ANNALES DE L'I. H. P., SECTION B

KARINE TRIBOULEY
GABRIELLE VIENNET

L_p adaptive density estimation in a β mixing framework

Annales de l'I. H. P., section B, tome 34, nº 2 (1998), p. 179-208
<http://www.numdam.org/item?id=AIHPB_1998__34_2_179_0>

Ann. Inst. Henri Poincaré,

Vol. 34, n° 2, 1998, p. 179-208. *Probabilités et Statistiques*

$\mathbb{L}_p$ adaptive density estimation in a β mixing framework

by

Karine TRIBOULEY and Gabrielle VIENNET

Laboratoire de modélisation stochastique et statistique, Bat 425,
Université Paris Sud, 91405 Orsay Cedex, France.

ABSTRACT. – We study the $\mathbb{L}_\pi$-integrated risk with $\pi \geq 2$ of an adaptive density estimator by wavelets method for absolutely regular observations. By a duality argument, the study of the risk is linked to the control of the supremum of the empirical process over a suitable class of functions. The main argument is a generalization to absolutely regular variables of a result of Talagrand stated for i.i.d. variables. Assuming that the sequence of the β-mixing coefficients $(\beta_l)_{l\geq 0}$ is arithmetically decreasing, we prove that our estimator is adaptive in a class of Besov spaces with unknown smoothness. © Elsevier, Paris

Key words and phrases: Adaptive estimation, absolutely regular variables, Besov spaces, density estimation, strictly stationary sequences, wavelet orthonormal basis.

RÉSUMÉ. – Dans un cadre des variables absolument régulières, on étudie le risque $\mathbb{L}_\pi$-intégré, $\pi \geq 2$, d'un estimateur par méthode d'ondelettes adaptatif. À l'aide d'un argument de dualité, l'étude du risque est liée au contrôle du supremum du processus empirique sur une classe adéquate de fonctions. L'argument principal est une généralisation à des variables absolument régulières d'un résultat de Talagrand enoncé pour des variables i.i.d. En supposant que la suite des coefficients de β-mélange $(\beta_l)_{l\geq 0}$ est arithmétiquement décroissante, on démontre que notre estimateur est adaptatif dans une classe d'espace de Besov de régularité inconnue. © Elsevier, Paris

1. INTRODUCTION

Let $(X_1, \ldots, X_n)$ be n observations drawn from some strictly stationary process $(X_i)_{i \in \mathbb{Z}}$. The aim of this paper is to study adaptive estimation of the stationary density f of the process under some adequate mixing assumptions. Assuming some prior knowledge on f (such as its degree of smoothness for instance) it is possible to prove the optimality of many estimators. But, from a practical point of view, this is not satisfactory: one would rather prefer to get optimal estimation without any extra knowledge on the density. Recently different procedures have been proposed in the i.i.d. case which all tend to reduce the prior information needed to estimate the unknown density. Roughly, we say that such procedures are adaptive, the precise sense being defined in each specific situation. Let us present the definition of adaptivity that we use in the sequel. We first recall the minimax risk associated to the loss function $\mathbb{L}_\pi, \pi \geq 1$ for a set of functions $\mathcal{F}_\alpha$:

$$R_n(\alpha) = \inf_{\hat{f}} \sup_{f \in \mathcal{F}_\alpha} \mathbb{E}||\hat{f} - f||_\pi^\pi, \tag{1}$$

where the infimum is taken over all estimators $\hat{f}$. We say that an estimator f^* is adaptive in a class of functions $\{\mathcal{F}_\alpha; \alpha \in A\}$ if and only if there exists a positive constant $C(\alpha)$ such that:

$$\forall \alpha \in A, \forall f \in \mathcal{F}_\alpha, \ \mathbb{E}||f^* - f||_\pi^\pi \leq C(\alpha)\, R_n(\alpha). \tag{2}$$

We aim at giving an account of the different constructions of adaptive estimators and of their performances in the independent framework. One of the most popular method is the cross-validation method. It consists in minimizing an empirical criterion which tends to estimate the unknown quadratic loss. The first example of adaptation to unknown smoothness in the sense of Definition (2) appears in a crucial paper of Efromovich and Pinsker [11]. They deal with the white noise model in the context of the Fourier basis and their procedure is based on thresholding. Efromovich [10] has adapted their method to density estimation. They obtain adaptation over a class $\mathcal{F}_\alpha$ of Sobolev ellipsoids relatively to $\mathbb{L}_2$ loss. The introduction of wavelet bases provides more accurate approximation than the Fourier basis for functions with spacially inhomogeneous smoothness. They allow adaptation over more complicated functions classes. Wavelet thresholding methods have been extensively developed these last years and we refer to Donoho *et al.* [7] for numerous references. Let us describe the method of local thresholding. Let f be the unknown density to be estimated. One assumes that f belongs to a ball of the Besov space $\mathcal{B}_{s,p,q}$ and is compactly supported. More precisely, one assumes that

$$f \in \mathcal{F}_{s,p,q}(M_1, M_2, B)$$
$$= \{f \text{ density}, \text{supp } (f) \in [-B; B], \|f\|_{spq} \leq M_1, \|f\|_\infty \leq M_2\},$$

where $\|\cdot\|_{s,p,q}$ denotes the Besov norm. The density is expanded on some wavelet basis into a sum of a low frequency term and a detail term, and one considers the projection estimator. The local thresholding method consists in keeping only the empirical details greater than a fixed level and the computation of the estimator requires the knowledge of the radius M_1. Let N be the number of vanishing moments of the wavelet. Donoho *et al.* show that the local thresholding estimator is adaptive (up to a power of $\log(n)$ for the small regularities) over the class $\{\mathcal{F}_{s,p,q}(M_1, M_2, B), 1/p < s < N+1, p \geq 1, q \geq 1, B > 0\}$ relatively to any $\mathbb{L}_\pi$-loss function, $\pi \geq 1$. A global thresholding procedure is studied by Kerkyacharian *et al.* [14]: they compute a statistic close to the l_π-norm of the empirical details at a fixed resolution level; they keep all the empirical details of the resolution level, if this quantity is greater than a fixed amount. The global procedure of Kerkyacharian *et al.* is adaptive over the class of functions $\{\mathcal{F}_{s,p,q}(M_1, M_2, B), 1/p < s < N+1, p \geq \pi,$ $q \geq 1, B > 0, M_1 > 0, M_2 > 0\}$ relatively to the loss function $\mathbb{L}_\pi, \pi \geq 2$. This latter procedure has been inspired by the work of Efromovich [10]. For the particular case $\pi = 2$, it is similar to the cross validation procedure with the advantage to provide an explicit estimator.

The problem of adaptive estimation in weakly dependent framework is quite new. In the case of the estimation of the regression function (in linear AR models), some results have been obtained by Dahlaus *et al.* [5], Hoffmann [12].

In this paper, we propose to extend the thresholding methods to the dependent framework. In fact, without any *a priori* independence assumption on the data, it is interesting to get some robustness results with respect to dependence. We focus on the situation where the data are absolutely regular. It covers a large class of examples and allows us to use coupling technics in the proofs. Examples of such processes may be found in Doukhan [8]. Even if the local thresholding method can easily be generalized in the weaker case of $\phi-$mixing, it is not adapted to the absolutely regular context. Nevertheless, the global thresholding method is well fitted for this dependent framework. Introducing a small modification of the threshold, we show that the global thresholding method preserves its adaptive properties in the minimax $\mathbb{L}_\pi$-sense, $\pi \geq 2$. Our results may be extend to other bases such as splines; we focus on the wavelet estimator which provides clearest proofs.

Let us present our results. Let $(X_i)_{i\in\mathbb{Z}}$ be a strictly stationary absolutely regular process, also called β-mixing with a sequence of β-mixing coefficients $(\beta_l)_{l\geq 0}$. We assume that the rate of decay of the sequence $(\beta_l)_{l\geq 0}$ is arithmetic. More precisely, we assume that there exists $\theta > \pi - 2$ and a positive constant B_β such that

$$\forall l \geq 1, \beta_l \leq B_\beta\, l^{-(1+\theta)},$$

and we denote by B_π the bound for the series

$$\sum_{l\geq 0}(l+1)^{\pi-2}\beta_l \leq B_\pi < \infty. \tag{3}$$

Such an assumption of arithmetic decay of the coefficients is often made in papers on density estimation in a dependent framework. Let us remark that for $\pi = 2$, the condition $B_2 < \infty$ is known to be a minimal condition for results like the central limit theorem (*see* Doukhan *et al.* [9]).

In this paper we consider a compactly supported wavelet basis with scaling function φ and wavelet function ψ and we denote by N the number of vanishing moments of the wavelet. We use an estimator $\hat{f}$ similar to the global threshold estimator introduced by Kerkyacharian *et al.* [14] which computation depends on $\|f\|_\infty$ and on B_π if $\pi > 2$, but only on B_2 if $\pi = 2$.

We show that if $2 < \pi$, as soon as

$$\theta > \pi - 2 + \frac{\pi^2}{2}, \tag{4}$$

our global threshold estimator is adaptive in the class:

$$\{\mathcal{F}_{s,p,q}(M_1, M_2, B), 1/p < s < N+1, p \geq \pi, q \geq 1, B > 0, M_1 > 0\}.$$

For $\pi = 2$, as soon as $\theta > 2$, the estimator is adaptive in the class:

$$\{\mathcal{F}_{s,p,q}(M_1, M_2, B), 1/p < s < N+1, p \geq 2, q \geq 1, B > 0, M_1 > 0, M_2 > 0\}.$$

The idea of the proof of adaptation is completely different as the one used in the independent case. It is based on an interesting result stated in Theorem 2.2 which relies on an important Theorem of Talagrand [22]. In the same spirit as in Birgé and Massart [4], it provides a control of the supremum of the empirical process over a suitable class of functions. This method has the advantage of providing simpler proofs and estimates than those of Kerkyacharian *et al.* [14] at the price of introducing the unknown constants B_π and $||f||_\infty$ if $\pi > 2$. Nevertheless, we propose a practical procedure using an over-estimate of the unknown quantity $||f||_\infty$. We prove that the error due to this practical procedure is, in probability, of the same order than the error due to the adaptive procedure. Under the same mixing conditions as before, this estimator is shown to be adaptive in probability in the class

$$\{\mathcal{F}_{s,p,q}(M_1, M_2, B), (s, p, q, B, M_1, M_2) \in A\},$$

with

$$A = \{(1/p \vee 1)/(2(1+\theta)) < s < N+1, p \geq \pi, q \geq 1, B > 0, M_1 > 0, M_2 > 0\},$$

that is

$$\forall \varepsilon > 0,\ \exists A_1(\varepsilon) > 0,\ \exists A_2(\varepsilon) > 0,$$

$$\forall \alpha \in A, \forall f \in \mathcal{F}_\alpha,\ \forall n \geq A_1(\varepsilon),\ \forall \lambda \geq A_2(\varepsilon)$$

$$P(||f^* - f||_\pi \geq \lambda R_n(\alpha)) \lesssim \varepsilon. \tag{5}$$

Let us remark that the *a priori* knowledge about the process needed to make the estimation is quite reasonable. Indeed, as noticed before, the assumption of arithmetic decay of the mixing coefficients is usual and suggest that B_π is not too large. It is also important to notice that in contrast to previous authors, no assumption is required on the joint law of (X_0, X_1). In fact as mentioned, our result could be understood in the following way: when we think that the process is nearly independent but when independency is debatable, a safe strategy to avoid too large errors is to increase the threshold.

