Density estimation for β-dependent sequences

Jérôme Dedecker, Florence Merlevède

Density estimation for β-dependent sequences

Jérôme Dedecker[a] and Florence Merlevède[b]

[a] Université Paris Descartes, Sorbonne Paris Cité, Laboratoire MAP5 (UMR 8145).
Email: jerome.dedecker@parisdescartes.fr

[b] Université Paris-Est, LAMA (UMR 8050), UPEM, CNRS, UPEC.
Email: florence.merlevede@u-pem.fr

Abstract

We study the $\mathbb{L}^p$-integrated risk of some classical estimators of the density, when the observations are drawn from a strictly stationary sequence. The results apply to a large class of sequences, which can be non-mixing in the sense of Rosenblatt and long-range dependent. The main probabilistic tool is a new Rosenthal-type inequality for partial sums of BV functions of the variables. As an application, we give the rates of convergence of regular Histograms, when estimating the invariant density of a class of expanding maps of the unit interval with a neutral fixed point at zero. These Histograms are plotted in the section devoted to the simulations.

Key words: density estimation, stationary processes, long-range dependence, expanding maps.
Mathematics Subject Classification (2010): Primary 62G07; Secondary 60G10.

1 Introduction

In this paper, we have four goals:

1. We wish to extend some of the results of Viennet [16] for stationary β-mixing sequences to the much larger class of β-dependent sequences, as introduced in [7], [8]. Viennet proved that, if the β-mixing coefficients $\beta(n)$ of a stationary sequence $(Y_i)_{i\in\mathbb{Z}}$ are such that

$$\sum_{k=0}^{\infty} k^{p-2}\beta(k) < \infty\,, \quad \text{for some } p \geq 2, \tag{1.1}$$

then the $\mathbb{L}^p$-integrated risk of the usual estimators of the density of Y_i behaves as in the independent and identically distributed (iid) case (as described in the paper by Bretagnolle and Huber [4]).

For Kernel estimators, we shall obtain a complete extension of Viennet's result (assuming only as an extra hypothesis that the Kernel has bounded variation). For projection estimators, the situation is more delicate, because our dependency coefficients cannot always give a good upper bound for the variance of the estimator (this was already pointed out in [7]). However, for estimators based on piecewise polynomials (including Histograms), the result of Viennet can again be fully extended.

2. We shall consider the $\mathbb{L}^p$-integrated risk for any $p \in [1, \infty)$, and not only for $p \geq 2$ (which was the range considered in [16]). Two main reasons for this: first the case $p = 1$ is of particular interest, because it gives some information on the total variation between the (possibly signed) measure with density f_n (the estimated density) and the distribution of Y_i. The variation distance is a true distance between measures, contrary to the $\mathbb{L}^p$-distance between densities, which depends on the dominating measure. Secondly, we have in mind applications to some classes of dynamical systems (see point 4 below), for which it is known either that the density has bounded variation over $[0, 1]$ or that it is non-decreasing on $(0, 1]$ (and blows up as $x \to 0$). In such cases, it turns out that the bias of our estimators is well controlled in $\mathbb{L}^1([0, 1], dx)$.

3. We want to know what happens if (1.1) is not satisfied, or if $\sum_{k \geq 0} \beta(k) = \infty$ in the case where $p \in [1, 2]$. Such results are not given in the paper [16], although Viennet could have done it by refining some computations. For β-dependent sequences and $p = 2$, the situation is clear (see [7]): the rate of convergence of the estimator depends on the regularity of f and of the behavior of $\sum_{k=0}^{n} \beta_{1,Y}(k)$ (the coefficients $\beta_{1,Y}(k)$ will be defined in the next section, and are weaker than the corresponding β-mixing coefficients). Hence, in that case, the consistency holds as soon as $\beta_{1,Y}(n)$ tends to zero as n tends to infinity, and one can compute the rates of convergence as soon as one knows the asymptotic behavior of $\beta_{1,Y}(n)$. This is the kind of result we want to extend to any $p \in [1, \infty)$. Once again, we have precise motivations for this, coming from dynamical systems that can exhibit long-range dependence (see point 4 below).

4. As already mentioned, our first motivation was to study the robustness of the usual estimators of the density, showing that they apply to a larger class of dependent processes than in [16]. But our second main objective was to be able to visualize the invariant density of the iterates of expanding maps of the unit interval. For uniformly expanding maps the invariant density has bounded variation, and one can estimate it in $\mathbb{L}^1([0, 1], dx)$ at the usual rate $n^{-1/3}$ by using an appropriate Histogram (see Subsection 5.1). The case of the intermittent map T_γ (as defined in (5.6) for $\gamma \in (0, 1)$) is even more interesting. In that case, one knows that the invariant density h_γ is equivalent to the density $x \to (1 - \gamma)x^{-\gamma}$ on $(0, 1)$ (see the inequality (6.1)). Such a map exhibits long-range dependence as soon as $\gamma \in (1/2, 1)$, but we can use our upper bound for the random part + bias of a regular Histogram, to compute the appropriate number of breaks of the Histogram (more precisely we give the order of the number of breaks as a function of n, up to an unknown constant). In Figure 5 of Section 6, we plot the Histograms of the invariant density h_γ, when $\gamma = 1/4$ (short-range dependent case), $\gamma = 1/2$ (the boundary case), and $\gamma = 3/4$ (long-range dependent case).

One word about the main probabilistic tool. As shown in [16], to control the $\mathbb{L}^p$-integrated risks for $p > 2$, the appropriate tool is a precise Rosenthal-type inequality. Such an inequality is not easy to prove in the β-mixing case, and the β-dependent case is even harder to handle because we cannot use Berbee's coupling (see [1]) as in [16]. A major step to get a good Rosenthal bound has been made by Merlevède and Peligrad [12]: they proved a very general inequality involving only conditional expectations of the random variables with respect to the past σ-algebra, which can be applied to many situations. However, it does not fit completely to our context, and leads to small losses when applied to kernel estimators (see Section 5 in [12]). In Section 2, we shall prove a taylor-made inequality, in the spirit of that of Viennet but expressed in terms of our weaker coefficients. This inequality will give the complete extension of Viennet's results for Kernel estimators and estimators based on piecewise polynomials, when $p > 2$.

2 A Rosenthal inequality for β-dependent sequences

From now, $(Y_i)_{i\in\mathbb{Z}}$ is a strictly stationary sequence of real-valued random variables. We define the β-dependence coefficients of $(Y_i)_{i\in\mathbb{Z}}$ as in [8]:

Definition 2.1. *Let P be the law of Y_0 and $P_{(Y_i,Y_j)}$ be the law of (Y_i, Y_j). Let $\mathcal{F}_\ell = \sigma(Y_i, i \le \ell)$, let $P_{Y_k|\mathcal{F}_\ell}$ be the conditional distribution of Y_k given $\mathcal{F}_\ell$, and let $P_{(Y_i,Y_j)|\mathcal{F}_\ell}$ be the conditional distribution of (Y_i, Y_j) given $\mathcal{F}_\ell$. Define the functions*

$$f_t = \mathbf{1}_{]-\infty,t]} \quad \textit{and} \quad f_t^{(0)} = f_t - P(f_t)\,,$$

and the the random variables

$$\begin{aligned} b_\ell(k) &= \sup_{t\in\mathbb{R}} \left| P_{Y_k|\mathcal{F}_\ell}(f_t) - P(f_t) \right| , \\ b_\ell(i,j) &= \sup_{(s,t)\in\mathbb{R}^2} \left| P_{(Y_i,Y_j)|\mathcal{F}_\ell}\left(f_t^{(0)} \otimes f_s^{(0)}\right) - P_{(Y_i,Y_j)}\left(f_t^{(0)} \otimes f_s^{(0)}\right) \right| . \end{aligned}$$

Define now the coefficients

$$\beta_{1,Y}(k) = \mathbb{E}(b_0(k)) \;\; \textit{and} \;\; \beta_{2,Y}(k) = \max\left\{\beta_1(k), \sup_{i>j\ge k} \mathbb{E}((b_0(i,j)))\right\} .$$

These coefficients are weaker than the usual β-mixing coefficients of $(Y_i)_{i\in\mathbb{Z}}$. Many examples of non-mixing process for which $\beta_{2,Y}(k)$ can be computed are given in [8]. Some of these examples will be studied in Sections 5 and 6.

Let us now give the main probabilistic tool of the paper. It is a Rosenthal-type inequality for partial sums of BV functions of Y_i (as usual, BV means "of bounded variation"). We shall use it to control the random part of the $\mathbb{L}^p$ integrated risk of the estimators of the density of Y_i when $p > 2$. The proof of this inequality is given in Subsection 8.2. Is is quite delicate, and relies on two intermediate results (see Subsection 8.1).

In all the paper, we shall use the notation $a_n \ll b_n$, which means that there exists a positive constant C not depending on n such that $a_n \le C b_n$, for all positive integers n.

Proposition 2.1. *Let $(Y_i)_{i\in\mathbb{Z}}$ be a strictly stationary sequence of real-valued random variables. For any $p > 2$ and any positive integer n, there exists a non-negative $\mathcal{F}_0$-measurable random variable $A_0(n,p)$ satisfying $\mathbb{E}(A_0(n,p)) \le \sum_{k=1}^n k^{p-2}\beta_{2,Y}(k)$ and such that: for any BV function h from $\mathbb{R}$ to $\mathbb{R}$, letting $X_i = h(Y_i) - \mathbb{E}(h(Y_i))$ and $S_n = \sum_{k=1}^n X_k$, we have*

$$\mathbb{E}\left(\sup_{1\le k\le n} |S_k|^p\right) \ll n^{p/2}\left(\sum_{i=0}^{n-1} |\mathrm{Cov}(X_0, X_i)|\right)^{p/2} + n\|dh\|^{p-1}\mathbb{E}\left(|h(Y_0)|A_0(n,p)\right) \\ + n\|dh\|^{p-1}\mathbb{E}\left(|h(Y_0)|\right)\sum_{k=0}^{n}(k+1)^{p-2}\beta_{2,Y}(k)\,, \quad (2.1)$$

where $\|dh\|$ is the total variation norm of the measure dh.

Remark 2.1. *For $p = 2$, using Proposition 1 in [9], the inequality can be simplified as follows. There exists a non-negative $\mathcal{F}_0$-measurable random variable $A_0(n, 2)$ satisfying $\mathbb{E}(A_0(n, 2)) \leq \sum_{k=1}^{n} \beta_{1,Y}(k)$ and such that: for any BV function h from $\mathbb{R}$ to $\mathbb{R}$, letting $X_i = h(Y_i) - \mathbb{E}(h(Y_i))$ and $S_n = \sum_{k=1}^{n} X_k$, we have*

$$\mathbb{E}\left(\sup_{1\leq k\leq n} S_k^2\right) \ll n\|dh\|\mathbb{E}\left(|h(Y_0)|(1 + A_0(n, 2))\right) . \tag{2.2}$$

3 $\mathbb{L}^p$-integrated risk for Kernel estimators

Let $(Y_i)_{i\in\mathbb{Z}}$ be a stationary sequence with unknown marginal density f. In this section, we wish to build an estimator of f based on the variables $Y_1, \dots, Y_n$.

Let K be a bounded-variation function in $\mathbb{L}^1(\mathbb{R}, \lambda)$, where λ is the Lebesgue measure. Let $\|dK\|$ be the variation norm of the measure dK, and $\|K\|_{1,\lambda}$ be the $\mathbb{L}^1$-norm of K with respect to λ.

Define then

$$X_{k,n}(x) = K(h_n^{-1}(x - Y_k)) \quad \text{and} \quad f_n(x) = \frac{1}{nh_n}\sum_{k=1}^{n} X_{k,n}(x) ,$$

where $(h_n)_{n\geq 1}$ is a sequence of positive real numbers.

The following proposition gives an upper bound of the term

$$\mathbb{E}\left(\int_{\mathbb{R}} |f_n(x) - \mathbb{E}(f_n(x))|^p dx\right) \tag{3.1}$$

when $p > 2$ and $f \in \mathbb{L}^p(\mathbb{R}, \lambda)$.

