doukhan. *Mixing: properties and examples*. Lecture Notes in Statistics 85, Springer-Verlag, 1994.
- [15] P. Doukhan. Models inequalities and limit theorems for stationary sequences. (*Doukhan et alii ed.*) *Birkhauser, Birkhauser*, pages 43–101, 2002.
- [16] P. Doukhan and G. Lang. Rates of convergence in the weak invariance principle for the empirical repartition process of weakly dependent sequences. *Statistical Inference for Stochastic Processes*, 5:199–228, 2002.
- [17] P. Doukhan and J. León. Déviation quadratique d’estimateur de densité par projection orthogonale. Technical report, Comptes rendus Académie des Sciences de Paris, 1990.

- [18] P. Doukhan and S. Louhichi. A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and their Applications*, 84:313–342, 1999.
- [19] P. Doukhan and M. Neumann. A Bernstein type inequality for times series. unpublished, 2005.
- [20] P. Doukhan and O. Wintenberger. A central limit theorem under non causal weak dependence and sharp moment assumptions. unpublished, 2005.
- [21] X. Guyon. *Random fields on a network: modeling, statistics, and applications*. New York: Springer-Verlag, 1995.
- [22] P. Hall and P. Patil. Formulae for mean integrated squared error of nonlinear wavelet-based density estimators. *Annals of Statistics*, 23(3):905–928, 1995.
- [23] W. Hardle, G. Kerkycharian, D. Picard, and A. Tsybakov. *Wavelets Approximation and Statistical Applications*, volume 129. Springer Verlag, Lecture Notes in Statistics, 1998.
- [24] A. Juditsky and S. Lambert-Lacroix. On minimax density estimation on \mathbb{R} . *Bernoulli*, 10(2):187–220, 2004.
- [25] G. Kerkycharian and D. Picard. Density estimation in besov spaces. *Statistics and Probability Letters*, 13:15–24, 1992.
- [26] G. Kerkycharian and D. Picard. Density estimation by kernel and wavelets methods: optimality of besov spaces. *Statistics and Probability Letters*, 18:327–336, 1993.
- [27] Y. Meyer. *Wavelets and operators*. Cambridge University Press, 1992.
- [28] C. Prieur. *Applications statistiques de suites faiblement dépendantes et de systèmes dynamiques*. PhD thesis, CREST, 2001.
- [29] C. Prieur. Density estimation for one-dimensional dynamical systems. *ESAIM , Probability & Statistics*, pages 51–76, 2001.
- [30] N. Ragache and O. Wintenberger. Convergence rates for density estimators of weakly dependent time series. unpublished, 2005.
- [31] M. Vannucci. Nonparametric density estimation using wavelets. Technical report, Texas A and M university, 1998.
- [32] M. Viana. Stochastic dynamics of deterministic systems. available at <http://w3.impa.br/~viana>, 1997.
- [33] G.G. Walter. Approximation of Delta function by wavelets. *Journal of Approximation Theory*, 71:329–343, 1992.
- [34] G.G. Walter. *Wavelets and others orthogonal systems with applications*. CRC press, Boca Raton, FL, 1994.
- [35] Wavelab. <http://www-stat.stanford.edu/~wavelab/>.
- [36] H. van Zanten and P. Zareba. A note on wavelet density deconvolution for weakly dependent data. unpublished, 1999.