The paper is organized as follows. In the section 2, we briefly recall some results about Besov spaces, wavelets and absolutely regular processes. We state our main tool Theorem 2.2. We introduce in section 3 the two estimators of interest and study their adaptive properties. All the proofs are given in section 4.

2. WAVELETS AND ABSOLUTELY REGULAR PROCESSES

We first review the very basic features of the multiresolution analysis of Meyer [17] and give useful elements of wavelets analysis. We recall then the definition of absolutely regular mixing coefficients and a result used on several occasions along this work. This result provides a sharp control of the variance of the empirical process and a Rosenthal type moment inequality for bounded absolutely regular variables. It is proved in Viennet [23]. Finally, we state in Theorem 2.2 a control on the probability tails for the supremum of the empirical process on a suitable class of functions. This result is based on a recent result of Talagrand [21] avaible for independent and identically distributed variables and is shown in the last section of this paper.

2.1. Wavelets and Besov spaces

One can construct a real function φ (the scaling function) such that

1. the sequence $\{\varphi_{0k} = \varphi(. - k) | k \in \mathbb{Z}\}$ is an orthonormal family of $\mathbb{L}_2(\mathbb{R})$. Let us call V_0 the subspace spanned by this sequence.

2. if V_j denotes the subspace spanned by the sequence $\{\varphi_{jk} = 2^{j/2}\varphi(2^j. - k) | k \in \mathbb{Z}\}$, then $\{V_j\}_{j\in\mathbb{Z}}$ is an increasing sequence of nested spaces such that

$$\cap_{j\in\mathbb{Z}} V_j = \{0\}, \qquad \text{and if} \quad \int \varphi = 1, \quad \overline{\cup_{j\in\mathbb{Z}} V_j} = \mathbb{L}_2.$$

It is possible to require in addition that φ is of class C^{r_0} and compactly supported (Daubechies wavelets [6]). We define the space W_j by the following: $V_{j+1} = V_j \oplus W_j$. Then, there also exists a function ψ (the wavelet) such that

1. ψ is of class C^{r_0} compactly supported,
2. $\{\psi_{0k} = \psi(. - k) | k \in \mathbb{Z}\}$ is an orthonormal basis of W_0,
3. $\{\psi_{jk} = 2^{j/2}\psi(2^j. - k) | k \in \mathbb{Z}, j \in \mathbb{Z}\}$ is an orthonormal basis of $\mathbb{L}_2$.

For $j_0 \in \mathbb{Z}$, the following decomposition is also true

$$\forall f \in \mathbb{L}_2(\mathbb{R}), \qquad f = \sum_{k\in\mathbb{Z}} \alpha_{j_0k}\varphi_{j_0k} + \sum_{j\geq j_0}\sum_{k\in\mathbb{Z}} \zeta_{jk}\psi_{jk} \tag{6}$$

where

$$\alpha_{jk} = \int f(x)\varphi_{jk}(x)dx \quad \text{and} \quad \zeta_{jk} = \int f(x)\psi_{jk}(x)dx. \tag{7}$$

According to a result of Meyer [17], we link the $\mathbb{L}_r$-norm of the details at the level j or the $\mathbb{L}_r$-norm of the low frequency part to the l_r-norm of the wavelets coefficients.

LEMMA 2.1. – *Let g be either φ or ψ, with the conditions above, let $\Gamma(x) = \Gamma_g(x) = \sum_{k\in\mathbb{Z}} |g(x-k)|$, and $\|\Gamma\|_r = \left(\int_0^1 |\Gamma(x)|^r dx\right)^{1/r}$. Let $f(x) = \sum_{k\in\mathbb{Z}} \lambda_k 2^{j/2} g(2^j x - k)$, then for any r, $1 \le r \le \infty$, and r_1 such that $1/r + 1/r_1 = 1$,*

$$\frac{1}{\|\Gamma\|_1^{1/r_1}\|\Gamma\|_\infty^{1/r}} 2^{j(1/2-1/r)}|\lambda|_r \le \|f\|_r \le \|\Gamma\|_r 2^{j(1/2-1/r)}|\lambda|_r$$

where $|\cdot|_r$ is the l_r-norm.

Besov spaces are characterized in terms of wavelet coefficients (*see* Meyer [17]); we do not use this characterization but just the following property (8). Let N be a positive integer. We consider in the sequel a wavelet basis such that the scaling function φ and the wavelet function ψ satisfy the following properties $\mathcal{P}(N)$:

(i) for any $m = 0, \ldots, N$

$$\int |\varphi(t)|^m dt < +\infty$$

(ii)

$$\sum_k |\varphi(x-k)| \in \mathbb{L}_\infty(\mathbb{R})$$

(iii) for any $m = 0, \ldots, N$

$$\int \psi(t) t^m dt = 0.$$

As a typical example, the Daubechies wavelets DB2N satisfy $\mathcal{P}(N)$ (Daubechies [6]). Then, for any $f \in B_{s,p,q}$, $0 < s < N+1$, $1 \le p \le \infty$, $1 \le q \le \infty$

$$\sum_{j\ge 0} (2^{j(s+1/2-1/p)}|\zeta_{j.}|_p)^q < +\infty. \tag{8}$$

We can notice moreover that, as soon as f is compactly supported, only a finite number of coefficients α_{0k}, (or ζ_{jk}), is nonzero. In fact, this number is less than $2^j A\, B^{-1}$ where $2B$ and $2A$ are the respective lengths of the supports of f and ψ.

2.2. Absolutely regular sequences

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. For any two σ-algebra $\mathcal{U}$ and $\mathcal{V}$ of $\mathcal{A}$, the absolutely regular mixing (or β mixing) coefficient $\beta(\mathcal{U}, \mathcal{V})$ is defined by

$$\beta(\mathcal{U}, \mathcal{V}) = \frac{1}{2} \sup \left(\sum_{i \in I} \sum_{j \in J} |\mathbb{P}(U_i)\mathbb{P}(V_j) - \mathbb{P}(U_i \cap V_j)| \right),$$

where the supremum is taken over all the finite partitions $(U_i)_{i \in I}$ and $(V_j)_{j \in J}$ of Ω, respectively $\mathcal{U}$ and $\mathcal{V}$ measurable (Kolmogorov and Rozanov [15]).

Let $(X_i)_{i \in \mathbb{Z}}$ be a strictly stationary process of $\mathbb{R}$-valued random elements of a Polish space $\mathcal{X}$. If we denote $\mathcal{F}_0 = \sigma(X_j, j \leq 0)$ and $\mathcal{F}_l = \sigma(X_j, j \geq l)$, the β-mixing coefficient β_l is defined by $\beta_l = \beta(\mathcal{F}_0, \mathcal{F}_l)$ for any integer l. The process $(X_i)_{i \in \mathbb{Z}}$ is called absolutely regular (or β mixing) if the sequence $(\beta_l)_{l \geq 0}$ of its β-mixing coefficients tend to zero when l tends to infinity.

Let us introduce some notation. Let P be the distribution of X_0 on $\mathcal{X}$. For any measurable function h which is P-integrable, $\int_{\mathcal{X}} h dP$ is denoted by $E_P(h)$. We denote by ν_n the centered empirical process $\nu_n = \mathbb{P}_n - P$ where $\mathbb{P}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$ is the empirical measure. Finally, for $r \geq 2$, let $\mathcal{L}(r, \beta, P)$ be the set of functions $b : \mathcal{X} \to \bar{\mathbb{R}}_+$ such that

$$b = \sum_{l \geq 0} (l+1)^{r-2} b_l \text{ with } 0 \leq b_l \leq 1 \text{ and } E_P(b_l) \leq \beta_l.$$

Let us recall that B_r is the bound of the series $\sum_{l \geq 0} (l+1)^{r-2} \beta_l$.

The following lemma which gives an evaluation of the $\mathbb{L}_r(P)$-norm of the function b is very useful in the sequel. For more details about Lemma (2.2) and Theorem (2.1) we refer to Viennet [23].

LEMMA 2.2. – *Let $1 \leq r < \infty$. As soon as $B_{r+1} < \infty$, for any function b in $\mathcal{L}(2, \beta, P)$,*

$$E_P(b^r) \leq r B_{r+1} < \infty.$$

Finally, Theorem 2.1 provides a sharp control of the variance of a sum of absolutely regular variables and states a Rosenthal type inequality for absolutely regular variables.

THEOREM 2.1. – *Let $(X_i)_{i \in \mathbb{Z}}$ be a strictly stationary absolutely regular process with sequence of β-mixing coefficients $(\beta_l)_{l \geq 0}$.*

• *If the β-mixing coefficients satisfy the summability condition $B_2 < \infty$, there exists a function b in $\mathcal{L}(2, \beta, P)$, such that for any positive integer n and any measurable function $h \in \mathbb{L}_2(P)$,*

$$n^2 \ \mathit{Var}\,(\nu_n(h)) \quad \leq \quad 2\, n\, E_P(b\, h^2). \tag{9}$$

• *Let $r \geq 2$. If the β-mixing coefficients satisfy the summability condition $B_r < \infty$, there exist two functions b and b' in $\mathcal{L}(2, \beta, P)$ and $\mathcal{L}(r, \beta, P)$ respectively, such that for any measurable bounded function h*

$$\begin{aligned} n^r\, \mathbb{E}(|\nu_n(h)|^r) \leq C_1(r) n^{r/2} E_P^{r/2}(bh^2) \\ + C_2(r) n ||h||_\infty^{r-2} E_P(b'h^2) \end{aligned} \tag{10}$$

where $C_1(r)$ and $C_2(r)$ are positive constants depending only on r.

The following theorem is our first main result. We explain in some remarks how to use it for our statistical purpose.

THEOREM 2.2. – *Let K be a set of indices of cardinality D. Let $\{\phi_k\}_{k \in K}$ be a real basis such that: for any $m' \geq 1$, there exists a constant $C_{m'} > 0$ depending on m' such that for any $(a_k)_{k \in K} \in \mathbb{R}^D$:*

$$|| \sum_{k \in K} a_k \phi_k ||_{m'} \leq C_{m'}\ D^{1/2 - 1/m'} |a|_{m'}. \tag{11}$$

For any $r \geq 2$, we denote by $\mathcal{F}_r$ the class of functions defined by

$$\mathcal{F}_r = \left\{ \sum_{k \in K} a_k \phi_k ; \sum_{k \in K} |a_k|^{r'} \leq 1 \right\}, \qquad \frac{1}{r} + \frac{1}{r'} = 1. \tag{12}$$

Let $\frac{r}{2} \leq m < \infty$ and $(X_i)_{i \geq 0}$ be a strictly stationary absolutely regular process such that its sequence of β-mixing coefficients satisfies

$$B_{2m} < \infty. \tag{13}$$

We assume that the stationary distribution P_f of $(X_i)_{i \geq 0}$ admits a density f with respect to the Lebesgue measure and that f is uniformly bounded. Then, there exists a positive constant K_1 depending on r, B_r and C_∞ and there exists a function b in $\mathbb{L}_m(P_f)$ with $||b(X_0)||_m \leq m B_{m+1} < \infty$ such that, as soon as $D \leq n$, for any integer q, $q \leq n$, for any $\lambda_1 > 0$ and for any $\lambda_2 > 0$

$$\begin{aligned} &\mathbb{P}\left(\sup_{h \in \mathcal{F}_r} |\nu_n(h)| \geq \lambda_1 + K_r \frac{D^{\frac{1}{r}}}{\sqrt{n}} + \lambda_2 \right) \\ &\leq \exp\left(- \frac{K_1 n}{1 + ||f||_\infty} \left(\frac{\lambda_1^2 D^{-\frac{1}{m}}}{||b(X_0)||_m} \wedge \frac{\lambda_1}{q\sqrt{D} \wedge \frac{\sqrt{n}\lambda_1^2}{q D^{\frac{1}{r}+\frac{1}{2}}}} \right) \right) + \frac{2C_\infty}{\lambda_2} \sqrt{D} \beta_q \end{aligned}$$

where

$$K_r = \Big(R_r C_\infty^2 \Big((2^r + C_1(r))\|f\|_\infty^{r/2-1} B_{r/2+1} + C_2(r) C_\infty^2 B_r\Big)\Big)^{1/r},$$

with R_r is the Rosenthal constant, C_∞ is defined in (11) *and $C_1(r)$ and $C_2(r)$ are the same as in Theorem* 2.1.