Proposition 3.1. *Let $p > 2$, and assume that f belongs to $\mathbb{L}^p(\mathbb{R}, \lambda)$. Let*

$$V_{1,p,Y}(n) = \sum_{k=0}^{n}(k + 1)^{p-2}\beta_{1,Y}(k) \quad \textit{and} \quad V_{2,p,Y}(n) = \sum_{k=0}^{n}(k + 1)^{p-2}\beta_{2,Y}(k) . \tag{3.2}$$

The following upper bounds holds

$$\mathbb{E}\left(\int_{\mathbb{R}} |f_n(x) - \mathbb{E}(f_n(x))|^p dx\right) \ll \left(\frac{\|dK\|\|K\|_{1,\lambda}\|f\|_{p,\lambda}^{\frac{(p-2)}{(p-1)}} (V_{1,p,Y}(n))^{\frac{1}{(p-1)}}}{nh_n}\right)^{\frac{p}{2}} + \frac{1}{(nh_n)^{p-1}}\|dK\|^{p-1}\|K\|_{1,\lambda}V_{2,p,Y}(n) . \tag{3.3}$$

Remark 3.1. *Note that, if $nh_n \to \infty$ as $n \to \infty$ and $\sum_{k=0}^{\infty}(k + 1)^{p-2}\beta_{2,Y}(k) < \infty$, then it follows from Proposition 3.1 that*

$$\limsup_{n\to\infty} (nh_n)^{p/2}\mathbb{E}\left(\int_{\mathbb{R}} |f_n(x) - \mathbb{E}(f_n(x))|^p dx\right) \leq C\|dK\|^{\frac{p}{2}}\|K\|_{1,\lambda}^{\frac{p}{2}}\|f\|_{p,\lambda}^{\frac{p(p-2)}{2(p-1)}} , \tag{3.4}$$

for some positive constant C. Note that (3.4) is comparable to the upper bound obtained by Viennet [16], with two differences: firstly our condition is written in terms of the coefficients $\beta_{2,Y}(n)$ (while Viennet used the usual β-mixing coefficients), and secondly we only require that f belongs to $\mathbb{L}^p(\mathbb{R}, \lambda)$ (while Viennet assumed that f is bounded). When $p \geq 4$, an upper bound similar to (3.4) is given in [12], Proposition 33 Item (2), under the slightly stronger condition $\beta_{2,Y}(n) = O(n^{-(p-1+\varepsilon)})$ for some $\varepsilon > 0$.

Remark 3.2. *The first term in the upper bound of Proposition 3.1 has been obtained by assuming only that f belongs to $\mathbb{L}^p(\mathbb{R},\lambda)$. As will be clear from the proof, a better upper bound can be obtained by assuming that f belongs to $\mathbb{L}^q(\mathbb{R},\lambda)$ for $q > p$. For instance, if f is bounded, the first term of the upper bound can be replaced by*

$$\|f\|_\infty^{\frac{p}{2}-1}\left(\frac{\|dK\|\|K\|_{1,\lambda}}{nh_n}\right)^{\frac{p}{2}}\sum_{k=0}^{n}(k+1)^{\frac{p}{2}-1}\beta_{1,Y}(k)\,.$$

This can lead to a substantial improvement of the upper bound of (3.1)*, for instance in the cases where $\sum_{k=0}^{\infty}(k+1)^{p-2}\beta_{2,Y}(k) = \infty$ but $\sum_{k=0}^{\infty}(k+1)^{p/2-1}\beta_{1,Y}(k) < \infty$.*

We now give an upper bound of the same quantity when $1 \le p \le 2$. Note that the case $p = 1$ is of special interest, since it enables to get the rate of convergence to the unknow probability μ (with density f) for the total variation distance.

Proposition 3.2. *As in* (3.2)*, let $V_{1,2,Y}(n) = \sum_{k=0}^{n}\beta_{1,Y}(k)$. The following upper bounds hold*

1. *For $p = 2$,* $\mathbb{E}\left(\int_{\mathbb{R}}|f_n(x)-\mathbb{E}(f_n(x))|^2dx\right) \ll \frac{1}{nh_n}\|dK\|\|K\|_{1,\lambda}V_{1,2,Y}(n)\,.$
2. *Let $1 \le p < 2$ and $\alpha > 1, q > 1$. Let also $U_{1,q,Y}(n) = \sum_{k=0}^{n}(k+1)^{\frac{1}{q-1}}\beta_{1,Y}(k)$. If*

$$M_{\alpha q,p}(f) := \int |x|^{\frac{\alpha q(2-p)}{p}}f(x)dx < \infty \quad \textit{and} \quad M_{\alpha,p}(K) := \int |x|^{\frac{\alpha(2-p)}{p}}|K(x)|dx < \infty\,,$$

then

$$\mathbb{E}\left(\int_{\mathbb{R}}|f_n(x)-\mathbb{E}(f_n(x))|^p dx\right) \ll \left(\frac{\|dK\|\|K\|_{1,\lambda}(M_{\alpha q,p}(f))^{\frac{1}{q}}(U_{1,q,Y}(n))^{\frac{q-1}{q}}}{nh_n}\right)^{\frac{p}{2}} + \left(\frac{\|dK\|\left(\|K\|_{1,\lambda}+h_n^{\frac{\alpha(2-p)}{p}}M_{\alpha,p}(K)\right)V_{1,2,Y}(n)}{nh_n}\right)^{\frac{p}{2}}.$$

Remark 3.3. *Note first that Item 1 of Proposition 3.2 is due to Dedecker and Prieur [7]. For $p \in [1,2)$, It follows from Item 2 that if $\mathbb{E}(|Y_0|^{\alpha q(2-p)/p}) < \infty$ for $\alpha > 1, q > 1$ and if $\sum_{k=0}^{\infty}(k+1)^{1/(q-1)}\beta_{1,Y}(k) < \infty$, then*

$$\limsup_{n\to\infty}(nh_n)^{\frac{p}{2}}\mathbb{E}\left(\int_{\mathbb{R}}|f_n(x)-\mathbb{E}(f_n(x))|^p dx\right) \le C\|dK\|^{\frac{p}{2}}\|K\|_{1,\lambda}^{\frac{p}{2}}\left(1+(M_{\alpha q,p}(f))^{\frac{p}{2q}}\right)$$

for some positive constant C, provided $nh_n \to \infty$ and $h_n \to 0$ as $n \to \infty$. As will be clear from the proof, this upper bound remains true when $q = \infty$, that is when $\|Y_0\|_\infty < \infty$ and $\sum_{k=0}^{\infty}\beta_{1,Y}(k) < \infty$.

With these two propositions, one can get the rates of convergence of f_n to f, when f belongs to the generalized Lipschitz spaces $\mathrm{Lip}^*(s,\mathbb{L}^p(\mathbb{R},\lambda))$ with $s > 0$, as defined in [10], Chapter 2, Paragraph 9. Recall that $\mathrm{Lip}^*(s,\mathbb{L}^p(\mathbb{R},\lambda))$ is a particular case of Besov spaces (precisely $\mathrm{Lip}^*(s,\mathbb{L}^p(\mathbb{R},\lambda)) = B_{s,p,\infty}(\mathbb{R})$). Moreover, if s is a positive integer, $\mathrm{Lip}^*(s,\mathbb{L}^p(\mathbb{R},\lambda))$ contains the Sobolev space $W^s(\mathbb{L}^p(\mathbb{R},\lambda))$ if $p > 1$ and $W^{s-1}(BV)$ if

$p = 1$ (see again [10], Chapter 2, paragraph 9). Recall that, if s is a positive integer the space $W^s(\mathbb{L}^p(\mathbb{R},\lambda))$ (resp. $W^s(BV)$) is the space of functions for which $f^{(s-1)}$ is absolutely continuous, with almost everywhere derivative $f^{(s)}$ belonging to $\mathbb{L}^p(\mathbb{R},\lambda)$ (resp. $f^{(s)}$ has bounded variation).

Let $K_h(\cdot) = h^{-1}K(\cdot/h)$, and r be a positive integer, and assume that, for any g in $W^r(\mathbb{L}^p(\mathbb{R},\lambda))$,

$$\int_{\mathbb{R}} |g(x) - g * K_h(x)|^p dx \leq C_1 h^{pr} \|g^{(r)}\|_{p,\lambda}^p \,, \tag{3.5}$$

for some constant C_1 depending only on r. For instance, (3.5) is satisfied for any Parzen kernel of order r (see Section 4 in [4]).

From (3.5) and Theorem 5.2 in [10], we infer that, for any $g \in \mathbb{L}^p(\mathbb{R},\lambda)$,

$$\int_{\mathbb{R}} |g(x) - g * K_h(x)|^p dx \leq C_2 (\omega_r(g,h)_p)^p \,,$$

for some constant C_2 depending only on r, where $\omega_r(g,\cdot)_p$ is the r-th modulus of regularity of g in $\mathbb{L}^p(\mathbb{R},\lambda)$ as defined in [10], Chapter 2, Paragraph 7. This last inequality implies that, if g belongs to $\mathrm{Lip}^*(s,\mathbb{L}^p(\mathbb{R},\lambda))$ for any $s \in [r-1,r)$, then

$$\int_{\mathbb{R}} |g(x) - g * K_h(x)|^p dx \leq C_2 h^{ps} \|g\|_{\mathrm{Lip}^*(s,\mathbb{L}^p(\mathbb{R},\lambda))}^p \,. \tag{3.6}$$

Combining Proposition 3.1 or 3.2 with the control of the bias given in (3.6), we obtain the following upper bounds for the $\mathbb{L}^p$-integrated risk of the kernel estimator.

Let K be a bounded variation function in $\mathbb{L}^1(\mathbb{R},\lambda)$, and assume that K satisfies (3.5) for some positive integer r.

- Let $p \geq 2$ and assume that $\sum_{k=0}^{n}(k+1)^{p-2}\beta_{2,Y}(k) = O(n^{\delta(p-1)})$ for some $\delta \in [0,1)$. Assume that f belongs to $\mathrm{Lip}^*(s,\mathbb{L}^p(\mathbb{R},\lambda))$ for $s \in [r-1,r)$ or to $W^s(\mathbb{L}^p(\mathbb{R},\lambda))$ for $s = r$. Then, taking $h_n = Cn^{-(1-\delta)/(2s+1)}$,
$$\mathbb{E}\left(\int_{\mathbb{R}} |f_n(x) - f(x)|^p dx\right) = O\left(n^{\frac{-ps(1-\delta)}{2s+1}}\right) \,.$$
Hence, if $\sum_{k=0}^{\infty}(k+1)^{p-2}\beta_{2,Y}(k) < \infty$ (case $\delta = 0$), we obtain the same rate as in the iid situation. This result generalizes the result of Viennet [16], who obtained the same rates under the condition $\sum_{k=0}^{\infty}(k+1)^{p-2}\beta(k) < \infty$, where the $\beta(k)$'s are the usual β-mixing coefficients.
- Let $1 \leq p < 2$. Assume that Y_0 has a moment of order $q(p-2)/p + \varepsilon$ for some $q > 1$ and $\varepsilon > 0$, and that $\sum_{k=0}^{n}(k+1)^{1/(q-1)}\beta_{1,Y}(k) = O(n^{\delta q/(q-1)})$ for some $\delta \in [0,1)$. Assume that f belongs to $\mathrm{Lip}^*(s,\mathbb{L}^p(\mathbb{R},\lambda))$ for $s \in [r-1,r)$ or to $W^s(\mathbb{L}^p(\mathbb{R},\lambda))$ for $s = r$. Then, taking $h_n = Cn^{-(1-\delta)/(2s+1)}$,
$$\mathbb{E}\left(\int_{\mathbb{R}} |f_n(x) - f(x)|^p dx\right) = O\left(n^{\frac{-ps(1-\delta)}{2s+1}}\right) \,.$$
Hence, if $\sum_{k=0}^{\infty}(k+1)^{1/(q-1)}\beta_{1,Y}(k) < \infty$ (case $\delta = 0$), we obtain the same rate as in the iid situation.

  Let us consider the particular case where $p = 1$. Let μ be the probability measure with density f, and let $\hat{\mu}_n$ be the random (signed) measure with density f_n. We have just proved that
$$\mathbb{E}\|\hat{\mu}_n - \mu\| = O\left(n^{\frac{-s(1-\delta)}{2s+1}}\right)$$

(recall that $\|\cdot\|$ is the variation norm). It is an easy exercice to modify $\hat{\mu}_n$ in order to get a random probability measure μ_n^* that converges to μ at the same rate (take the positive part of f_n and renormalize).