According to Ledoux and Talagrand [16], the Rosenthal constant is smaller than 4^r. The following remarks will be detailed in the proof of Theorem 2.2.

Remark 1. – We will apply this theorem in our wavelet framework for the following class of functions

$$\mathcal{F}_r = \{h = \sum_k a_k \psi_{jk}; \sum_k |a_k|^{r'} \leq 1\}, \qquad \frac{1}{r} + \frac{1}{r'} = 1, \tag{14}$$

In that case, condition (11) is satisfied for any $m' \geq 1$ (*see* Lemma 2.1) with $C_{m'} = \| \sum_k |\psi_{0k}| \ \|_{m'}$.

Remark 2. – Let us comment the theorem in the classical case $r = 2$. The constant $K_r = K_2$ is then more tractable, namely $K_2 = C_\infty B_2^{1/2}$. When $(X_i)_{i\in\mathbb{Z}}$ is a uniformly mixing process (also called Φ-mixing) such that its sequence of mixing coefficients $(\Phi_l)_{l\geq 0}$ satisfies the summability condition $\sum_{l\geq 0} \Phi_l^{1/2} < \infty$, the function b may be taken as a constant, namely $b = \sum_{l\geq 0} \Phi_l^{1/2}$. Then, the conclusion of Theorem 2.2 is again valid for $m = \infty$ and with $\|b(X_0)\|_\infty = \|f\|_\infty \sum_{l\geq 0} \Phi_l^{1/2}$.

Remark 3. – When $(X_i)_{i\in\mathbb{Z}}$ is an independent process one can take $K_r = (R_r C_\infty^2)^{1/r} \|f\|_\infty^{1/2-1/r}$.

3. ESTIMATION

In the first part of this section, we present our estimation procedure. We give in Theorem 3.1 the main statistical result of the paper concerning the adaptation (in the sense defined in the introduction) of our estimator. We explain in the second part how to compute our estimator in practice.

3.1. Estimation and result

Let $\pi \geq 2$. We consider in the sequel $(X_i)_{i\geq 0}$ a strictly stationary absolutely regular process with β-mixing coefficients satisfying: $B_\pi < \infty$

where B_π is defined in (3). These assumptions are quite natural in view of applying Theorem 2.1 and Lemma 2.2. We assume moreover that the mixing coefficients are arithmetically decreasing, more precisely:

$$\exists \theta > \pi - 2 \, , \; \forall l \geq 0, \; \beta_l = O(l^{-(1+\theta)}).$$

Let P_f be the marginal distribution of the process, absolutely continuous with respect to the Lebesgue measure. We denote by f its unknown density to estimate. Because of the expansions (6) and (7), we consider the weighted estimator:

$$\hat{f} = \sum_k \hat{\alpha}_{j_0 k} \varphi_{j_0 k} + \sum_{j=j_0}^{j_1} \hat{\eta}_j \sum_k \hat{\zeta}_{j_k} \psi_{jk} \tag{15}$$

where $\hat{\alpha}$ and $\hat{\zeta}$ are the empirical coefficients:

$$\hat{\alpha}_{j_0 k} = \frac{1}{n} \sum_{i=1}^{n} \varphi_{j_0 k}(X_i) \qquad \hat{\zeta}_{jk} = \frac{1}{n} \sum_{i=1}^{n} \psi_{jk}(X_i). \tag{16}$$

and where $\hat{\eta}_j$ is a thresholding statistic:

$$\hat{\eta}_j = I\!I_{\{\sum_k |\hat{\zeta}_{jk}|^\pi > s_j^\pi\}}$$

for a threshold s_j to be determined below.

This estimator is nearly the same as the one introduced by Kerkyacharian *et al.* [14] in an independent and identically distributed framework. Under some Besov regular assumption on f, the first sum in (15) is an estimation of the low-frequency part of f and we choose the level $j_0(n) = 0$ in order to make the term of linear variance $\mathbb{E}\| \sum_k (\hat{\alpha}_{j_0 k} - \alpha_{j_0 k}) \varphi_{j_0 k} \|_\pi$ negligible in the global error. In the same spirit, the level $j_1(n) = O(\log_2(n))$ can also be chosen such that the bias term $\mathbb{E}\| \sum_{j \geq j_1} \sum_k \zeta_{jk} \psi_{jk} \|_\pi$ will never contribute in the global error.

We use the same idea as in the i.i.d. case to determine the threshold statistic: we keep all the details of the level j if, at this level, the l_π-norm $(\sum_k |\zeta_{jk}|^\pi)^{1/\pi}$ of the coefficients is greater than the threshold s_j. We have now to estimate the quantity $\sum_k |\zeta_{jk}|^\pi$. In the i.i.d. case, the study of the properties of the density estimator is based on the computation of the moments of the estimator of $\sum_k |\zeta_{jk}|^\pi$: it is then necessary to estimate $\sum_k |\zeta_{jk}|^\pi$ with its associated U-statistic because of the crossing terms. This method allows to choose $s_j^\pi = C \frac{2^j}{n^{\pi/2}}$ with $C = 1$. Since it is unreasonable to compute the moments of the U-statistic in a mixing setting, we use an other approach. The advantage is that we can use the natural estimator $\sum_k |\hat{\zeta}_{jk}|^\pi$; the price to pay for this simplicity of implementation is that the constant C is now depending on the quantities $||f||_\infty$ and B_π.

The main idea is contained in the following remark: the study of $\sum_k |\hat{\zeta}_{j,k} - \zeta_{j,k}|^\pi$ is linked to the study of the supremum of the empirical process on a suitable function class. Indeed, if $\mathcal{F}_\pi$ is the class of functions defined by (14) (with $r = \pi$), we have $|\hat{\zeta}_{j,k} - \zeta_{j,k}| = |\nu_n(\psi_{j,k})|$. It follows by duality arguments that

$$\sum_k |\hat{\zeta}_{j,k} - \zeta_{j,k}|^\pi = \sup_{a_k, \sum_k |a_k|^{\pi'} \leq 1} |\sum_k a_k \nu_n(\psi_{j,k})|^\pi = \sup_{h \in \mathcal{F}_\pi} |\nu_n(h)|^\pi \quad (17)$$

where $\pi^{-1} + \pi'^{-1} = 1$. This linearization is crucial in our mixing framework; indeed it allows to take advantage of Theorem 2.2. Let us now state our statistical result.

THEOREM 3.1. – *Let $\pi \geq 2$. We assume that f belongs to the class*

$$\mathcal{F}_{s,p,q}(M_1, M_2, B) = \{f \text{density}, f \in B_{s,p,q}, \\ \text{supp}\,(f) \subset [-B, B], ||f||_{s,p,q} \leq M_1, ||f||_\infty \leq M_2\}$$

where $1/p < s < N + 1$ and $q \geq 1$. Let

$$\hat{f} = \sum_k \hat{\alpha}_{j_0 k} \varphi_{j_0 k} + \sum_{j=j_0}^{j_1} I\!I_{\{\sum_k |\hat{\zeta}_{jk}|^\pi > s_j^\pi\}} \sum_k \hat{\zeta}_{j_k} \psi_{jk}$$

with:

$$2^{j_0} = 1, \quad n \leq 2^{j_1} \leq 2n,$$
$$s_j^\pi = (2K_\pi)^\pi \frac{2^j}{n^{\pi/2}},$$

where

$$K_\pi = \left(R_\pi C_\infty^2\, B_\pi \left(||f||_\infty^{\pi/2-1}(2^\pi + C_1(\pi)) + C_2(\pi)\right)\right)^{1/\pi}$$

with R_π is the Rosenthal constant, C_∞ is defined in (11) *and $C_1(\pi)$ and $C_2(\pi)$ are the same as in Theorem* 2.1.

Then for $s \in]1/p, N + 1[, q \in [1, +\infty], p \geq \pi$ and for mixing coefficients arithmetically decreasing such that

$$\theta > \pi - 2 + \frac{\pi^2}{2},$$

there exists a positive constant C which is an increasing function of M_1 and M_2 such that:

$$\mathbb{E}\|\hat{f} - f\|_\pi^\pi \leq C\, n^{-\frac{s\pi}{1+2s}}.$$

Let us recall (*see* Donoho *et al.* [7] for the density, Nemirovskii [18] for the regression), that the optimal theoretical rate of the minimax risk $R_n(\alpha)$ defined in (1) for the set of functions $\mathcal{F}_{spq}(M_1, B)$ is $-\frac{s\pi}{1+2s}$ for the loss function $\mathbb{L}_\pi, \pi \geq 2$. We immediately deduce the following corollary:

COROLLARY 3.1. – *Under the same assumptions as in Theorem 3.2, $\hat{f}$ is adaptive in the class $\{\mathcal{F}_{spq}(M_1, M_2, B), 1/p < s < N+1, p \geq \pi, q \geq 1, B > 0, M_1 > 0\}$.*

3.2. Practical estimation

When $\pi > 2$, $\hat{f}$ depends on the quantity M_2. We propose hereafter a new estimator $\tilde{f}$ the computation of which does not need prior knowledge of M_2. Let j_0 and j_1 be defined as previously. We assume now that f belongs to a Besov space $B_{s,p,\infty}$. Let S be a positive constant $(0 < S < \infty)$. We consider

$$\hat{M}_2 = C'_\infty 2^{j_1/2} \sup_k |\hat{\alpha}_{j_1,k}| + 2S.$$

We introduce the estimator $\tilde{f}$ and specify its adaptivity property in the following theorem.

THEOREM 3.2. – *Let $\pi \geq 2$. We assume that f belongs to the class*

$$\mathcal{F}_{s,p,q}(M_1, M_2, B) = \Big\{ f\,density, f \in B_{s,p,q},$$

$$\text{supp}\,(f) \subset [-B, B], \|f\|_{s,p,q} \leq M_1, \|f\|_\infty \leq M_2 \Big\}$$

where $1/p < s < N+1$ and $q \geq 1$. Let

$$\tilde{f} = \sum_k \hat{\alpha}_{j_0 k}\varphi_{j_0 k} + \sum_{j=j_0}^{j_1} I\!I_{\{\sum_k |\hat{\zeta}_{jk}|^\pi > \tilde{s}_j^\pi\}} \sum_k \hat{\zeta}_{jk}\psi_{jk}$$

with:

$$2^{j_0} = 1, \quad n \leq 2^{j_1} \leq 2n,$$

$$\tilde{s}_j = (2\tilde{K}_\pi)\frac{2^{j/\pi}}{\sqrt{n}},$$

where

$$\tilde{K}_\pi = \Big(R_\pi C_\infty^2 \Big((2^\pi + C_1(\pi))\hat{M}_2^{\pi/2-1} + C_2(\pi)\Big) B_\pi\Big)^{1/\pi}$$

with R_π *is the Rosenthal constant,* C_∞ *is defined in* (11) *and* $C_1(\pi)$ *and* $C_2(\pi)$ *are the same as in Theorem* 2.1.