4 $\mathbb{L}^p$-integrated risk for estimators based on piecewise polynomials

Let $(Y_i)_{i\in\mathbb{Z}}$ be a stationary sequence with unknown marginal density f. In this section, we wish to estimate f on a compact interval I with the help of the variables $Y_1, \ldots, Y_n$. Without loss of generality, we shall assume here that $I = [0,1]$.

We shall consider the piecewise polynomial basis on a a regular partition of $[0,1]$, defined as follows. Let $(Q_i)_{1\le i\le r+1}$ be an orthonormal basis of the space of polynomials of order r on $[0,1]$, and define the function R_i on $\mathbb{R}$ by: $R_i(x) = Q_i(x)$ if x belongs to $]0,1]$ and 0 otherwise. Consider now the regular partition of $]0,1]$ into m_n intervals $(](j-1)/m_n, j/m_n])_{1\le j\le m_n}$. Define the functions $\varphi_{i,j}(x) = \sqrt{m_n}R_i(m_n x-(j-1))$. Clearly the family $(\varphi_{i,j})_{1\le i\le r+1}$ is an orthonormal basis of the space of polynomials of order r on the interval $[(j-1)/m_n, j/m_n]$. Since the supports of $\varphi_{i,j}$ and $\varphi_{k,\ell}$ are disjoints for $\ell \neq j$, the family $(\varphi_{i,j})_{1\le i\le r+1, 1\le j\le m}$ is then an orthonormal system of $\mathbb{L}^2([0,1],\lambda)$. The case of regular Histograms corresponds to $r = 0$.

Define then

$$X_{i,j,n} = \frac{1}{n}\sum_{k=1}^{n}\varphi_{i,j}(Y_k) \quad \text{and} \quad f_n = \sum_{i=1}^{r+1}\sum_{j=1}^{m_n} X_{i,j,n}\varphi_{i,j}\,.$$

The following proposition gives an upper bound of

$$\mathbb{E}\left(\int_0^1 |f_n(x) - \mathbb{E}(f_n(x))|^p dx\right)$$

when $p > 2$ and $f\mathbf{1}_{[0,1]} \in \mathbb{L}^p([0,1],\lambda)$.

Proposition 4.1. *Let $p > 2$, and assume that $f\mathbf{1}_{[0,1]}$ belongs to $\mathbb{L}^p([0,1],\lambda)$. Let*

$$C_{1,p} = \sum_{i=1}^{r+1} \|R_i\|_\infty^{\frac{3p}{2}} \|dR_i\|^{\frac{p}{2}} \quad \text{and} \quad C_{2,p} = \sum_{i=1}^{r+1} \|R_i\|_\infty^{p+1}\|dR_i\|^{p-1}\,.$$

and recall that $V_{1,p,Y}(n)$ and $V_{2,p,Y}(n)$ have been defined in (3.2). The following upper bounds holds

$$\mathbb{E}\left(\int_0^1 |f_n(x) - \mathbb{E}(f_n(x))|^p dx\right) \ll \left(\frac{m_n}{n}\right)^{\frac{p}{2}} C_{1,p}\left(\|f\mathbf{1}_{[0,1]}\|_{p,\lambda}^{\frac{(p-2)}{(p-1)}} \left(V_{1,p,Y}(n)\right)^{\frac{1}{(p-1)}}\right)^{\frac{p}{2}} + \left(\frac{m_n}{n}\right)^{p-1} C_{2,p}V_{2,p,Y}(n)\,.$$

Remark 4.1. *Note that, if $n/m_n \to \infty$ as $n \to \infty$ and $\sum_{k=0}^{\infty}(k+1)^{p-2}\beta_{2,Y}(k) < \infty$, then it follows from Proposition 4.1 that*

$$\limsup_{n\to\infty} \left(\frac{n}{m_n}\right)^{\frac{p}{2}} \mathbb{E}\left(\int_0^1 |f_n(x) - \mathbb{E}(f_n(x))|^p dx\right) \le C\cdot C_{1,p}\|f\mathbf{1}_{[0,1]}\|_{p,\lambda}^{\frac{p(p-2)}{2(p-1)}}\,, \qquad (4.1)$$

for some positive constant C. This bound is comparable to the upper bound obtained by Viennet ([16], Theorem 3.2) for the usual β-mixing coefficients. Note however that Viennet's results is valid for a much broader class of projection estimators. As a comparison, it seems very difficult to deal with the trigonometric basis in our setting.

We now give an upper bound of the same quantity when $1 \leq p \leq 2$.

Proposition 4.2. *Let $1 \leq p \leq 2$. The following upper bounds holds*

$$\mathbb{E}\left(\int_0^1 |f_n(x) - \mathbb{E}(f_n(x))|^p dx\right) \ll \left(\frac{m_n}{n}\right)^{\frac{p}{2}} C_{1,2}^{\frac{p}{2}} \left(V_{1,2,Y}(n)\right)^{\frac{p}{2}} .$$

Remark 4.2. *Note first that the upper bounf for $p = 2$ is due to Dedecker and Prieur [7]. Note also that, if $n/m_n \to \infty$ as $n \to \infty$ and $\sum_{k=0}^{\infty} \beta_{1,Y}(k) < \infty$, then it follows from Proposition 4.2 that*

$$\limsup_{n\to\infty} \left(\frac{n}{m_n}\right)^{\frac{p}{2}} \mathbb{E}\left(\int_0^1 |f_n(x) - \mathbb{E}(f_n(x))|^p dx\right) \leq C \cdot C_{1,2}^{\frac{p}{2}} , \tag{4.2}$$

for some positive constant C.

With these two propositions, one can get the rates of convergence of f_n to $f\mathbf{1}_{[0,1]}$ when $f\mathbf{1}_{[0,1]}$ belongs to to the generalized Lipschitz spaces $\mathrm{Lip}^*(s, \mathbb{L}^p([0,1], \lambda))$ with $s > 0$.

Applying the Bramble-Hilbert lemma (see [3]), we know that, for any f such that $f\mathbf{1}_{[0,1]}$ belongs to $W^{r+1}(\mathbb{L}^p([0,1], \lambda))$

$$\int_0^1 |f(x) - \mathbb{E}(f_n(x))|^p dx \leq C_1 m_n^{-p(r+1)} \left\| f^{(r)} \mathbf{1}_{[0,1]} \right\|_{p,\lambda} ,$$

for some constant C_1 depending only on r. From [10], page 359, we know that, if $f\mathbf{1}_{[0,1]}$ belongs to $\mathrm{Lip}^*(s, \mathbb{L}^p([0,1], \lambda))$ and if the degree r is such that $r > s - 1$,

$$\int_0^1 |f(x) - \mathbb{E}(f_n(x))|^p dx \leq C_2 m_n^{-ps} \| f\mathbf{1}_{[0,1]} \|_{\mathrm{Lip}^*(s, \mathbb{L}^p([0,1], \lambda))} , \tag{4.3}$$

for some constant C_2 depending only on r. Combining Proposition 4.1 or 4.2 with the control of the bias given in (4.3), we obtain the following upper bounds for the $\mathbb{L}^p$-integrated risk.

- Let $p > 2$ and assume that $\sum_{k=0}^{n} (k+1)^{p-2} \beta_{2,Y}(k) = O(n^{\delta(p-1)})$ for some $\delta \in [0,1)$. Assume that $f\mathbf{1}_{[0,1]}$ belongs to $\mathrm{Lip}^*(s, \mathbb{L}^p([0,1], \lambda))$ for $s < r+1$ or to $W^s(\mathbb{L}^p([0,1]))$ for $s = r+1$. Then, taking $m_n = [Cn^{(1-\delta)/(2s+1)}]$,

  $$\mathbb{E}\left(\int_0^1 |f_n(x) - f(x)|^p dx\right) = O\left(n^{\frac{-ps(1-\delta)}{2s+1}}\right) .$$

  Hence, if $\sum_{k=0}^{\infty} (k+1)^{p-2} \beta_{2,Y}(k) < \infty$ (case $\delta = 0$), we obtain the same rate as in the iid situation. This result generalizes the result of Viennet [16], who obtained the same rates under the condition $\sum_{k=0}^{\infty} (k+1)^{p-2} \beta(k) < \infty$, where the $\beta(k)$'s are the usual β-mixing coefficients.

- Let $1 \leq p \leq 2$ and assume that $\sum_{k=0}^{n} \beta_{1,Y}(k) = O(n^{\delta})$ for some $\delta \in [0, 1)$. Assume that $f\mathbf{1}_{[0,1]}$ belongs to $\mathrm{Lip}^*(s, \mathbb{L}^p([0,1], \lambda))$ for $s < r+1$ or to $W^s(\mathbb{L}^p([0,1], \lambda))$ for $s = r+1$. Then, taking $m_n = [Cn^{(1-\delta)/(2s+1)}]$,

$$\mathbb{E}\left(\int_0^1 |f_n(x) - f(x)|^p dx\right) = O\left(n^{\frac{-ps(1-\delta)}{2s+1}}\right).$$

  Hence, if $\sum_{k=0}^{\infty} \beta_{1,Y}(k) < \infty$ (case $\delta = 0$), we obtain the same rate as in the iid situation.

  Let us consider the particular case where $p = 1$ and f is supported on $[0, 1]$. Let μ be the probability measure with density f, and let $\hat{\mu}_n$ be the random measure with density f_n. We have just proved that

$$\mathbb{E}\|\hat{\mu}_n - \mu\| = O\left(n^{\frac{-s(1-\delta)}{2s+1}}\right)$$

  (recall that $\|\cdot\|$ is the variation norm).

5 Application to density estimation of expanding maps

5.1 Uniformly expanding maps

Several classes of uniformly expanding maps of the interval are considered in the literature. We recall here the definition given in [6] (see the references therein for more informations).

Definition 5.1. *A map $T : [0, 1] \to [0, 1]$ is uniformly expanding, mixing and with density bounded from below if it satisfies the following properties:*

1. *There is a (finite or countable) partition of T into subintervals I_n on which T is strictly monotonic, with a C^2 extension to its closure $\overline{I_n}$, satisfying Adler's condition $|T''|/|T'|^2 \leq C$, and with $|T'| \geq \lambda$ (where $C > 0$ and $\lambda > 1$ do not depend on I_n).*
2. *The length of $T(I_n)$ is bounded from below.*
3. *In this case, T has finitely many absolutely continuous invariant measures, and each of them is mixing up to a finite cycle. We assume that T has a single absolutely continuous invariant probability measure ν, and that it is mixing.*
4. *Finally, we require that the density h of ν is bounded from below on its support.*

From this point on, we will simply refer to such maps as *uniformly expanding*. It is well known, that, for such classes, the density h has bounded variation.

We wish to estimates h with the help of the first iterates $T, T^2, \ldots, T^n$. Since the bias term of a density having bounded variation is well controlled in $\mathbb{L}^1([0,1], \lambda)$, we shall give the rates in terms of the $\mathbb{L}^1$-integrated risk. We shall use an Histogram, as defined in Section 4. More precisely, our estimator h_n of h is given by

$$h_n(x, y) = \sum_{i=1}^{m_n} \alpha_{i,n}(y)\varphi_i(x), \tag{5.1}$$

where

$$\varphi_i = \sqrt{m_n}\mathbf{1}_{](i-1)/m_n, i/m_n]} \quad \text{and} \quad \alpha_{i,n}(y) = \frac{1}{n}\sum_{k=1}^{n} \varphi_i\left(T^k(y)\right).$$

As usual, the bias term is of order

$$\int_0^1 |h(x) - \nu(h_n(x,\cdot))|\, dx = O\left(\frac{1}{m_n}\right) . \tag{5.2}$$

On another hand, one can apply Proposition 4.2 to get

$$\int_0^1 \int_0^1 |h_n(x,y) - \nu(h_n(x,\cdot))| \nu(dy)\, dx = O\left(\sqrt{\frac{m_n}{n}}\right) . \tag{5.3}$$

Choosing $m_n = [Cn^{1/3}]$ for some $C > 0$, it follows from (5.2) and (5.3) that

$$\int_0^1 \int_0^1 |h_n(x,y) - h(x)| \nu(dy)\, dx = O\left(n^{-1/3}\right) .$$

Now, if $\nu_n(y)$ is the probability measure with density $h_n(\cdot, y)$, we have just proved that

$$\int_0^1 \|\nu_n(y) - \nu\|\, \nu(dy) = O\left(n^{-1/3}\right) .$$

Let us briefly explain how to derive (5.3) from Proposition 4.2. To do this, we go back to the Markov chain associated with T, as we describe now. Let first K be the Perron-Frobenius operator of T with respect to ν, defined as follows: for any functions u, v in $\mathbb{L}^2([0,1],\nu)$

$$\nu(u \cdot v \circ T) = \nu(K(u) \cdot v) . \tag{5.4}$$

The relation (5.4) states that K is the adjoint operator of the isometry $U : u \mapsto u \circ T$ acting on $\mathbb{L}^2([0,1],\nu)$. It is easy to see that the operator K is a transition kernel, and that ν is invariant by K. Let now $(Y_i)_{i\geq 0}$ be a stationary Markov chain with invariant measure ν and transition kernel K. It is well known that on the probability space $([0,1],\nu)$, the random vector $(T, T^2, \ldots, T^n)$ is distributed as $(Y_n, Y_{n-1}, \ldots, Y_1)$. Hence (5.3) is equivalent to

$$\mathbb{E}\left(\int_0^1 \left|\tilde{h}_n(x) - \mathbb{E}\left(\tilde{h}_n(x)\right)\right| dx\right) = O\left(\sqrt{\frac{m_n}{n}}\right) \tag{5.5}$$

where

$$\tilde{h}_n(x) = \sum_{i=1}^{m_n} X_{i,n}\varphi_i(x) , \quad \text{with} \quad X_{i,n} = \frac{1}{n}\sum_{k=1}^{n} \varphi_i(Y_k) .$$

Now (5.5) follows easily from Proposition 4.2 ans the fact the $\beta_{1,Y}(n) = O(a^n)$ for some $a \in (0,1)$ (see Section 6.3 in [8]).

5.2 Intermittent maps

For γ in $(0,1)$, we consider the intermittent map T_γ (or simply T) from $[0,1]$ to $[0,1]$, introduced by Liverani, Saussol and Vaienti [11]:

$$T_\gamma(x) = \begin{cases} x(1+2^\gamma x^\gamma) & \text{if } x \in [0,1/2[\\ 2x-1 & \text{if } x \in [1/2,1]. \end{cases} \tag{5.6}$$

It follows from [15] that there exists a unique absolutely continuous T_γ-invariant probability measure ν_γ (or simply ν), with density h_γ (or simply h). From [15], Theorem 1, we infer that the function $x \mapsto x^\gamma h_\gamma(x)$ is bounded from above and below. From Lemma 2.3 in [11], we know that h is non-increasing with $h_\gamma(1) > 0$, and that it is Lipshitz on any interval $[a, 1]$ with $a > 0$.

We wish to estimate h with the help of the first iterates $T, T^2, \dots, T^n$. To do this, we shall use the Histogram h_n defined in (5.1).

Using the properties of h, it is easy to see that

$$\int_0^1 |h(x) - \nu(h_n(x, \cdot))| \, dx = O\left(\frac{1}{m_n^{1-\gamma}}\right) . \tag{5.7}$$

On another hand, one can apply Proposition 4.2 to get

$$\int_0^1 \int_0^1 |h_n(x, y) - \nu(h_n(x, \cdot))| \nu(dy) \, dx = \begin{cases} O\left(\sqrt{m_n/n}\right) & \text{if } \gamma < 1/2 \\ O\left(\sqrt{m_n \log(n)/n}\right) & \text{if } \gamma = 1/2 \\ O\left(\sqrt{m_n/n^{(1-\gamma)/\gamma}}\right) & \text{if } \gamma > 1/2 \,. \end{cases} \tag{5.8}$$

Starting from (5.7) and (5.8), the appropriate choices of m_n lead to the rates

$$\int_0^1 \int_0^1 |h_n(x, y) - h(x)| \nu(dy) \, dx = \begin{cases} O\left(n^{-(1-\gamma)/(3-2\gamma)}\right) & \text{if } \gamma < 1/2 \\ O\left((n/\log(n))^{-1/4}\right) & \text{if } \gamma = 1/2 \\ O\left(n^{-(1-\gamma)^2/\gamma(3-2\gamma)}\right) & \text{if } \gamma > 1/2 \,. \end{cases} \tag{5.9}$$

Keeping the same notations as in Section 5.1, the bound (5.8) follows from Proposition 4.2 by noting that the coefficients $\beta_{1,Y}(n)$ of the chain $(Y_i)_{i \geq 0}$ associated with T satisfy $\beta_{1,Y}(n) = O(n^{-(1-\gamma)/\gamma})$ (see [5]).

6 Simulations

6.1 Functions of an AR(1) process

In this subsection, we first simulate the simple AR(1) process

$$X_{n+1} = \frac{1}{2}\left(X_n + \varepsilon_{n+1}\right) ,$$

where X_0 is uniformly distributed over $[0, 1]$, and $(\varepsilon_i)_{i \geq 1}$ is a sequence of iid random variables with distribution $\mathcal{B}(1/2)$, independent of X_0.

One can check that the transition Kernel of this chain is

$$K(f)(x) = \frac{1}{2}\left(f(x) + f(x+1)\right) ,$$

and that the uniform distribution on $[0, 1]$ is the unique invariant distribution by K. Hence, the chain $(X_i)_{i \geq 0}$ is strictly stationary.

It is well known that this chain is not α-mixing in the sense of Rosenblatt [14] (see for instance [2]). In fact, the kernel K is the Perron-Frobenius operator of the uniformly expanding map T_0 defined in (5.6), which is another way to see that this non-irreducible chain cannot be mixing in the sense of Rosenblatt.

However, one can prove that the coefficients $\beta_{2,X}$ of the chain $(X_i)_{i\geq 0}$ are such that

$$\beta_{2,X}(k) \leq 2^{-k}$$

(see for instance Section 6.1 in [8]).

Let now Q_{μ,σ^2} be the inverse of the cumulative distribution function of the law $\mathcal{N}(\mu,\sigma^2)$. Let then

$$Y_i = Q_{\mu,\sigma^2}(X_i)\,.$$

The sequence $(Y_i)_{i\geq 0}$ is also a stationary Markov chain (as an invertible function of a stationary Markov chain), and one can easily check that $\beta_{2,Y}(k) = \beta_{2,X}(k)$. By construction, Y_i is $\mathcal{N}(\mu,\sigma^2)$-distributed, but the sequence $(Y_i)_{i\geq 0}$ is not a Gaussian process (otherwise it would be mixing in the sense of Rosenblatt).

Figure 1 shows two graphs of the kernel estimator of the density of Y_i, for $\mu = 10$ and $\sigma^2 = 2$, based on the simulated sample $Y_1,\ldots,Y_n$. The kernel K is the Epanechnikov kernel (which is a Parzen kernel of order 2, thus providing theoretically a good estimation when the density belongs to the the Sobolev space of order 2). Here we do not interfer, and let the software R choose an appropriate bandwidth, to see that the default procedure delivers a correct estimation of the density, even in this non-mixing framework.

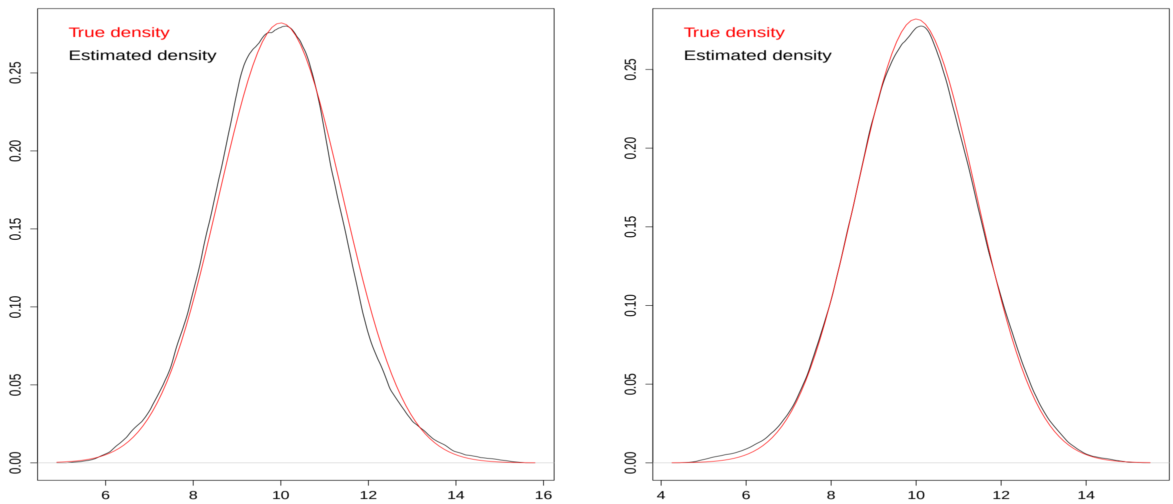

Figure 1: Estimation of the density of Y_i ($\mu = 10$ and $\sigma^2 = 2$) *via* the Epanechnikov kernel: $n = 1000$ (left) and $n = 5000$ (right)

We continue with another example. Let $Q : [0,1] \mapsto [0,1]$ be the inverse of the cumulative distribution function of the density f over $[0,1]$ defined by: $f \equiv 1/2$ on $[0,1/4]\cup[3/4,1]$ and $f \equiv 3/2$ on $(1/4,3/4)$. Let then

$$Y_i = Q(X_i)\,.$$

The same reasoning as before shows that $(Y_i)_{i\geq 0}$ is a stationary Markov chain satisfying $\beta_{2,Y}(k) = \beta_{2,X}(k)$. By construction, the density of the distribution of the X_i's is the density f. Since f belongs to the class of bounded variation functions over $[0,1]$, the bias of the Histogram will be well controlled in $\mathbb{L}^1([0,1],\lambda)$, and the computations of Section 4 show that a reasonable choice for m_n is $m_n = [n^{1/3}]$.

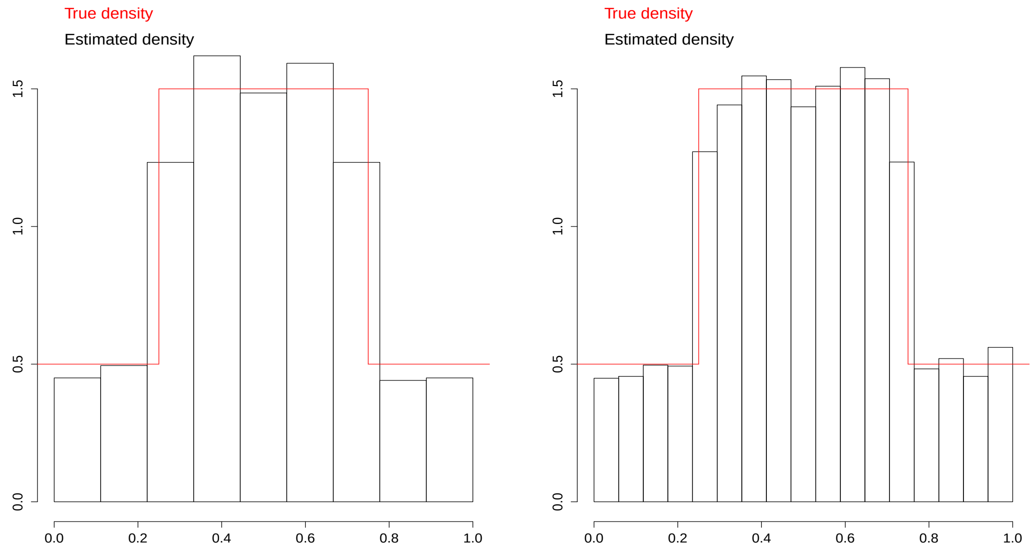

Figure 2: Estimation of the density f by an Histogram: $n = 1000$ (left) and $n = 5000$ (right)

Figure 2 shows two Histograms based on the simulated sample $Y_1, \ldots, Y_n$, with $m_n = [n^{1/3}]$ and two different values of n. We shall now study this example from a numerical point of view, by giving an estimation of the $\mathbb{L}^1$-integrated risk of the Histogram. To see the asymptotic behavior, we let n run from 5000 to 110000, with an increment of size 5000. The $\mathbb{L}^1$-integrated risk is estimated via a classical Monte-Carlo procedure, by averaging the variation distance between the true density and the estimated density over $N = 300$ independent trials. The results are given in the table below:

n	$\mathbb{L}^1$-integ. risk	n	$\mathbb{L}^1$-integ. risk
5000	0.0477	60000	0.0227
10000	0.0381	65000	0.0177
15000	0.0265	70000	0.0217
20000	0.0316	75000	0.0231
25000	0.0293	80000	0.0209
30000	0.0292	85000	0.0202
35000	0.0207	90000	0.0156
40000	0.0277	95000	0.0197
45000	0.0245	100000	0.0209
50000	0.0191	105000	0.0193
55000	0.0251	110000	0.0189

Figure 3 (left) shows the value of the $\mathbb{L}^1$-integrated risk as n increases. One can see that the $\mathbb{L}^1$-integrated risk is smaller for some particular sizes of n. This is due to the fact that for such n, two of the breaks of the Histogram are located precisely at 1/4 and 3/4, that is at the two discontinuity points of the density. In that case the variation distance between the Histogram and the true density is particularly small, as illustrated by Figure 3 (right). But of course, we are not supposed to know where the discontinuity are located.