Then for $s \in]1/p \vee \frac{1}{2(1+\theta)}, N+1[, p \geq \pi$ *and for mixing coefficients arithmetically decreasing such that*

$$\theta > \pi - 2 + \frac{\pi^2}{2},$$

$$\lim_{n \to +\infty} \lim_{\lambda \to +\infty} \mathbb{P}\Big(\|\tilde{f} - f\|_\pi^\pi \geq \lambda\, n^{-\frac{s\pi}{1+2s}}\Big) = 0.$$

We immediately deduce the following corollary:

COROLLARY 3.2. – *Under the same assumptions as in Theorem* 3.2 $\tilde{f}$ *is adaptive in probability in the class*

$$\{\mathcal{F}_{sp\infty}(M_1, M_2, B), 1/p \vee \frac{1}{2(1+\theta)} < s < N+1,$$

$$p \geq \pi, B > 0, M_1 > 0, M_2 > 0\}.$$

Remark. – We recall that the computation of $\tilde{f}$ only requires the knowledge of B_π, the upper bound for $\sum_{l \geq 0}(l+1)^{\pi-2}\beta_l$. The price to pay for the reduction of the prior knowledge is the diminution of the adaptive regularity bandwidth. In the independent setting ($\theta = \infty$), our approach shows that without any prior knowledge $\tilde{f}$ is adaptive in $1/p < s < N+1$.

4. PROOFS

C denotes throughout a constant whose value may change from line to line and may depend on $B_\pi, \pi, p, s, M_1, M_2$. The constant R_π denotes the Rosenthal constant of Lemma 4.3 and $C_1(\pi)$ and $C_2(\pi)$ are the contants introduced in Theorem 2.1. We first state some preliminary results used to establish the theorems.

4.1. Preliminary results

We first need a bound for the l_π-norm of wavelet coefficients at a fixed level j. The following lemma is a direct consequence of Theorem 2.1.

PROPOSITION 4.1. – *Let $\pi \geq 2$. Under the summability condition $B_\pi < \infty$, there exists constants C_1 and C_2 depending on B_π and $\|f\|_\infty$ such that for any j with $2^j \leq n$:*

$$\mathbb{E}|\hat{\alpha}_{j.} - \alpha_{j.}|_\pi^\pi \leq C_1\left(\frac{2^j}{n^{\pi/2}}\right) \quad \textit{and} \quad \mathbb{E}|\hat{\zeta}_{j.} - \zeta_{j.}|_\pi^\pi \leq C_2\left(\frac{2^j}{n^{\pi/2}}\right)$$

where $\hat{\alpha}$ and $\hat{\zeta}$ are the empirical wavelet coefficients.

In a second time, we derive from Theorem 2.2 a proposition which is its direct application to the wavelet framework.

PROPOSITION 4.2. – *Let $\pi \geq 2$. Assuming that the sequence of β mixing coefficients satisfies $B_\pi < \infty$, there exists a positive constant C such that, under the condition $\theta > \pi$, we have for any j*

$$\mathbb{P}\left(|\hat{\zeta}_{j.} - \zeta_{j.}|_\pi \geq 2K_\pi \frac{2^{j/\pi}}{n^{1/2}}\right) \leq C_2 n^{-\frac{\theta}{2}} 2^{-j(\theta+2)(\frac{1}{\pi}-\frac{1}{2})}(j)^{(1+\theta)}.$$

where

$$K_\pi = \left(R_\pi C_\infty^2 B_\pi \left(\|f\|_\infty^{\pi/2-1}(2^\pi + C_1(\pi)) + C_2(\pi)\right)\right)^{1/\pi}.$$

As a particular case, if j is such that $2^{j_0} = 1 \leq 2^j \leq 2^{j_s} = n^{\frac{1}{1+2s}}$:

$$\mathbb{P}\left(|\hat{\zeta}_{j.} - \zeta_{j.}|_\pi \geq 2K_\pi \frac{2^{j/\pi}}{n^{1/2}}\right) \leq C 2^{j_s \pi}\left(\frac{2^j}{n}\right)^{\pi/2}.$$

Finally, we use repeatedly in the proofs the following lemma which is simply due to a combination of the triangular inequality and of Lemma 2.1.

LEMMA 4.1. – *Let $\pi \geq 2$ and η be an appropriately chosen constant.*

$$\left\|\sum_{j_0}^{j_1}\sum_k \beta_{jk}\psi_{jk}\right\|_\pi^\pi \leq \begin{cases} 2^{j_1\eta\frac{\pi}{2}} \sum_{j_0}^{j_1} 2^{-j\eta\frac{\pi}{2}} \|\sum_k \beta_{jk}\psi_{jk}\|_\pi^\pi & \textit{if } \eta > 0 \\ 2^{j_0\eta\frac{\pi}{2}} \sum_{j_0}^{j_1} 2^{-j\eta\frac{\pi}{2}} \|\sum_k \beta_{jk}\psi_{jk}\|_\pi^\pi & \textit{if } \eta < 0 \end{cases}$$

4.2. Proof of Theorem 2.2

The scheme of this proof is quite classical. We take advantage of the absolute regularity using a corollary of Berbee's coupling lemma (Lemma 4.2). Thanks to this lemma, we approximate our original process by a sequence of independent variables. We decompose the centered empirical process into an error term and a centered empirical process associated to independent variables. The control of the first term is quite direct, whereas the control of the second one requires a closer attention.

We use in the following four essential results, the sub-cited corollary of Berbee's lemma, the following version of a Talagrand result (Talagrand [22], the Rosenthal inequality (Rosenthal [19]) and the moment inequality for absolutely regular variables (2.1). As we shall see, for $r = 2$ the proof is simpler, we use variance inequalities instead of Rosenthal inequalities.

LEMMA 4.2. – *Let $(X_i)_{i>0}$ be a sequence of random variables taking their values in a Polish space $\mathcal{X}$. Then, there exists a sequence $(X_i^*)_{i>0}$ of independent random variables such that for any positive integer i, X_i^* has the same distribution as X_i and*

$$\mathbb{P}(X_i \neq X_i^*) \leq \beta(\sigma(X_i), \sigma(X_j; j > i)).$$

The result of Talagrand is not stated in this form but one can actually write it as follows.

THEOREM 4.1. – *Let $X_1, \ldots, X_n$, be n independent identically distributed random variables, and $\mathcal{F}$ a family of functions that are uniformly bounded by some constant M_1. Let H and v be defined by*

$$H \geq \frac{1}{n}\mathbb{E}\left(\sup_{h\in\mathcal{F}}\left|\sum_{i=1}^{n} h(X_i)\right|\right), \qquad V = \mathbb{E}\left(\sup_{h\in\mathcal{F}}\sum_{i=0}^{n} h^2(X_i)\right).$$

Then, there exists universal constant K_1 such that for any $\lambda_1 > 0$,

$$\mathbb{P}(\sup_{h\in\mathcal{F}}|\nu_n(h)| \geq \lambda_1 + H) \leq \exp\left(-K_1 n\left(\frac{\lambda_1^2}{V} \wedge \frac{\lambda_1}{M_1}\right)\right).$$

Moreover, according to Corollary 3.4 in Talagrand [21], the following bound holds

$$V \leq v + 8M_1\tilde{H},$$

where v and $\tilde{H}$ are defined by

$$v \geq \sup_{h\in\mathcal{F}} \mathrm{Var}(h(X)), \qquad \tilde{H} \geq \frac{1}{n}\mathbb{E}\left(\sup_{h\in\mathcal{F}}\sum_{i=1}^{n} \epsilon_i h(X_i)\right),$$

with $\epsilon_1, \ldots, \epsilon_n$ n independent Rademacher variables. We derive then the following inequality: there exist a positive constant K_1 such that for any $\lambda_1 > 0$,

$$\mathbb{P}(\sup_{h\in\mathcal{F}}\nu_n(h) \geq \lambda_1 + H) \leq \exp\left(-K_1 n\left(\frac{\lambda_1^2}{v} \wedge \frac{\lambda_1}{M_1} \wedge \frac{\lambda_1^2}{8M_1\tilde{H}}\right)\right). \quad (19)$$

Let us now start the proof. For sake of simplicity, we assume in the following that $X_i = 0$ if $i > n$, and that $E_P(h) = 0$. It is also useful to notice that the condition (11) for $m' = \infty$ implies

$$\|\sum_k \psi_{j,k}^2\|_\infty \leq C_\infty^2 \, 2^j. \tag{20}$$

(*see* Birgé and Massart [3]). Let q be a fixed integer, $q = [\sqrt{n}]$, where $[\cdot]$ denotes the integer part. According to Lemma 4.2, we build a sequence of independent variables $(X_i^*)_{i\geq 0}$ such that $Y_k = (X_{qk+1}, \ldots, X_{q(k+1)})$ and $Y_k^* = (X_{qk+1}^*, \ldots, X_{q(k+1)}^*)$ fulfill the conditions: for any $k \geq 0$, Y_k and Y_k^* have the same distributions, for any $k \geq 0$, $\mathbb{P}(Y_k \neq Y_k^*) \leq \beta_q$, the random variables $(Y_{2k}^*)_{k\geq 0}$ are independent, and the $(Y_{2k+1}^*)_{k\geq 0}$ are independent. The centered empirical process $\nu_n(h)$ is then decomposed as

$$\nu_n(h) = \nu_n{}^*(h) + [\nu_n(h) - \nu_n{}^*(h)],$$

where $\nu_n{}^*(h)$ is the empirical process associated with the random variables $(X_i^*)_{i\geq 0}$. Since

$$\mathbb{P}\left(\sup_{h\in\mathcal{F}_r} |\nu_n(h)| \geq \lambda_1 + K_r \frac{D^{1/r}}{\sqrt{n}} + \lambda_2\right)$$
$$\leq \mathbb{P}\left(\sup_{h\in\mathcal{F}} |\nu_n(h) - \nu_n^*(h)| \geq \lambda_2\right) + \mathbb{P}\left(\sup_{h\in\mathcal{F}} |\nu_n^*(h)| \geq \lambda_1 + K_r \frac{D^{1/r}}{\sqrt{n}}\right),$$

we just have to control theses two quantities to prove Theorem 2.2. First, since

$$|\nu_n(h) - \nu_n^*(h)| \leq \frac{2}{n}\|h\|_\infty \sum_{i=1}^n 1\!\!1_{X_i \neq X_i^*},$$

according to the assumption (11) with $m = \infty$, we have

$$\mathbb{E}(|\nu_n(h) - \nu_n{}^*(h)|) \leq \frac{2}{n}\|h\|_\infty n\beta_q \leq 2C_\infty D^{1/2}\beta_q.$$