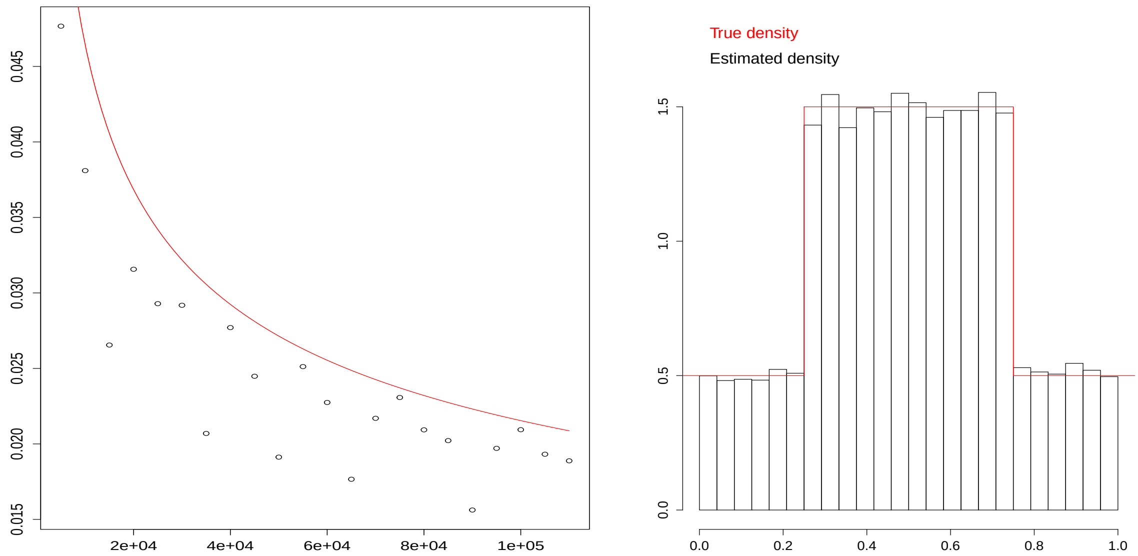

Figure 3: Left: Graph of the $\mathbb{L}^1$-integrated risk as a function of n. The red curve is the theoretical rate for estimating BV functions : $n \to n^{-1/3}$. Right: Estimation of the density f by an Histogram: $n = 15000$

6.2 Intermittent maps

In this section, our goal is to visualize the invariant density h_γ of the intermittent maps T_γ defined in (5.6). Hence, we shall consider very large n in order to get a good picture. As indicated in Subsection 5.2, we choose $m_n = [n^{1/(3-2\gamma)}]$ if $\gamma \in (0, 1/2]$, and $m_n = [n^{(1-\gamma)/(\gamma(3-2\gamma))}]$ if $\gamma \in (1/2, 1)$. We shall consider three cases: $\gamma = 1/4$, $\gamma = 1/2$ and $\gamma = 3/4$. Recall that $\gamma < 1/2$ corresponds to the short-range dependent case, $\gamma > 1/2$ to the long-range dependent case, and $\gamma = 1/2$ is the boundary case (see for instance [5]).

As one can see from Figure 4, due to the behavior of T_γ around zero, the process $(T_\gamma^i)_{i\geq 0}$ spends much more time in the neighborhood of 0 when $\gamma = 3/4$ than when $\gamma = 1/4$.

We do not have an explicit expression of the invariant density h_γ, but, as already mentioned in Subsection 5.2, we know the qualitative behavior of h_γ in the neighborhood of 0. More precisely, one can introduce an *equivalent* density f_γ such that $f_\gamma(x) = (1-\gamma)x^{-\gamma}$ if $x \in (0, 1]$ and $f_\gamma(x) = 0$ elsewhere. By equivalent, we mean that there exists two positive constants a, b, such that, on $(0, 1]$,

$$0 < a \leq h_\gamma / f_\gamma \leq b < \infty \,. \tag{6.1}$$

Figure 5 shows three Histograms based on $(T_i^\gamma)_{1\leq i\leq n}$ for $\gamma = 1/4$, $\gamma = 1/2$, $\gamma = 3/4$, and very large values of n. Since the rates are very slow if γ is much larger than 1/2 (see (5.9)), we have chosen $n = 10^7$ for the estimation of $h_{3/4}$. In each cases, we plotted the equivalent density on the same graph, to see that the behavior of h_γ around 0 is as expected.

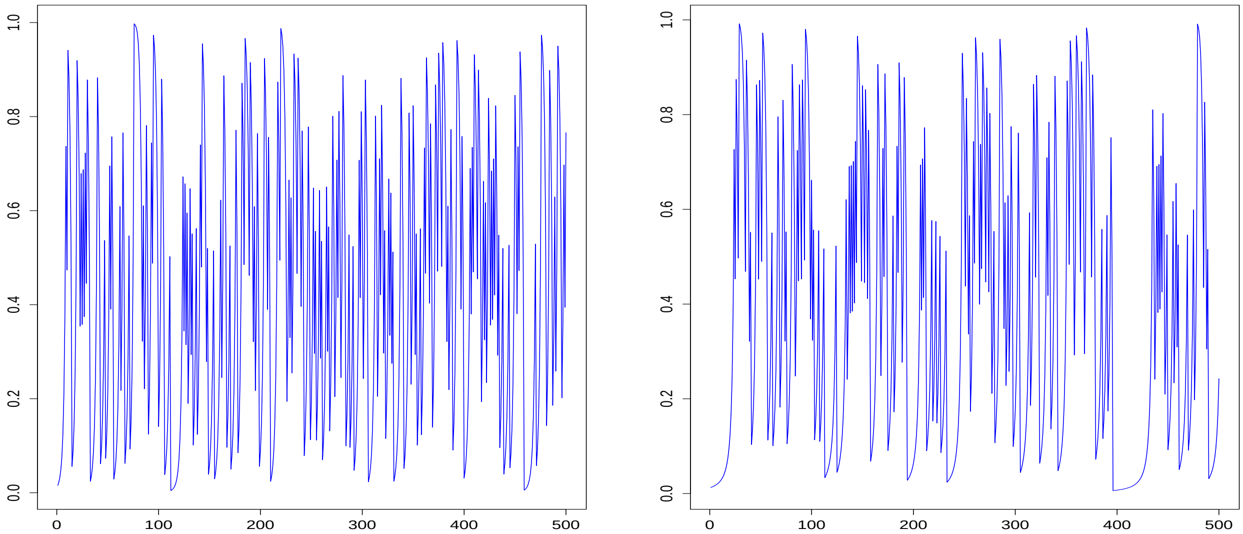

Figure 4: Graphs of 500 iterations of the map T_γ for $\gamma = 1/4$ (left) and $\gamma = 3/4$ (right)

7 Proof of the results of Sections 3 and 4

7.1 Proof of Proposition 3.1

Setting $Y_{i,n}(x) = K((x - Y_i)/h_n)$ and $X_{i,n}(x) = Y_{i,n}(x) - \mathbb{E}(Y_{i,n}(x))$, we have that

$$\mathbb{E}\left(\int_{\mathbb{R}} |f_n(x) - \mathbb{E}(f_n(x))|^p dx\right) \le (nh_n)^{-p} \int_{\mathbb{R}} \mathbb{E}\left(\left|\sum_{i=1}^{n} X_{i,n}(x)\right|^p\right) dx\,. \qquad (7.1)$$

Starting from (7.1) and applying Proposition 2.1, we get

$$\begin{aligned}\mathbb{E}\left(\int_{\mathbb{R}} |f_n(x) - \mathbb{E}(f_n(x))|^p dx\right) &\ll (nh_n^2)^{-p/2} \int \left(\sum_{i=0}^{n-1} |\mathrm{Cov}(X_{0,n}(x), X_{i,n}(x))|\right)^{p/2} dx \\ &+ n(nh_n)^{-p} \|dK\|^{p-1} \int \mathbb{E}\left(|Y_{0,n}(x)| A_0(n,p)\right) dx \\ &+ n(nh_n)^{-p} \|dK\|^{p-1} \left(\int \mathbb{E}\left(|Y_{0,n}(x)|\right) dx\right) \sum_{k=0}^{n} (k+1)^{p-2} \beta_{2,Y}(k)\,. \qquad (7.2)\end{aligned}$$

Since

$$\int |Y_{0,n}(x)| dx \le h_n \|K\|_{1,\lambda} \quad \text{and} \quad \mathbb{E}(A_0(n,p)) \le \sum_{k=1}^{n} k^{p-2} \beta_{2,Y}(k)\,,$$

the two last terms on the right hand side of (7.2) are bounded by the second term on the right hand side of (3.3).

To complete the proof, it remains to handle the first term on the right hand side of (7.2). By Item 1 of Lemma 8.1 (see Subsection 8.2) applied to $Z = X_{0,n}(x)$ and

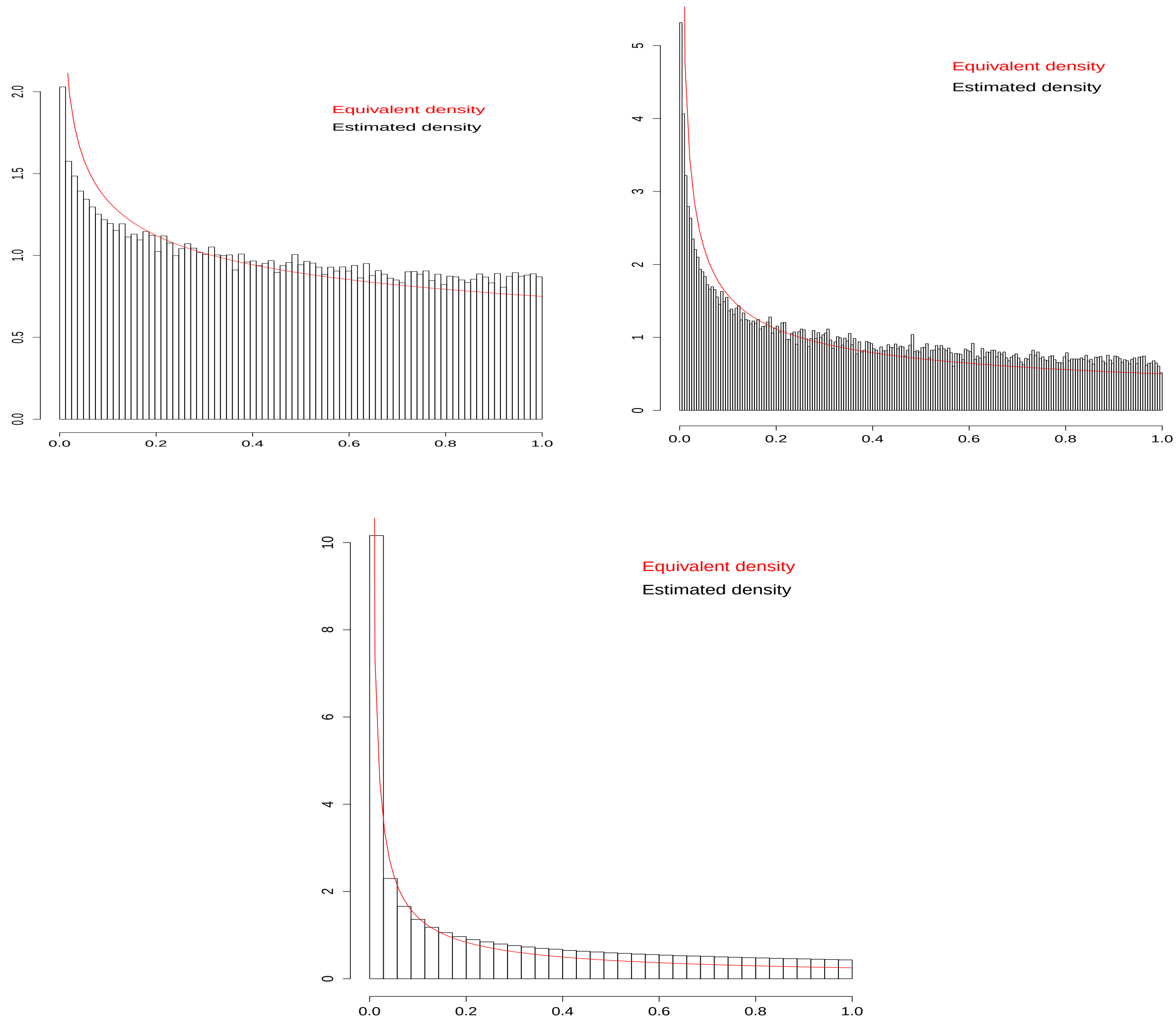

Figure 5: Topleft: Estimation of the density $h_{1/4}$, $n = 60000$. Topright: Estimation of the density $h_{1/2}$, $n = 40000$. Bottom: Estimation of the density $h_{3/4}$, $n = 10^7$. The equivalent density is plotted in red