By Markov inequality, we deduce

$$\forall \lambda_2 > 0, \quad \mathbb{P}\left(\sup_{h\in\mathcal{F}} |\nu_n(h) - \nu_n^*(h)| \geq \lambda_2\right) \leq \frac{2C_\infty}{\lambda_2} D^{1/2}\beta_q. \tag{21}$$

Let us now control $\mathbb{P}\left(\sup_{h\in\mathcal{F}} |\nu_n{}^*(h)| \geq \lambda_1 + K_r \frac{D^{1/r}}{\sqrt{n}}\right)$. Let $p(n)$ be the greatest integer such that $n = qp(n) + x$ for $0 \leq x < q$. Then

$$\nu_n^*(h) = \frac{p(n)}{n}\left[\frac{1}{p(n)} \sum_{l=0}^{p(n)} \left(\sum_{i=2ql+1}^{q(2l+1)} h(X_i^*) + \sum_{i=q(2l+1)+1}^{q(2l+2)} h(X_i^*)\right)\right].$$

The control is made in two steps, considering odd terms and even ones. They are both treated in the same way; we only detail the even part. Since the variables $\left(\sum_{i=2ql+1}^{q(2l+1)} h(X_i^*)\right)_{0\leq l\leq p(n)}$ are independent by construction, we are allowed to apply the inequality (19) to $\nu_{p(n)}^* = \frac{1}{p(n)}\sum_{l=0}^{p(n)}\sum_{i=2ql+1}^{q(2l+1)} h(X_i^*)$ with adequate choices of λ_1, $M_1, H, \tilde{H}$ and v. Let $\mathcal{F}_r$ be the family of functions defined in (12). We have to determine the quantities $M_1, H, \tilde{H}$ and v. Since

$$\left\| \sum_{i=2ql+1}^{q(2l+1)} h(X_i^*) \right\|_\infty \leq C_\infty D^{1/2} q,$$

we put

$$M_1 = C_\infty D^{1/2} q. \tag{22}$$

Under the summability condition on the mixing coefficients $B_{2m} < \infty$, Theorem 2.1 and Lemma 2.2 ensure that there exists a function b belonging to $\mathcal{L}(2, \beta, P_f) \cap \mathbb{L}_m(P)$ such that, using the Hölder inequality with $\frac{1}{m} + \frac{1}{m'} = 1$ and a convexity inequality ($2m' \geq r'$), we have

$$\begin{aligned}
\operatorname{var}&\left(\sum_{i=2lq+1}^{q(2l+1)} h(X_i^*) - \int h dP\right)\\
&\leq 4q \int bf(\sum_{k\in K} a_k\phi_k(x))^2 d\mu\\
&\leq 4q\left(\int (bf)^m\right)^{1/m}\left(\int (\sum_{k\in K} a_k\phi_k)^{2m'}\right)^{1/m'}\\
&\leq 4qC_{2m'}^2 \|f\|_\infty^{\frac{m-1}{m}} \left(\int b^m f\right)^{1/m} D^{(1-\frac{1}{m'})}\left(\sum_{k\in K} a_k^{2m'}\right)^2\\
&\leq 4qC_{2m'}^2 \|f\|_\infty^{\frac{m-1}{m}} \|b(X_0)\|_m \left(\sum_{k\in K} |a_k|^{r'}\right)^{m'/r'} D^{\frac{1}{m}}.
\end{aligned}$$

Thus we set

$$v = 4qC_{2m'}^2 \|f\|_\infty^{\frac{m-1}{m}} \|b(X_0)\|_m \, D^{\frac{1}{m}}. \tag{23}$$

Let us now determine the quantity H. Applying Hölder's inequality with $\frac{1}{r} + \frac{1}{r'} = 1$ and Jensen's inequality, we have

$$\mathbb{E}\left[\sup_{h \in \mathcal{F}_r} \left| \sum_{l=0}^{p(n)} \sum_{i=2ql+1}^{q(2l+1)} \sum_{k \in K} a_k \phi_k(X_i^*) \right| \right]$$
$$\leq \mathbb{E}\left[\sup_{h \in \mathcal{F}_r} \left((\sum_{k \in K} |a_k|^{r'})^{1/r'} \left(\sum_{k \in K} \left| \sum_{l=0}^{p(n)} \sum_{i=2ql+1}^{q(2l+1)} \phi_k(X_i^*) \right|^r \right)^{1/r} \right) \right]$$
$$\leq \mathbb{E}^{1/r}\left[\sum_{k \in K} \left| \sum_{l=0}^{p(n)} \sum_{i=2ql+1}^{q(2l+1)} \phi_k(X_i^*) \right|^r \right].$$

In order to bound this last term, we use Rosenthal's inequality.

LEMMA 4.3. – *Let $Y_1, \cdots, Y_n$ be n real independent identically distributed variables, such that for any $0 \leq i \leq n$*

$$|Y_i| = K, \quad \mathbb{E}(Y_i) = 0, \ \text{Var}(Y_i) = \sigma^2.$$

Then, there exists a constant R_r such that for any $r \geq 2$

$$\mathbb{E}\left(\left| \sum_{i=1}^{n} Y_i \right|^r \right) \leq R_r \Big(n \|Y_1\|_r^r + n^{r/2} \sigma^r \Big).$$

We apply this lemma to the variables $\{Y_{l,k} = \sum_{i=2ql+1}^{q(2l+1)} \phi_k(X_i^*)\}_{0 \leq l \leq p(n)}$ which are independent and satisfy

$$\mathbb{E}(Y_{l,k}) = 0, \qquad |Y_{l,k}| \leq q\|\phi_k\|_\infty, \qquad \text{Var}\,(Y_{l,k}) \leq 4q \int bf\phi_k^2,$$

$$\|Y_{l,k}\|_r^r \leq C_1(r) q^{r/2} \Big(\int b^{r/2} f^{r/2} \phi_k^2 \Big) + C_2(r) q \|\phi_k\|_\infty^{r-2} E_f\big(b'\phi_k^2\big).$$

The variance is bounded by inequality (9) and the control of the $\mathbb{L}_r$-norm is obtained combining inequality (10) and Hölder inequality with $\left(\frac{r}{2}\right)^{-1} + \left(\frac{r}{r-2}\right)^{-1} = 1$:

$$\Big(\int bf\phi_k^2 \Big)^{r/2} \leq \Big(\int b^{r/2} f^{r/2} \phi_k^2 \Big) \Big(\int \phi_k^2 \Big)^{r/2-1}.$$

Thus,

$$\mathbb{E}\left[\sum_{k\in K}\left|\sum_{l=0}^{p(n)}\sum_{i=2ql+1}^{q(2l+1)}\phi_k(X_i^*)\right|^r\right]$$
$$\leq \sum_{k\in K}\left[C_1(r)R_r p(n)q^{r/2}\int b^{r/2}f^{r/2}\phi_k^2 + C_2(r)R_r p(n)q\|\phi_k\|_\infty^{r-2}E_f(b'\phi_k^2)\right]$$
$$+\sum_{k\in K}R_r 2^r q^{r/2}p(n)^{r/2}\left(\int bf\phi_k^2\right)^{r/2}$$
$$\leq C_1(r)R_r p(n)q^{r/2}DC_\infty^2\|f\|_\infty^{r/2-1}\int b^{r/2}f + C_2(r)R_r nD^{\frac{r-2}{2}}C_\infty^2 D\int b'f$$
$$+2^r n^{r/2}C_\infty^2 DR_r\|f\|_\infty^{r/2-1}\int b^{r/2}f,$$

and since $q \leq n$ and $D \leq n$,

$$\mathbb{E}\left[\sum_{k\in K}\left|\sum_{l=0}^{p(n)}\sum_{i=2ql+1}^{q(2l+1)}\phi_k(X_i^*)\right|^r\right]$$
$$\leq R_r C_\infty^2\left(\|f\|_\infty^{r/2-1}(2^r + C_1(r))\int b^{r/2}f + C_2(r)\int b'f\right)n^{r/2}D.$$

We deduce then

$$\frac{1}{p(n)}E_f\left(\sup_{h\in\mathcal{F}_r}\left|\sum_{l=0}^{p(n)}\sum_{i=2lq+1}^{q(2l+1)}h(X_i^*)\right|\right) \leq \frac{\sqrt{n}}{p(n)}D^{1/r}K_r,$$

and we take

$$H = K_r\sqrt{\frac{q}{p(n)}}D^{1/r} \tag{24}$$

with

$$K_r = \left(R_r C_\infty^2\left(\|f\|_\infty^{r/2-1}B_{r/2+1}\beta_l(2^r + C_1(r)) + C_2(r)B_r\right)\right)^{1/r}.$$

If $r = 2$, then $K_2 = C_\infty B_2^{1/2}$, indeed,

$$\mathbb{E}\left[\sup_{h\in\mathcal{F}_r}\left|\sum_{l=0}^{p(n)}\sum_{i=2ql+1}^{q(2l+1)}\sum_{k\in K}a_k\phi_k(X_i^*)\right|\right] \leq \mathbb{E}\left[\sum_{k\in K}\left|\sum_{l=0}^{p(n)}\sum_{i=2ql+1}^{q(2l+1)}\phi_k(X_i^*)\right|^r\right]^{1/2},$$

and by inequality (9) and assumption (11)

$$\mathbb{E}\left[\sum_{k\in K}\left|\sum_{l=0}^{p(n)}\sum_{i=2ql+1}^{q(2l+1)}\phi_k(X_i^*)\right|^2\right] \leq \sum_k p(n)q\int b\phi_k^2 f \leq C_\infty^2 nD\int bf.$$

Finally, we notice that

$$\frac{1}{n}\mathbb{E}\left(\sup_{h\in\mathcal{F}_r}\sum_{i=1}^n \epsilon_i h(X_i)\right) \leq K_r\sqrt{\frac{q}{p(n)}}D^{1/r}, \tag{25}$$

and we take $H = \tilde{H}$. Now, applying inequality (19) with M_1, v, H and $\tilde{H}$ defined in (22), (23), (24) and (25), and reminding the choice of $p(n)$, we obtain for any $m > r/2 - 1$

$$\mathbb{P}\left(\sup_{h\in\mathcal{F}_r}\left|\frac{1}{p(n)}\sum_{l=0}^{p(n)}\sum_{i=2ql+1}^{q(2l+1)}h(X_i^*)\right| \geq q\lambda_1 + K_r\sqrt{\frac{q}{p(n)}}D^{\frac{1}{r}}\right)$$
$$\leq \exp\left(-\frac{K_1 p(n)}{1+\|f\|_\infty}\left(\frac{q^2\lambda_1^2 D^{-1/m}}{q\|b(X_0)\|_m} \wedge \frac{q\lambda_1}{qD^{1/2}} \wedge \frac{q\lambda_1^2}{D^{1/2+1/r}}\sqrt{\frac{n}{q^2}}\right)\right),$$

where K_1 is a positive constant depending on r, B_r and C_∞. Similarly, we find for the odd part

$$\mathbb{P}\left(\sup_{h\in\mathcal{F}_r}\left|\frac{1}{p(n)}\sum_{l=0}^{p(n)}\sum_{i=q(2l+1)+1}^{q(2l+1)}h(X_i^*)\right| \geq q\lambda_1 + K_r\sqrt{\frac{q}{p(n)}}D^{\frac{1}{r}}\right)$$
$$\leq \exp\left(-\frac{K_1 p(n)}{1+\|f\|_\infty}\left(\frac{q\lambda_1^2 D^{-1/m}}{\|b(X_0)\|_m} \wedge \frac{\lambda_1}{D^{1/2}} \wedge \frac{q\lambda_1^2}{D^{1/2+1/p}}\sqrt{\frac{n}{q^2}}\right)\right).$$

Regrouping (21), (26), (27), the proof of Theorem 22 is complete.

Proof of Remark 2. – If $(X_i)_{i\in\mathbb{Z}}$ is a uniformly mixing process such that its sequence of mixing coefficients $(\Phi_l)_{l\geq 0}$ satisfies the summability condition $\sum_{l\geq 0}\Phi_l^{1/2} < \infty$, according to Viennet [23] the function b may be taken as a constant, namely $b = \sum_{l\geq 0}\Phi_l^{1/2}$. Thus, we set

$$v = 4qC_2^2\sum_{l\geq 0}\Phi_l^{1/2}\|f\|_\infty,$$

and

$$T_2 = C_\infty\sum_{l\geq 0}\Phi_l^{1/2}.$$

The conclusion of Theorem 2.2 is then valid for $m = \infty$ and with $\|b(X_0)\|_\infty = \|f\|_\infty \sum_{k\geq 0} \Phi_k^{1/2}$.

4.3. Proof of Proposition 4.2

We apply Theorem 2.2 to the basis $\{\psi_{jk}\}_k$ with

$$\pi/2 < m \leq (\pi - 1) \vee (1+\epsilon),\ \lambda_1 = \lambda_2 = \frac{K_\pi}{2} \frac{2^{j/\pi}}{n^{1/2}}$$

and $q = \left[\sqrt{n} 2^{j(1/\pi - 1/2)} (\eta j)^{-1}\right]$ where η is a positive constant to choose. Clearly under the summability condition on the mixing coefficients, $b(X_0) \in \mathbb{L}_m$. Thus, there exists some positive constants C and C' such that

$$\begin{aligned}
&\mathbb{P}\left(\sum_k |\hat{\zeta}_{jk} - \zeta_{jk}|^\pi \geq (2K_\pi)^\pi \frac{2^j}{n^{\pi/2}}\right) \\
&\leq \exp\left(-C\left(\frac{2^{j(2/\pi - 1/m)}}{\|b(X_0)\|_m} \wedge \frac{2^{j(1/\pi - 1/2)}}{q}\sqrt{n} \wedge \frac{2^{j(1/\pi-1/2)}}{q}\sqrt{n}\right)\right) \\
&+ C' 2^{j(1/2 - 1/\pi)} \sqrt{n} q^{-(1+\theta)} \\
&\leq \exp\left(-C\left(2^{j(2/\pi - 1/m)} \wedge \eta j\right)\right) + C' n^{-\theta/2} 2^{j(1 - \frac{2}{\pi} - \frac{\theta}{\pi} + \frac{\theta}{2})} (\eta j)^{1+\theta}.
\end{aligned}$$

Since $m > \pi/2$, if we choose η big enough, the exponential term is of smaller order than the mixing term and we obtain for some positive constant C_2:

$$\mathbb{P}\left(|\hat{\zeta}_{j.} - \zeta_{j.}|_\pi \geq 2K_\pi \frac{2^{j/\pi}}{n^{1/2}}\right) \leq C_2 n^{-\frac{\theta}{2}} 2^{-j(\theta+2)(\frac{1}{\pi} - \frac{1}{2})} (j)^{(1+\theta)}.$$

In order to complete the proof, we have to bound this term by a quantity of order $2^{js\pi}\left(\frac{2^j}{n}\right)^{\pi/2}$. We have to prove that:

$$\forall j_0 \leq j \leq j_s, \quad n^{(\pi-\theta)} 2^{-j((\theta+2)(2/\pi - 1) + \pi(2s+1))} = o(1)$$

Under the assumption $\theta > \pi$, the exponent of 2^j is positive and we just have to verify the inequality for $j = j_s$. Since $\theta > \pi - 2$, we obtain the assertion.

4.4. Proof of Proposition 4.1

We only detail $\mathbb{E}\left[|\hat{\alpha}_{j.} - \alpha_{j.}|_\pi^\pi\right]$, the other term is bounded exactly in the same way. This lemma is a direct application of the first part of

Theorem 2.1. There exists two functions b and b' respectively in $\mathcal{L}(2, \beta, P_f)$ and $\mathcal{L}(\pi, \beta, P_f)$, and a positive constant C depending on π such that

$$\mathbb{E}\left[|\hat{\alpha}_{j.} - \alpha_{j.}|_\pi^\pi\right] \le n^{-\pi} \sum_k \mathbb{E}\left(\left|\sum_{i=1}^n \varphi_{j,k}(X_i) - E_f(\varphi_{j,k})\right|^\pi\right)$$
$$\le C n^{-\pi} \sum_k \left(n^{\pi/2} E_f^{\pi/2}(b\varphi_{j,k}^2) + n\|\varphi_{j,k}\|_\infty^{\pi-2} E_f\left(b'\varphi_{j,k}^2\right)\right).$$

Using (20) and Hölder inequality, we deduce

$$\mathbb{E}\left[|\hat{\alpha}_{j.} - \alpha_{j.}|_\pi^\pi\right] \le C n^{-\pi/2} \|f\|_\infty^{\pi/2-1} 2^j E_f(b^{\pi/2}) + C n^{-\pi+1} 2^{j\pi/2} E_f(b')$$
$$\le C \frac{2^j}{n^{\pi/2}} \|f\|_\infty^{\pi/2-1} B_\pi.$$

4.5. Proof of Theorem 3.1

Let us denote

$$C_1 = \|\sum_k |\psi_{0k}|\,\|_1^{1/\pi'} \|\sum_k |\psi_{0k}|\,\|_\infty^{1/\pi}, \qquad C_2 = \|\sum_k |\psi_{0k}|\,\|_p$$

where $\frac{1}{\pi} + \frac{1}{\pi'} = 1$ and let j_s be such that $2^{j_s} = \left(\frac{C_1 C_2^2 M_1}{2K_\pi}\right)^{\frac{2}{1+2s}} n^{\frac{1}{1+2s}}$.

Classically, we decompose $\mathbb{E}\|\hat{f} - f\|_\pi^\pi$ into three terms: a bias term, a linear stochastic term and a non linear stochastic term:

$$\mathbb{E}\|\hat{f} - f\|_\pi^\pi \le 3^{\pi-1} \Bigg[\mathbb{E}\|\sum_k (\hat{\alpha}_{j_0 k} - \alpha_{j_0 k})\varphi_{j_0 k}\|_\pi^\pi + \|\sum_{j_1 < j} \sum_k \zeta_{jk}\psi_{jk}\|_\pi^\pi$$
$$+ \mathbb{E}\|\sum_{j_0 \le j \le j_1} \sum_k (\hat{\zeta}_{jk}\, \mathbb{I}_{\{\sum_k |\hat{\zeta}_{jk}|^\pi \ge s_j^\pi\}} - \zeta_{jk})\psi_{jk}\|_\pi^\pi\Bigg].$$

We have to prove that each term of the above bound is bounded by $C\, n^{-\frac{s\pi}{1+2s}}$.

1. Bias term: it is the same term as for the independent variables case. By a direct application of the caracterization (8) of Besov spaces, we have:

$$\|\sum_{j_1 < j} \sum_k \zeta_{jk}\psi_{jk}\|_\pi^\pi \le \left(\sum_{j_1 \le j} \|\sum_k \zeta_{jk}\psi_{jk}\|_\pi\right)^\pi$$
$$\le \left(\sum_{j_1 \le j} \|\sum_k \zeta_{jk}\psi_{jk}\|_p\right)^\pi$$
$$\le (M\, 2^{-j_1 s})^\pi \le C n^{-s\pi} \le C n^{-\frac{s\pi}{1+2s}}.$$

2. Linear stochastic term: using Lemma 2.1 and Proposition 4.1 for the $\hat{\alpha}_{jk}$, we have:

$$\mathbb{E}\|\sum_k(\hat{\alpha}_{j_0k}-\alpha_{j_0k})\varphi_{j_0k}\|_\pi^\pi \leq C\,\mathbb{E}\Big[2^{j_0(\frac{1}{2}-\frac{1}{\pi})}|\hat{\alpha}_{j_0.}-\alpha_{j_0.}|_\pi\Big]^\pi$$
$$\leq C\,2^{j_0(\frac{\pi}{2}-\frac{\pi}{\pi})}\left(\frac{2^{j_0}}{n^{\pi/2}}\right)^{\pi/\pi} \leq Cn^{-\frac{s\pi}{1+2s}}.$$

3. Non linear stochastic term: we decompose this term into four terms.

$$\mathbb{E}\|\sum_{j_0\leq j\leq j_1}\sum_k(\hat{\zeta}_{jk}\,\mathbb{1}_{\{\sum_k|\hat{\zeta}_{jk}|^\pi\geq s_j^\pi\}}-\zeta_{jk})\psi_{jk}\|_\pi^\pi \leq 4^{\pi-1}(\mathbb{E}_1+\mathbb{E}_2+\mathbb{E}_3+\mathbb{E}_4).$$

where

$$\mathbb{E}_1=\mathbb{E}\|\sum_{j_0\leq j\leq j_s}\sum_k(\hat{\zeta}_{jk}-\zeta_{jk})\psi_{jk}\,\mathbb{1}_{\{\sum_k|\hat{\zeta}_{jk}|^\pi\geq s_j^\pi\}}\|_\pi^\pi,$$
$$\mathbb{E}_2=\mathbb{E}\|\sum_{j_0\leq j< j_s}\sum_k(\hat{\zeta}_{jk}-\zeta_{jk})\psi_{jk}\,\mathbb{1}_{\{\sum_k|\hat{\zeta}_{jk}|^\pi\leq s_j^\pi\}}\|_\pi^\pi,$$
$$\mathbb{E}_3=\mathbb{E}\|\sum_{j_s\leq j\leq j_1}\sum_k(\hat{\zeta}_{jk}-\zeta_{jk})\psi_{jk}\mathbb{1}_{\{\sum_k|\hat{\zeta}_{jk}|^\pi\geq 2^\pi s_j^\pi\}}\|_\pi^\pi,$$
$$\mathbb{E}_4=\mathbb{E}\|\sum_{j_s\leq j\leq j_1}\sum_k(\hat{\zeta}_{jk}-\zeta_{jk})\psi_{jk}\mathbb{1}_{\{\sum_k|\hat{\zeta}_{jk}|^\pi< 2^\pi s_j^\pi\}}\|_\pi^\pi.$$

Study of the term $\mathbb{E}_4=\mathbb{E}\|\sum_{j=j_s}^{j_1}\sum_k(\hat{\zeta}_{jk}-\zeta_{jk})\psi_{jk}\mathbb{1}_{\{\sum_k|\hat{\zeta}_{jk}|^\pi<2^\pi s_j^\pi\}}\|_\pi^\pi$**:** in the same way as for the bias term, we can derive the upper bound:

$$\mathbb{E}_4=\mathbb{E}\|\sum_{j=j_s}^{j_1}\sum_k\zeta_{jk}\psi_{jk}\mathbb{1}_{\{\sum_k|\hat{\zeta}_{jk}|^\pi<2^\pi s_j\}}\|_\pi^\pi$$
$$\leq\|\sum_{j=j_s}^{j_1}\sum_k\zeta_{jk}\psi_{jk}\|_\pi^\pi\leq\|\sum_{j=j_s}^{j_1}\sum_k\zeta_{jk}\psi_{jk}\|_p^\pi$$
$$\leq C2^{-j_s s\pi}\leq Cn^{-\frac{s\pi}{1+2s}}.$$