$\mathcal{F}_0 = \sigma(Y_0)$,

$$\int \left(\sum_{i=0}^{n-1} |\mathrm{Cov}(X_{0,n}(x), X_{i,n}(x))| \right)^{p/2} dx$$
$$\leq \|dK\|^{p/2} \int \left(\int B(y,n) |K((x-y)/h_n)| f(y) dy \right)^{p/2} dx \,, \quad (7.3)$$

where $B(y,n) = b(y,0) + \cdots + b(y,n-1)$ and $b(Y_0,n) = \sup_{t \in \mathbb{R}} |P_{Y_n|Y_0}(f_t) - P(f_t)|$ (keeping the same notations as in Definition 2.1 for the conditional probabilities). By

Jensen's inequality

$$\left(\int B(y,n)|K((x-y)/h_n)|f(y)dy\right)^{p/2} \\ \le h_n^{p/2}\|K\|_{1,\lambda}^{p/2-1}\int B(y,n)^{p/2}f(y)^{p/2}h_n^{-1}|K((x-y)/h_n)|dy\,.$$

Integrating with respect to x, we get

$$\int\left(\int B(y,n)|K((x-y)/h_n)|f(y)dy\right)^{p/2}dx \le h_n^{p/2}\|K\|_{1,\lambda}^{p/2}\int B(y,n)^{p/2}f(y)^{p/2}dy\,.$$

Together with (7.3), this gives

$$(nh_n^2)^{-p/2}\int\left(\sum_{i=0}^{n-1}|\mathrm{Cov}(X_{0,n}(x),X_{i,n}(x))|\right)^{p/2}dx \\ \le (nh_n)^{-p/2}\|dK\|^{p/2}\|K\|_{1,\lambda}^{p/2}\int B(y,n)^{p/2}f(y)^{p/2}dy\,. \quad (7.4)$$

Applying Hölder's inequality,

$$\int B(y,n)^{p/2}f(y)^{p/2}dy \le \|f\|_{p,\lambda}^{p(p-2)/2(p-1)}\left(\int B(y,n)^{p-1}f(y)dy\right)^{p/2(p-1)}. \quad (7.5)$$

Now

$$\left(\int B(y,n)^{p-1}f(y)dy\right)^{1/(p-1)} = \left(\mathbb{E}\left(B(Y_0,n)^{p-1}\right)\right)^{1/(p-1)} = \mathbb{E}\left(Z_n(Y_0)B(Y_0,n)\right)\,,$$

where

$$Z_n(Y_0) = \frac{B(Y_0,n)^{p-2}}{\left(\mathbb{E}\left(B(Y_0,n)^{p-1}\right)\right)^{(p-2)/(p-1)}}\,.$$

Note that $\|Z_n(Y_0)\|_{(p-1)/(p-2)} = 1$. Arguing as in [13] (applying Remark 1.6 with $g^2(x) = Z_n(x)$ and Inequality C.5), we infer that

$$\left(\int B(y,n)^{p-1}f(y)dy\right)^{p/2(p-1)} \ll (V_{1,p,Y}(n))^{p/2(p-1)}\,. \quad (7.6)$$

Combining (7.4), (7.5) and (7.6), the proof of Proposition 3.1 is complete.

7.2 Proof of Proposition 3.2

We keep the same notations as in Subsection 7.1. The case $p = 2$ (Item 1 of Proposition 3.2) has been treated in [7].

For $p \in [1,2)$, we start from the elementary inequality

$$\mathbb{E}\left(\int_{\mathbb{R}}|f_n(x)-\mathbb{E}(f_n(x))|^p dx\right) \le \int\left(\mathbb{E}\left(|f_n(x)-\mathbb{E}(f_n(x))|^2\right)\right)^{p/2}dx\,.$$

Let $\alpha > 1$. Applying Hölder's inequality,

$$\mathbb{E}\left(\int_{\mathbb{R}} |f_n(x) - \mathbb{E}(f_n(x))|^p dx\right) \ll \left(\int \left(|x|^{\alpha(2-p)/p} + 1\right) \mathbb{E}\left(|f_n(x) - \mathbb{E}(f_n(x))|^2\right) dx\right)^{p/2} .$$

As in (7.3), we infer that

$$\mathbb{E}\left(\int_{\mathbb{R}} |f_n(x) - \mathbb{E}(f_n(x))|^p dx\right)$$
$$\ll (nh_n)^{-p/2}\|dK\|^{p/2}\left(\int \mathbb{E}\left(\left(|x|^{\alpha(2-p)/p} + 1\right) B(Y_0, n) h_n^{-1}|K((x - Y_0)/h_n)|\right) dx\right)^{p/2} . \tag{7.7}$$

Now $|x|^{\alpha(2-p)/p} \le C h_n^{\alpha(2-p)/p}|(x-Y_0)/h_n|^{\alpha(2-p)/p} + C|Y_0|^{\alpha(2-p)/p}$ for some positive constant C. Plugging this upper bound in (7.7) and integrating with respect to x we get

$$\mathbb{E}\left(\int_{\mathbb{R}} |f_n(x) - \mathbb{E}(f_n(x))|^p dx\right)$$
$$\ll (nh_n)^{-p/2}\|dK\|^{p/2}\|K\|_{1,\lambda}^{p/2}\left(\mathbb{E}\left(|Y_0|^{\alpha(2-p)/p}B(Y_0, n)\right)\right)^{p/2}$$
$$+ (nh_n)^{-p/2}\|dK\|^{p/2}\left(\left(h_n^{\alpha(2-p)/p}M_{\alpha,p}(K) + \|K\|_{1,\lambda}\right) V_{1,2,Y}(n)\right)^{p/2} . \tag{7.8}$$

To complete the proof, it remains to handle the first term in the right hand side of (7.8). We use once more Hölder's inequality: for any $q > 1$,

$$\mathbb{E}\left(|Y_0|^{\alpha(2-p)/p}B(Y_0, n)\right) \le (M_{\alpha q,p}(f))^{1/q}\left(\mathbb{E}\left(B(Y_0, n)^{q/(q-1)}\right)\right)^{(q-1)/q}$$
$$\ll (M_{\alpha q,p}(f))^{1/q}\left(U_{1,q,Y}(n)\right)^{(q-1)/q} , \tag{7.9}$$

where the last upper bound is proved as in (7.6). Combining (7.8) and (7.9), the proof of Proposition 3.2 is complete.

7.3 Proof of Proposition 4.1

We shall use the following notation

$$\nu_n(g) = \frac{1}{n}\sum_{k=1}^{n}\left(g(Y_k) - \mathbb{E}(g(Y_0))\right) .$$

With this notation, we have

$$\int_0^1 |f_n(x) - \mathbb{E}(f_n(x))|^p dx = \int_0^1 \sum_{j=1}^{m_n}\left|\sum_{i=1}^{r+1}\nu_n(\varphi_{i,j})\varphi_{i,j}(x)\right|^p dx$$
$$\le (r+1)^{p-1}\sum_{j=1}^{m_n}\sum_{i=1}^{r+1}\left(\int_0^1 |\varphi_{i,j}(x)|^p dx\right)|\nu_n(\varphi_{i,j})|^p . \tag{7.10}$$

Now, by definition of $\varphi_{i,j}$,

$$\int_0^1 |\varphi_{i,j}(x)|^p dx \le \|R_i\|_\infty^p m_n^{p/2-1} .$$

Consequently

$$\mathbb{E}\left(\int_0^1 |f_n(x)-\mathbb{E}(f_n(x))|^p dx\right) \ll m_n^{p/2-1} \sum_{i=1}^{r+1} \|R_i\|_\infty^p \sum_{j=1}^{m_n} \mathbb{E}\left(|\nu_n(\varphi_{i,j})|^p\right) . \qquad (7.11)$$

Applying Proposition 2.1, we get

$$\begin{aligned}\mathbb{E}\left(|\nu_n(\varphi_{i,j})|^p\right) \ll \frac{1}{n^{p/2}} &\left(\sum_{k=0}^{n-1} |\mathrm{Cov}(\varphi_{i,j}(Y_0), \varphi_{i,j}(Y_k))|\right)^{p/2} \\ &+ \frac{1}{n^{p-1}} \|d\varphi_{i,j}\|^{p-1} \mathbb{E}(\varphi_{i,j}(Y_0) A_0(n,p)) \\ &\quad + \frac{1}{n^{p-1}} \|d\varphi_{i,j}\|^{p-1} \mathbb{E}(\varphi_{i,j}(Y_0)) \sum_{k=0}^{n} (k+1)^{p-2} \beta_{2,Y}(k) . \qquad (7.12)\end{aligned}$$

Since

$$\|d\varphi_{i,j}\| = \sqrt{m_n}\|dR_i\| , \quad \sum_{j=1}^{m_n} |\varphi_{i,j}| \le \sqrt{m_n}\|R_i\|_\infty \quad \text{and} \quad \mathbb{E}(A_0(n)) \le \sum_{k=1}^{n} k^{p-2}\beta_{2,Y}(k) , \qquad (7.13)$$

the two last terms of the right hand side can be easily bounded, and we obtain

$$\begin{aligned}\sum_{j=1}^{m_n} \mathbb{E}\left(|\nu_n(\varphi_{i,j})|^p\right) \ll \frac{1}{n^{p/2}} \sum_{j=1}^{m_n} &\left(\sum_{k=0}^{n-1} |\mathrm{Cov}(\varphi_{i,j}(Y_0), \varphi_{i,j}(Y_k))|\right)^{p/2} \\ &+ \frac{{m_n}^{p/2}}{n^{p-1}} \|dR_i\|^{p-1} \|R_i\|_\infty \sum_{k=0}^{n} (k+1)^{p-2}\beta_{2,Y}(k) . \qquad (7.14)\end{aligned}$$

It remains to control the first term on the right hand side of (7.14). Applying Lemma 8.1 as in (7.3), we get

$$\left(\sum_{k=0}^{n-1} |\mathrm{Cov}(\varphi_{i,j}(Y_0), \varphi_{i,j}(Y_k))|\right)^{p/2} \le m_n^{p/4} \|dR_i\|^{p/2} \left(\mathbb{E}\left(B(Y_0,n)\varphi_{i,j}(Y_0)\right)\right)^{p/2} , \qquad (7.15)$$

where $B(y,n)$ has been defined right after (7.3). Now, by Jensen's inequality,

$$\left(\mathbb{E}\left(B(Y_0,n)\varphi_{i,j}(Y_0)\right)\right)^{p/2} \le \left(\int |\varphi_{i,j}(x)|dx\right)^{p/2-1} \int_0^1 B(x,n)^{p/2} f(x)^{p/2} |\varphi_{i,j}(x)| dx ,$$

and consequently, using the first part of (7.13),

$$m_n^{p/4} \sum_{j=1}^{m_n} \left(\mathbb{E}\left(B(Y_0,n)\varphi_{i,j}(Y_0)\right)\right)^{p/2} \le \|R_i\|_\infty^{p/2} m_n \int_0^1 B(x,n)^{p/2} f(x)^{p/2} dx . \qquad (7.16)$$

From (7.14), (7.15), (7.16) and arguing as in (7.5)-(7.6), we get that

$$\begin{aligned}\sum_{j=1}^{m_n} \mathbb{E}\left(|\nu_n(\varphi_{i,j})|^p\right) \ll \frac{m_n}{n^{p/2}} &\|dR_i\|^{p/2} \|R_i\|_\infty^{p/2} \|f \mathbf{1}_{[0,1]}\|_{p,\lambda}^{p(p-2)/2(p-1)} (V_{1,p,Y}(n))^{p/2(p-1)} \\ &+ \frac{{m_n}^{p/2}}{n^{p-1}} \|dR_i\|^{p-1} \|R_i\|_\infty \sum_{k=0}^{n} (k+1)^{p-2}\beta_{2,Y}(k) . \qquad (7.17)\end{aligned}$$

Combining (7.11) and (7.17), the result follows.