Study of the term $\mathbb{E}_1=\mathbb{E}\|\sum_{j=j_0}^{j_s}\sum_k(\hat{\zeta}_{jk}-\zeta_{jk})\psi_{jk}\mathbb{1}_{\{\sum_k|\hat{\zeta}_{jk}|^\pi\geq s_j^\pi\}}\|_\pi^\pi$**:** we bound this term as the linear stochastic term. Using Lemma 4.1 for some $\eta > 0$, Lemma 4.1 and Proposition 4.1 for the $\hat{\zeta}_{jk}$, we get:

$$\begin{aligned}
\mathbb{E}_1 &\leq 2^{j_s \eta\pi} \mathbb{E}\left(\sum_{j=j_0}^{j_s} 2^{-j\eta\pi} \| \sum_k (\hat{\zeta}_{jk} - \zeta_{jk})\psi_{jk} \|_\pi^\pi \right) \\
&\leq C\ 2^{j_s\eta\pi} \sum_{j=j_0}^{j_s} 2^{-j\eta\pi} 2^{j(\frac{\pi}{2} - \frac{\pi}{\pi})} \mathbb{E}\left(|\hat{\zeta}_{j.} - \zeta_{j.}|_\pi^\pi \right) \\
&\leq C\ 2^{j_s\eta\pi} \sum_{j=j_0}^{j_s} 2^{-j\eta\pi} 2^{j(\frac{\pi}{2} - \frac{\pi}{\pi})} \left(\frac{2^j}{n^{\pi/2}} \right)^{\pi/\pi} \\
&\leq C \left(\frac{2^{j_s}}{n} \right)^{\pi/2} \leq\ C n^{\frac{s\pi}{1+2s}}.
\end{aligned}$$

Study of the term $\mathbb{E}_2 = \mathbb{E} \| \sum_{j=j_0}^{j_s} \sum_k (\hat{\zeta}_{jk} - \zeta_{jk})\psi_{jk} \mathbb{1}_{\{\sum_k |\hat{\zeta}_{jk}|^\pi < s_j^\pi\}} \|_\pi^\pi$**:** using Lemma 4.1 for some $\eta > 0$ and Lemma 2.1, we get:

$$\begin{aligned}
\mathbb{E}_2 &\leq 2^{j_s\eta\pi} \sum_{j=j_0}^{j_s} 2^{-j\eta\pi} \| \sum_k \zeta_{jk}\psi_{jk} \|_\pi^\pi \mathbb{E}\left(I\!I_{\{\sum_k |\hat{\zeta}_{jk}|^\pi < s_j^\pi\}} \right) \\
&\leq 2^{j_s\eta\pi} \sum_{j=j_0}^{j_s} 2^{-j\eta\pi} \| \sum_k \zeta_{jk}\psi_{jk} \|_\pi^\pi \\
&\times \mathbb{E}\left[I\!I_{\{\sum_k |\hat{\zeta}_{jk}|^\pi < s_j^\pi\}} \left(I\!I_{\{\sum_k |\zeta_{jk}|^\pi < 2^\pi s_j^\pi\}} + I\!I_{\{\sum_k |\zeta_{jk}|^\pi \geq 2^\pi s_j^\pi\}} \right) \right] \\
&\leq \mathbb{E}_{21} + \mathbb{E}_{22}.
\end{aligned}$$

On the one hand:

$$\begin{aligned}
\mathbb{E}_{21} &\leq C 2^{j_s\eta\pi} \sum_{j=j_0}^{j_s} 2^{-j\eta\pi} \| \sum_k \zeta_{jk}\psi_{jk} \|_\pi^\pi \left(I\!I_{\sum_k |\zeta_{jk}|^\pi < 2^\pi s_j^\pi\}} \right) \\
&\leq C 2^{j_s\eta\pi} \sum_{j=j_0}^{j_s} 2^{-j\eta\pi} 2^{j(\frac{\pi}{2} - \frac{\pi}{\pi})} \left(\sum_k |\zeta_{jk}|^\pi \right)^{\pi/\pi} I\!I_{\{\sum_k |\zeta_{jk}|^\pi < 2^\pi s_j^\pi\}} \\
&\leq C 2^{j_s\eta\pi} \sum_{j=j_0}^{j_s} 2^{-j\eta\pi} 2^{j(\frac{\pi}{2} - \frac{\pi}{\pi})} 2^\pi (s_j^\pi)^{\pi/\pi} \\
&\leq C 2^{j_s\eta\pi} 2^\pi \sum_{j=j_0}^{j_s} 2^{-j\eta\pi} 2^{j(\frac{\pi}{2} - \frac{\pi}{\pi})} \left(\frac{2^j}{n^{\pi/2}} \right)^{\pi/\pi} \\
&\leq\ C \left(\frac{2^{j_s}}{n} \right)^{\pi/2} \leq\ C n^{-\frac{s\pi}{1+2s}}.
\end{aligned}$$

On the other hand, according to the triangular inequality

$$\mathbb{1}_{\{\sum_k |\hat{\zeta}_{jk}|^\pi < s_j^\pi\}} \mathbb{1}_{\{\sum_k |\zeta_{jk}|^\pi \geq 2^\pi s_j^\pi\}} \leq \mathbb{1}_{\{\sum_k |\hat{\zeta}_{jk} - \zeta_{jk}|^\pi \geq s_j^\pi\}}$$

and using the caracterization (8) of the Besov spaces and applying then Proposition 4.2, we get:

$$\begin{aligned}
\mathbb{E}_{22} &= C 2^{j_s \eta \pi} \sum_{j=j_0}^{j_s} 2^{-j\eta\pi} \| \sum_k \zeta_{jk} \psi_{jk} \|_\pi^\pi \\
&\times \mathbb{E}\left(\mathbb{1}_{\{\sum_k |\hat{\zeta}_{jk}|^\pi < s_j^\pi\}} \mathbb{1}_{\{\sum_k |\zeta_{jk}|^\pi \geq 2^\pi s_j^\pi\}} \right) \\
&\leq C 2^{j_s \eta \pi} \sum_{j=j_0}^{j_s} 2^{-j\eta\pi} \| \sum_k \zeta_{jk} \psi_{jk} \|_p^\pi \mathbb{E}\left(\mathbb{1}_{\{\sum_k |\hat{\zeta}_{jk} - \zeta_{jk}|^\pi \geq s_j^\pi\}} \right) \\
&\leq C 2^{j_s \eta \pi} \sum_{j=j_0}^{j_s} 2^{-j\eta\pi} 2^{-js\pi} \mathbb{P}\left(\sum_k |\hat{\zeta}_{jk} - \zeta_{jk}|^\pi \geq s_j^\pi \right) \\
&\leq C 2^{j_s \eta \pi} \sum_{j=j_0}^{j_s} 2^{-j\eta\pi} 2^{-js\pi} \left[2^{js\pi} \left(\frac{2^j}{n} \right)^{\pi/2} \right] \\
&\leq C \left(\frac{2^{j_s}}{n} \right)^{\pi/2} \leq C n^{-\frac{s\pi}{1+2s}}.
\end{aligned}$$

Study of the term $\mathbb{E}_3 = \mathbb{E} \| \sum_{j=j_s}^{j_1} \sum_k (\hat{\zeta}_{jk} - \zeta_{jk}) \psi_{jk} \mathbb{1}_{\{\sum_k |\hat{\zeta}_{jk}|^\pi \geq 2^\pi s_j^\pi\}} \|_\pi^\pi$: using the caracterization (8) of the Besov spaces and the definition of j_s, we have, for all j in $[j_s, j_1]$:

$$|\zeta_{j.}|_p \leq M_1 C_2 2^{-js} 2^{j(1/p-1/2)} \leq M_1 C_2 2^{-j_s(s+1/2)} 2^{j/p} \leq \frac{2K_\pi}{C_1 C_2} \frac{2^{j/p}}{n^{1/2}}. \tag{28}$$

Let us now applying Lemma 2.1:

$$C_1^{-1} 2^{j(\frac{1}{2}-\frac{1}{\pi})} |\zeta_{j.}|_\pi \leq \| \sum_k \zeta_{jk} \psi_{jk} \|_\pi \leq \| \sum_k \zeta_{jk} \psi_{jk} \|_p \leq C_2 2^{j(\frac{1}{2}-\frac{1}{p})} |\zeta_{j.}|_p,$$

which imply

$$\sum_k |\zeta_{jk}|^\pi \leq (C_1 C_2)^\pi 2^{j(1-\frac{\pi}{p})} (\sum_k |\zeta_{jk}|^p)^{\frac{\pi}{p}}. \tag{29}$$

Combining (28) and (29), we deduce then

$$\text{for } j \geq j_s, \qquad \sum_k |\zeta_{jk}|^\pi \leq (2K_\pi)^\pi \frac{2^j}{n^{\pi/2}} = s_j^\pi.$$

According to the triangular inequality

$$\mathbb{I}_{\{\sum_k |\hat{\zeta}_{jk}|^\pi \geq 2^\pi s_j^\pi\}} \mathbb{I}_{\{\sum_k |\zeta_{jk}|^\pi < s_j^\pi\}} \leq \mathbb{I}_{\{\sum_k |\hat{\zeta}_{jk} - \zeta_{jk}|^\pi \geq s_j^\pi\}}$$

and using Lemma 4.1 for some $\eta > 0$, Lemma 2.1 and the Holder inequality for $\frac{1}{m} + \frac{1}{m\prime} = 1$, we get:

$$\mathbb{E}_3 \leq C 2^{j_1 \eta \pi} \sum_{j=j_s}^{j_1} 2^{-j\eta\pi} 2^{j(\frac{\pi}{2}-1)} \mathbb{E}\Big(|\hat{\zeta}_{j.} - \zeta_{j.}|_\pi^\pi \mathbb{I}_{\{\sum_k |\hat{\zeta}_{jk} - \zeta_{jk}|^\pi \geq s_j^\pi\}} \Big)$$

$$\leq C 2^{j_1 \eta \pi} \sum_{j=j_s}^{j_1} 2^{-j\eta\pi} 2^{j(\frac{\pi}{2}-1)} \mathbb{E}\Big(|\hat{\zeta}_{j.} - \zeta_{j.}|_\pi^{\pi m\prime} \Big)^{1/m\prime} P\Big(|\hat{\zeta}_{j.} - \zeta_{j.}|_\pi \geq s_j \Big)^{1/m}$$

$$\leq C 2^{j_1 \eta \pi} \sum_{j=j_s}^{j_1} 2^{-j\eta\pi} 2^{j(\frac{\pi}{2}-1)} \left(\frac{2^j}{n^{\pi/2}} \right) \left(n^{-\frac{\theta}{2}} 2^{-j(\theta+2)(\frac{1}{\pi}-\frac{1}{2})} (\eta j)^{(1+\theta)} \right)^{1/m}$$

$$\leq C 2^{j_1 (\frac{1}{m}(\theta+2)(\frac{1}{2}-\frac{1}{\pi}) + \frac{\pi}{2})} (\eta j_1)^{\frac{1}{m}(1+\theta)} n^{-(\frac{\pi}{2}+\frac{\theta}{2m})}$$

This quantity is bounded by $C n^{-\frac{s\pi}{1+2s}}$ as soon as $\theta > \frac{s\pi^2}{1+2s} + \pi - 2$. Using the fact that $s < N+1$ and choosing and $m > 1$ as small as possible, we have the result under the assumption $\theta > \frac{(N+1)m\pi\pi}{2(N+1)+1} + \pi - 2$. The choice $m = 1 + \frac{1}{2(N+1)}$ completes the proof of Theorem 3.1.