7.4 Proof of Proposition 4.2

If $p \in [1,2]$, the following inequality holds:

$$\mathbb{E}\left(\int_0^1 |f_n(x) - \mathbb{E}(f_n(x))|^p dx\right) \leq \left(\mathbb{E}\left(\int_0^1 |f_n(x) - \mathbb{E}(f_n(x))|^2 dx\right)\right)^{p/2} . \tag{7.18}$$

To control the right hand term of (7.18), we apply (7.11) with $p = 2$. For $p = 2$, the upper bound (7.12) becomes simply

$$\mathbb{E}\left(|\nu_n(\varphi_{i,j})|^2\right) \ll \frac{1}{n}\sum_{k=0}^{n-1} |\mathrm{Cov}(\varphi_{i,j}(Y_0), \varphi_{i,j}(Y_k))| \ ,$$

and, following the computations of Subsection 7.3, we get

$$\sum_{j=1}^{m_n} \mathbb{E}\left(|\nu_n(\varphi_{i,j})|^2\right) \ll \frac{m_n}{n} \|dR_i\| \|R_i\|_\infty V_{1,2,Y}(n) . \tag{7.19}$$

The result follows from (7.18), (7.11) with $p = 2$, and (7.19).

8 Deviation and Rosenthal bounds for partial sums of bounded random variables

Before proving Proposition 2.1 in Subsection 8.2, we shall state and prove two intermediate results in Subsection 8.1: a deviation inequality for stationary sequences of bounded random variables (see Proposition 8.1) and a Rosenthal-type inequality in the same context (see Corollary 8.1).

8.1 A deviation inequality and a Rosenthal inequality

In this subsection, $(X_i)_{i\in\mathbb{Z}}$ is a strictly stationary sequence of real-valued random variables such that $|X_0| \leq M$ almost surely and $\mathbb{E}(X_0) = 0$. We denote by $\mathcal{F}_i$ the σ-algebra $\mathcal{F}_i = \sigma(X_k, k \leq i)$, and by $\mathbb{E}_i(\cdot)$ the conditional expectation with respect to $\mathcal{F}_i$.

Proposition 8.1. *Let $S_n = \sum_{k=1}^n X_k$. For any $x \geq M$, $r > 2$, $\beta \in]r-2, r[$ and any integer $q \in [1,n]$ such that $qM \leq x$, one has for any $x \geq M$,*

$$\begin{aligned}\mathbb{P}\left(\sup_{1\leq k\leq n} |S_k| \geq 5x\right) \ll \frac{n^{r/2}}{x^r}\left(\sum_{i=0}^{q-1} |\mathrm{Cov}(X_0, X_i)|\right)^{r/2} + \frac{n}{x^r}\|X_1\|_r^r \\ + \frac{n}{x^2 q}\sum_{k=q+1}^{2q}\sum_{\ell=q+1}^{n+q} \|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\|_1 \\ + \frac{n}{x^r} q^{r/2-1}\sum_{i=1}^{q} i^{r/2-2}\left\{ i\|X_0\mathbb{E}_0(X_i)\|_{r/2}^{r/2} + \sum_{j=0}^{i-1} \|\mathbb{E}_0(X_iX_{i+j}) - \mathbb{E}(X_iX_{i+j})\|_{r/2}^{r/2}\right\} \\ + \frac{n}{x^r} q^{r-2-\beta/2}\sum_{\ell=0}^{q-1}\sum_{j=q+1}^{n} j^{\beta/2-1}\|\mathbb{E}_0(X_jX_{j+\ell}) - \mathbb{E}(X_jX_{j+\ell})\|_{r/2}^{r/2} .\end{aligned} \tag{8.1}$$

As a consequence we obtain the following Rosenthal-type inequality:

Corollary 8.1. *Let* $S_n = \sum_{k=1}^n X_k$. *For any* $p > 2$, *any* $r \in]2p-2, 2p[$ *and any* $\beta \in]r-2, 2p-2[$, *one has*

$$\mathbb{E}\Big(\sup_{1\le k\le n} |S_k|^p\Big) \ll n^{p/2}\Big(\sum_{i=0}^{n-1} |\mathrm{Cov}(X_0, X_i)|\Big)^{p/2} + nM^{p-2}\sum_{\ell=0}^{n}\sum_{k=0}^{\ell}(k+1)^{p-3}\,\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\|_1 + nM^{p-2}\sum_{k=1}^{n} k^{p-2}\,\|\mathbb{E}_0(X_k)\|_2^2 + nM^{p-r}\sum_{i=1}^{n} i^{p-2}\|X_0\mathbb{E}_0(X_i)\|_{r/2}^{r/2} + nM^{p-r}\sum_{j=1}^{n} j^{\beta/2-1}\sum_{\ell=0}^{j-1}(\ell+1)^{p-2-\beta/2}\,\|\mathbb{E}_0(X_jX_{j+\ell}) - \mathbb{E}(X_jX_{j+\ell})\|_{r/2}^{r/2}\,. \quad (8.2)$$

Remark 8.1. *Note that the constants that are implicitely involved in Proposition 8.1 and Corollary 8.1 depend only on* r *and* β.

Proof of Proposition 8.1. Let $q \in [1, n]$ be an integer such that $qM \le x$. For any integer i, define the random variables

$$U_i = \sum_{k=(i-1)q+1}^{iq} X_k\,.$$

Consider now the σ-algebras $\mathcal{G}_i = \mathcal{F}_{iq}$ and define the variables $\tilde U_i$ as follows: $\tilde U_{2i-1} = U_{2i-1} - \mathbb{E}(U_{2i-1}|\mathcal{G}_{2(i-1)-1})$ and $\tilde U_{2i} = U_{2i} - \mathbb{E}(U_{2i}|\mathcal{G}_{2(i-1)})$. The following inequality is then valid

$$\max_{1\le k\le n}|S_k| \le 2qM + \max_{2\le 2j\le [n/q]}\Big|\sum_{i=1}^{j}\tilde U_{2i}\Big| + \max_{1\le 2j-1\le [n/q]}\Big|\sum_{i=1}^{j}\tilde U_{2i-1}\Big| + \max_{1\le j\le [n/q]}\Big|\sum_{i=1}^{j}(U_i - \tilde U_i)\Big|\,.$$

It follows that

$$\mathbb{P}\Big(\max_{1\le k\le n}|S_k| \ge 5x\Big) \le \mathbb{P}\Big(\max_{2\le 2j\le [n/q]}\Big|\sum_{i=1}^{j}\tilde U_{2i}\Big| \ge x\Big) + \mathbb{P}\Big(\max_{1\le 2j-1\le [n/q]}\Big|\sum_{i=1}^{j}\tilde U_{2i-1}\Big| \ge x\Big) + \mathbb{P}\Big(\max_{1\le j\le [n/q]}\Big|\sum_{i=1}^{j}(U_i-\tilde U_i)\Big| \ge x\Big)\,. \quad (8.3)$$

Note that the two first terms on the right hand side of (8.3) can be treated similarly, so that we shall only prove an upper bound for the first one.

Let us first deal with the last term on the right hand side of (8.3). By Markov's inequality followed by Proposition 1 in [9], we have

$$\mathbb{P}\Big(\max_{1\le j\le [n/q]}\Big|\sum_{i=1}^{j}(U_i-\tilde U_i)\Big| \ge x\Big) \le \frac{4}{x^2}\sum_{i=1}^{[n/q]}\|\mathbb{E}(U_i|\mathcal{G}_{i-2})\|_2^2 + \frac{8}{x^2}\sum_{i=1}^{[n/q]-1}\Big\|\mathbb{E}(U_i|\mathcal{G}_{i-2})\mathbb{E}\Big(\sum_{j=i+1}^{[n/q]} U_j|\mathcal{G}_{i-2}\Big)\Big\|_1\,.$$

Therefore, by stationarity,

$$\mathbb{P}\left(\max_{1\le j\le [n/q]}\Big|\sum_{i=1}^{j}(U_i-\tilde U_i)\Big|\ge x\right)$$
$$\le \frac{8}{x^2}\sum_{i=1}^{[n/q]}\sum_{k=(i-1)q+1}^{iq}\sum_{j=i}^{[n/q]}\sum_{\ell=(j-1)q+1}^{jq}\left\|\mathbb{E}_{(i-2)q}(X_k)\mathbb{E}_{(i-2)q}(X_\ell)\right\|_1$$
$$\le \frac{8}{x^2}\sum_{i=1}^{[n/q]}\sum_{k=q+1}^{2q}\sum_{j=i}^{[n/q]}\sum_{\ell=(j-i)q+q+1}^{(j-i+2)q}\left\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\right\|_1$$
$$\le \frac{8}{x^2}\sum_{i=0}^{[n/q]-1}\sum_{k=q+1}^{2q}\sum_{j=0}^{i}\sum_{\ell=(j+1)q+1}^{(j+2)q}\left\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\right\|_1 .$$

So, overall,

$$\mathbb{P}\left(\max_{1\le j\le [n/q]}\left|\sum_{i=1}^{j}(U_i-\tilde U_i)\right|\ge x\right)\le \frac{8n}{qx^2}\sum_{k=q+1}^{2q}\sum_{\ell=q+1}^{n+q}\left\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\right\|_1 . \tag{8.4}$$

Now, we handle the first term on the right hand side of (8.3). Using Markov's inequality, we obtain

$$\mathbb{P}\left(\max_{2\le 2j\le [n/q]}\left|\sum_{i=1}^{j}\tilde U_{2i}\right|\ge x\right)\le x^{-r}\left\|\max_{2\le 2j\le [n/q]}\left|\sum_{i=1}^{j}\tilde U_{2i}\right|\right\|_r^r . \tag{8.5}$$

Note that $(\tilde U_{2i})_{i\in\mathbb{Z}}$ (resp. $(\tilde U_{2i-1})_{i\in\mathbb{Z}}$) is a stationary sequence of martingale differences with respect to the filtration $(\mathcal{G}_{2i})_{i\in\mathbb{Z}}$ (resp. $(\mathcal{G}_{2i-1})_{i\in\mathbb{Z}}$). Applying Theorem 6 in [12], we get

$$\left\|\max_{2\le 2j\le [n/q]}\left|\sum_{i=1}^{j}\tilde U_{2i}\right|\right\|_r$$
$$\ll (n/q)^{1/r}\|\tilde U_2\|_r+(n/q)^{1/r}\left(\sum_{k=1}^{[n/(2q)]}\frac{1}{k^{1+2\delta/r}}\left\|\mathbb{E}_0\left(\left(\sum_{i=1}^{k}\tilde U_{2i}\right)^2\right)\right\|_{r/2}^{\delta}\right)^{1/(2\delta)} , \tag{8.6}$$

where $\delta=\min(1,1/(r-2))$. Since $(\tilde U_{2i})_{i\in\mathbb{Z}}$ is a stationary sequence of martingale differences with respect to the filtration $(\mathcal{G}_{2i})_{i\in\mathbb{Z}}$,

$$\mathbb{E}_0\left(\left(\sum_{i=1}^{k}\tilde U_{2i}\right)^2\right)=\sum_{i=1}^{k}\mathbb{E}_0\left(\tilde U_{2i}^2\right) .$$