4.6. Proof of Theorem 3.2

We first state a lemma which is proved immediately after. It ensures that, for an adequate choice of j_1, the statistic $\hat{M}_2$ is bounded in probability. According to Lemma 2.1, we define $\tilde{M}_2$ an upper bound for $\|f\|_\infty$ by

$$\|f\|_\infty \leq C_\infty \sum_{j=j_1}^{\infty} 2^{j/2} \sup_k |\zeta_{j,k}| + C'_\infty 2^{j_1/2} \sup_k |\alpha_{j_1,k}| = \tilde{M}_2$$

where $C_\infty = \| \sum_k |\psi_{0k}| \|_\infty$ and $C'_\infty = \| \sum_k |\phi_{0k}| \|_\infty$.

LEMMA 4.4. – *Let* $n^{\frac{1}{1+2\gamma_1}} \leq 2^{j_1} \leq 2n^{\frac{1}{1+2\gamma_1}}$ *where* $\gamma_1 > \frac{1}{2(1+\theta)}$. *Then,*

$$\lim_{n\to\infty} P\Big(|\hat{M}_2 - \tilde{M}_2| \geq \frac{1}{2} \tilde{M}_2 \Big) = 0$$

We have the following decomposition

$$
\begin{aligned}
&P\Big(\|\tilde f - f\|_\pi \ge \lambda n^{-s/(2s+1)}\Big)\\
&\le P\Big(\|\tilde f - f\|_\pi \ge \lambda n^{-s/(2s+1)}, |\hat M_2 - \tilde M_2| \ge \tilde M_2/2\Big)\\
&\quad + P\Big(\|\tilde f - f\|_\pi \ge \lambda n^{-s/(2s+1)}, |\hat M_2 - \tilde M_2| \le \tilde M_2/2\Big)\\
&\le P\Big(|\hat M_2 - \tilde M_2| \ge \tilde M_2/2\Big)\\
&\quad + P\Big(\|\tilde f - f\|_\pi \ge \lambda n^{-s/(2s+1)}, \tilde M_2/2 \le \hat M_2 \le 3/2\tilde M_2\Big)
\end{aligned}
$$

According to Lemma 4.4, the first term tends to 0. Let us notice that the inequality $\tilde M_2/2 \le \hat M_2 \le 3/2\tilde M_2$ implies that there exists $s_j^{\max}$ of the same order as s_j such that $s_j \le \hat s_j \le s_j^{\max}$. Thus, Theorem 3.1 ensures that the second term also tends to 0.

Proof of Lemma 4.4. – Let $\pi > 2$. Since j_1 is a large index of resolution, for n big enough

$$C_\infty \sum_{j=j_1}^{\infty} 2^{j/2} \sup_k |\zeta_{j,k}| \le S \le 1.$$

We first notice that by remark (17) and Markov inequality

$$
\begin{aligned}
&P\left(|\hat M_2 - \tilde M_2| \ge \frac{1}{2}\tilde M_2\right)\\
&\le P\left(C'_\infty 2^{j_1/2} \sup_k |\hat\alpha_{j_1 k} - \alpha_{j_1 k}| \ge \frac{1}{2} C'_\infty 2^{j_1/2} \sup_k |\alpha_{j_1 k}| + 1\right)\\
&\le C 2^{j_1/2} \mathbb{E}\left(\sup_k |\nu_n(\varphi_{j_1 k})|\right).
\end{aligned}
\tag{30}
$$

To control this term we introduce the following lemma derived from Lemma 4 of Doukhan *et al.* [9] combined with inequality (9).

LEMMA 4.5. – *Let a and δ be positive reals and $\mathcal{G}$ be a finite subclass of $\mathcal{L}_{2,\beta}(P)$ satisfying the following assumptions*

(i) *for any* $g \in \mathcal{G}$, $\mathbb{E}(g(X_0)) = 0$;

(ii) *for any* $g \in \mathcal{G}$, $\|g\|_\infty \le a$ *and* $\int bg^2 dP \le \delta^2$.

Let $L(\mathcal{G}) = \max(1, \log|\mathcal{G}|)$, where $|\mathcal{G}|$ *denotes the cardinality of* $\mathcal{G}$. *There exists some universal positive constant C such that, for any* $q \in [1, n]$

$$\mathbb{E}\left(\sup_{g\in\mathcal{G}} |\nu_n(g)|\right) \le \frac{C}{\sqrt{n}}\left(\delta\sqrt{L(\mathcal{G})} + aq\frac{L(\mathcal{G})}{\sqrt{n}} + a\beta_q\sqrt{n}\right). \tag{31}$$

We apply this lemma to the class of function $\mathcal{G} = \{\varphi_{j_1 k}\}_k$ with

$$\mathrm{card}(\mathcal{G}) \le C 2^{j_1}, \quad a = 2^{j_1/2}, \quad \delta = C\Big(2^{j_1(\frac{1}{\pi-1})}\Big). \tag{32}$$

Let us detail the computation of δ. Since $B_\pi < \infty$, $b \in \mathbb{L}_{\pi-1}$ thus, by Hölder's inequality with $1/(\pi-1) + (\pi-2)/(\pi-1) = 1$ we have

$$\left(\int b\varphi_{j_1 k}^2 f\right)^{1/2} \le \left(\int b^{\pi-1} f^{\pi-1}\right)^{1/2(\pi-1)} \left(\int \varphi_{j_1 k}^{2\frac{\pi-1}{\pi-2}}\right)^{\frac{\pi-2}{2(\pi-1)}}$$

$$\le C_{2\frac{\pi-1}{\pi-2}} 2^{j_1(\frac{1}{2} - \frac{\pi-2}{2(\pi-1)})} \|b\|_{\pi-1}^{1/2} \|f\|_\infty^{\frac{\pi-2}{2(\pi-1)}}.$$

Finally, applying (31) with the choices (32), inequality (30) becomes

$$P\Big(|\hat{M}_2 - \tilde{M}_2| \ge \frac{1}{2}\tilde{M}_2\Big) \le C\frac{\sqrt{j_1}}{\sqrt{n}} 2^{j_1 \frac{\pi}{2(\pi-1)})} + C\frac{2^{j_1}}{n} q j_1 + 2^{j_1}\beta_q.$$

Choosing $q = C n^{\frac{1}{2+\theta}} j_1^{-1}$,

$$P(|\hat{M}_2 - \tilde{M}_2| \ge \frac{1}{2}\tilde{M}_2) \le C\left(\frac{\sqrt{j_1}}{\sqrt{n}} 2^{j_1 \frac{\pi}{2(\pi-1)}} + 2^{j_1} n^{-\frac{1+\theta}{2+\theta}} j_1^{1+\theta}\right),$$

and by the definition of j_1

$$\lim_{n\to\infty} \left(\frac{\sqrt{j_1}}{\sqrt{n}} 2^{j_1 \frac{\pi}{2(\pi-1)}} + 2^{j_1} n^{-\frac{1+\theta}{2+\theta}} j_1^{1+\theta}\right) = 0.$$

ACKNOWLEDGMENTS

We would like to thank Pascal Massart for constructive suggestions and warm encouragement.

REFERENCES

[1] P. ANGO NZE and P. DOUKHAN, Functional estimation for time series : a general approach, Prépublication, Université Paris Sud, France, Vol. **93-43**, 1993.
[2] H. C. P. BERBEE, Random walks with stationary increments and renewal theory, *Cent. Math. Tracts*, Amsterdam, 1979.
[3] L. BIRGÉ and P. MASSART, Rates of convergence for minimum contrast estimators, *Theor. Probab. Appl.*, Vol. **97**, 1993, pp. 113-150.
[4] L. BIRGÉ and P. MASSART, From model selection to adaptative estimation, Prépublication, Université Paris Sud, France, Vol. **95-41**, 1995.

[5] R. Dalhaus, M. Neumman and R. von Sachs, Nonlinear wavelet estimation of time varying autoregressive processes, WIAS, Vol. **159**, 1995.

[6] I. Daubechies, Orthogonal bases of compactly supported wavelets, *Comm. in Pure and Applied Math.*, Vol. **41**, 1988, pp 906-996.

[7] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian and D. Picard, Density estimation by wavelet thresholding, *Annals of Stat.*, Vol. **24**, 1996, pp. 508-539.

[8] P. Doukhan, *Mixing : properties and examples*, Lecture Notes in Statistics, Springer Verlag, Vol. **72**, 1994.

[9] P. Doukhan, P. Massart and E. Rio, The functional central limit theorem for strongly mixing processes, *Ann. Inst. Henri Poincaré, Probab. Stat.*, Vol. **30**, 1994, pp. 63-82.

[10] S. Y. Efroimovitch, Nonparametric estimation of a density of unknown smoothness, *Theory Prob. Appl*, Vol. **30**, 1985, pp. 557-661.

[11] S. Y. Efroimovich and M. S. Pinsker, Learning algorithm for nonparametric filtering, Automat. Remote Control, Vol. **11**, 1984, pp. 1434-1440.

[12] M. Hoffmann, On adaptive estimation in nonlinear AR(1) models, 1997 (Submitted).

[13] G. Kerkyacharian and D. Picard, Density estimation in Besov Spaces, Statistics and Probability Letters, Vol. **13**, 1992a pp. 15-24.

[14] G. Kerkyacharian, D. Picard and K. Tribouley, *L_p adaptative density estimation*, Bernoulli, Vol. **2**, 1996, pp. 229-247.

[15] A. N. Kolmogorov and Y. A. Rozanov, On the strong mixing conditions for stationary gaussian sequences, *Theor. Probab. Appl.*, Vol. **5**, 1960, pp. 204-207.

[16] M. Ledoux and M. Talagrand, Probability in Banach spaces, *A series of Modern Surveys in Mathematics*, 1990, Springer, New York.

[17] Y. Meyer, *Ondelettes*, Paris, 1990, Hermann.

[18] A. S. Nemirovskii, Nonparametric estimation of smooth regression functions, *J. Comput. Syst. Sci.*, Vol. **23**, 1986, pp. 1-11.

[19] H. P. Rosenthal,On the subspace of L_p, $p > 2$, spanned by sequences of independent random variables, *Israel J. Math.*, Vol. **8**, 1970, pp. 273-303.

[20] B. W. Silverman, Density Estimation for Statistics and Data Analysis, Chapman and Hall, London, 1986.

[21] M. Talagrand, Sharper bounds for gaussian and empirical processes, *Ann. Probab.*, Vol. **22**, 1994, pp. 28-76.

[22] M. Talagrand, New concentration inequalities in product spaces, *Technical Report*, Universite Paris VI, Ohio State University, 1995.

[23] G. Viennet, Inequalities for absolutely regular sequences: application to density estimation, *Probab. Th. Rel. Fields*, Vol. **107**, 1997, pp 467-492.

(Manuscript received October 22, 1996;
Revised version June 24, 1997;
final version October 30, 1997.)