Moreover, $\mathbb{E}_0(\tilde U_{2i}^2)\le \mathbb{E}_0(U_{2i}^2)$. Therefore

$$\left\|\mathbb{E}_0\left(\left(\sum_{i=1}^{k}\tilde U_{2i}\right)^2\right)\right\|_{r/2}\le \sum_{i=1}^{k}\left\|\mathbb{E}_0\left(U_{2i}^2\right)-\mathbb{E}\left(U_{2i}^2\right)\right\|_{r/2}+\sum_{i=1}^{k}\mathbb{E}\left(U_{2i}^2\right) .$$

By stationarity

$$\sum_{i=1}^{k} \mathbb{E}\left(U_{2i}^2\right) = k\|S_q\|_2^2 .$$

Moreover $\|\tilde{U}_2\|_r \le 2\|S_q\|_r$. From (8.6) and the computations we have made, it follows that

$$\left\| \max_{2\le 2j\le [n/q]} \left| \sum_{i=1}^{j} \tilde{U}_{2i} \right| \right\|_r^r \ll \frac{n}{q}\|S_q\|_r^r + \left(\frac{n}{q}\right)^{r/2} \|S_q\|_2^r + \frac{n}{q} \left(\sum_{k=1}^{[n/(2q)]} \frac{1}{k^{1+2\delta/r}} D_{k,q}^{\delta} \right)^{r/(2\delta)} ,$$

where

$$D_{k,q} = \sum_{i=1}^{k} \left\| \mathbb{E}_0\left(U_{2i}^2\right) - \mathbb{E}\left(U_{2i}^2\right) \right\|_{r/2} .$$

Hence,

$$\left\| \max_{2\le 2j\le [n/q]} \left| \sum_{i=1}^{j} \tilde{U}_{2i} \right| \right\|_r^r \ll \frac{n}{q}\|S_q\|_r^r + n^{r/2} \left(\sum_{i=0}^{q-1} |\mathrm{Cov}(X_0, X_i)| \right)^{r/2} + \frac{n}{q} \left(\sum_{k=1}^{[n/(2q)]} \frac{1}{k^{1+2\delta/r}} D_{k,q}^{\delta} \right)^{r/(2\delta)} . \quad (8.7)$$

Notice that

$$\begin{aligned} D_{k,q} &\le \sum_{i=1}^{k} \sum_{j,k=(2i-1)q+1}^{2iq} \|\mathbb{E}_0(X_j X_k) - \mathbb{E}(X_j X_k)\|_{r/2} \\ &\le 2 \sum_{i=1}^{k} \sum_{j=(2i-1)q+1}^{2iq} \sum_{\ell=0}^{2iq-j} \|\mathbb{E}_0(X_j X_{j+\ell}) - \mathbb{E}(X_j X_{j+\ell})\|_{r/2} . \end{aligned}$$

Let $\eta = (\beta-2)/r$ and recall that $r > 2$ and $r-2 < \beta < r$. Since $\eta < (r-2)/r$, applying Hölder's inequality, we then get that

$$\begin{aligned} D_{k,q} &\ll k^{-\eta+(r-2)/r} \left(\sum_{i=1}^{k} i^{\beta/2-1} \left(\sum_{j=(2i-1)q+1}^{2iq} \sum_{\ell=0}^{q-1} \|\mathbb{E}_0(X_j X_{j+\ell}) - \mathbb{E}(X_j X_{j+\ell})\|_{r/2} \right)^{\frac{r}{2}} \right)^{\frac{2}{r}} \\ &\ll q^{2-4/r} k^{-\eta+(r-2)/r} \left(\sum_{i=1}^{k} i^{\beta/2-1} \sum_{j=(2i-1)q+1}^{2iq} \sum_{\ell=0}^{q-1} \|\mathbb{E}_0(X_j X_{j+\ell}) - \mathbb{E}(X_j X_{j+\ell})\|_{r/2}^{r/2} \right)^{\frac{2}{r}} . \end{aligned}$$

Since $2\delta/r > -\delta\eta + \delta(r-2)/r$ (indeed $-\eta + (r-2)/r = (r-\beta)/r$ and $r - \beta < 2$), it follows that

$$\left(\sum_{k=1}^{[n/(2q)]}\frac{1}{k^{1+2\delta/r}}D_{k,q}^{\delta}\right)^{r/(2\delta)}$$
$$\ll q^{r-2}\sum_{\ell=0}^{q-1}\sum_{i=1}^{[n/(2q)]} i^{\beta/2-1}\sum_{j=(2i-1)q+1}^{2iq}\|\mathbb{E}_0(X_jX_{j+\ell})-\mathbb{E}(X_jX_{j+\ell})\|_{r/2}^{r/2}$$
$$\ll q^{r-\beta/2-1}\sum_{\ell=0}^{q-1}\sum_{i=1}^{[n/(2q)]}\sum_{j=(2i-1)q+1}^{2iq} j^{\beta/2-1}\|\mathbb{E}_0(X_jX_{j+\ell})-\mathbb{E}(X_jX_{j+\ell})\|_{r/2}^{r/2}.$$

So, overall,

$$\frac{n}{q}\left(\sum_{k=1}^{[n/(2q)]}\frac{1}{k^{1+2\delta/r}}D_{k,q}^{\delta}\right)^{r/(2\delta)} \ll nq^{r-\beta/2-2}\sum_{\ell=0}^{q-1}\sum_{j=q+1}^{n} j^{\beta/2-1}\|\mathbb{E}_0(X_jX_{j+\ell})-\mathbb{E}(X_jX_{j+\ell})\|_{r/2}^{r/2}. \tag{8.8}$$

Combining (8.3), (8.4), (8.7) and (8.8), Proposition 8.1 will be proved if we show that

$$\frac{n}{q}\|S_q\|_r^r \ll n\|X_1\|_r^r + nq^{r/2-1}\left(\sum_{i=0}^{q-1}|\mathbb{E}(X_0X_i)|\right)^{r/2}$$
$$+ nq^{r/2-1}\sum_{i=1}^{q} i^{r/2-2}\left\{i\|X_0\mathbb{E}_0(X_i)\|_{r/2}^{r/2}+\sum_{j=0}^{i-1}\|\mathbb{E}_0(X_iX_{i+j})-\mathbb{E}(X_iX_{i+j})\|_{r/2}^{r/2}\right\}. \tag{8.9}$$

By Theorem 6 in [12] again, and taking into account their Comment 7 (Item 4) together with the fact that $\|\mathbb{E}_0(S_k)\|_r \le \|\mathbb{E}_0(S_k^2)\|_{r/2}^{1/2}$, we have

$$\|S_q\|_r^r \ll q\|X_1\|_r^r + q\left(\sum_{k=1}^{q}\frac{1}{k^{1+2\delta/r}}\|\mathbb{E}_0(S_k^2)\|_{r/2}^{\delta}\right)^{r/(2\delta)},$$

where $\delta=\min(1/2,1/(p-2))$. Now,

$$\mathbb{E}(S_k^2)\le 2k\sum_{i=0}^{k-1}|\mathbb{E}(X_0X_i)|.$$

Hence, since $r>2$,

$$\|S_q\|_r^r \ll q\|X_1\|_r^r + q^{r/2}\left(\sum_{i=0}^{q-1}|\mathbb{E}(X_0X_i)|\right)^{r/2} + q\left(\sum_{k=1}^{q}\frac{1}{k^{1+2\delta/r}}\|\mathbb{E}_0(S_k^2)-\mathbb{E}(S_k^2)\|_{r/2}^{\delta}\right)^{r/(2\delta)}. \tag{8.10}$$

Now

$$\|\mathbb{E}_0(S_k^2)-\mathbb{E}(S_k^2)\|_{r/2} \le 2\sum_{i=1}^{k}\sum_{j=0}^{k-i}\|\mathbb{E}_0(X_iX_{i+j})-\mathbb{E}(X_iX_{i+j})\|_{r/2}$$
$$\le 2\sum_{i=1}^{k}\sum_{j=0}^{i-1}\|\mathbb{E}_0(X_iX_{i+j})-\mathbb{E}(X_iX_{i+j})\|_{r/2} + 2\sum_{i=1}^{k}\sum_{j=i}^{k}\|\mathbb{E}_0(X_iX_{i+j})-\mathbb{E}(X_iX_{i+j})\|_{r/2}.$$

Note that, by stationarity,

$$\|\mathbb{E}_0(X_iX_{i+j}) - \mathbb{E}(X_iX_{i+j})\|_{r/2} \le 2\|X_i\mathbb{E}_i(X_{i+j})\|_{r/2} = 2\|X_0\mathbb{E}_0(X_j)\|_{r/2}\,.$$

Therefore

$$\left\|\mathbb{E}_0(S_k^2) - \mathbb{E}(S_k^2)\right\|_{r/2} \le 2\sum_{i=1}^{k}\sum_{j=0}^{i-1}\|\mathbb{E}_0(X_iX_{i+j}) - \mathbb{E}(X_iX_{i+j})\|_{r/2} + 4\sum_{j=1}^{k} j\|X_0\mathbb{E}_0(X_j)\|_{r/2}\,.$$

Applying Hölder's inequality,

$$\sum_{j=1}^{k} j\|X_0\mathbb{E}_0(X_j)\|_{r/2} \le k\left(\sum_{j=1}^{k} j^{r/2-1}\|X_0\mathbb{E}_0(X_j)\|_{r/2}^{r/2}\right)^{2/r}\,,$$

and

$$\begin{aligned}
&\sum_{i=1}^{k}\sum_{j=0}^{i-1}\|\mathbb{E}_0(X_iX_{i+j}) - \mathbb{E}(X_iX_{i+j})\|_{r/2} \\
&\qquad\le k^{1-2/r}\left(\sum_{i=1}^{k}\left(\sum_{j=0}^{i-1}\|\mathbb{E}_0(X_iX_{i+j}) - \mathbb{E}(X_iX_{i+j})\|_{r/2}\right)^{r/2}\right)^{2/r} \\
&\qquad\le k^{1-2/r}\left(\sum_{i=1}^{k} i^{r/2-1}\sum_{j=0}^{i-1}\|\mathbb{E}_0(X_iX_{i+j}) - \mathbb{E}(X_iX_{i+j})\|_{r/2}^{r/2}\right)^{2/r} \\
&\qquad\le k\left(\sum_{i=1}^{k} i^{r/2-2}\sum_{j=0}^{i-1}\|\mathbb{E}_0(X_iX_{i+j}) - \mathbb{E}(X_iX_{i+j})\|_{r/2}^{r/2}\right)^{2/r}\,.
\end{aligned}$$

So, overall,

$$\begin{aligned}
&q\left(\sum_{k=1}^{q}\frac{1}{k^{1+2\delta/r}}\|\mathbb{E}_0(S_k^2) - \mathbb{E}(S_k^2)\|_{r/2}^{\delta}\right)^{r/(2\delta)} \\
&\ll q^{r/2}\left\{\sum_{j=1}^{q} j^{r/2-1}\|X_0\mathbb{E}_0(X_j)\|_{r/2}^{r/2} + \sum_{i=1}^{q} i^{r/2-2}\sum_{j=0}^{i-1}\|\mathbb{E}_0(X_iX_{i+j}) - \mathbb{E}(X_iX_{i+j})\|_{r/2}^{r/2}\right\}\,.
\end{aligned}$$

Taking into account this last upper bound in (8.10), we obtain (8.9). The proof of Proposition 8.1 is complete.

Proof of Corollary 8.1. Setting

$$s_n^2 = \max\left(n\sum_{i=0}^{n-1}|\mathrm{Cov}(X_0, X_i)|, M^2\right)\,, \tag{8.11}$$

we have

$$\begin{aligned}
\mathbb{E}\left(\sup_{1\le k\le n}|S_k|^p\right) &= p\int_0^{nM} x^{p-1}\mathbb{P}\left(\sup_{1\le k\le n}|S_k|\ge x\right)dx \\
&= 5^p p\int_0^{nM/5} x^{p-1}\mathbb{P}\left(\sup_{1\le k\le n}|S_k|\ge 5x\right)dx \\
&\le 5^p p\int_0^{s_n} x^{p-1}\mathbb{P}\left(\sup_{1\le k\le n}|S_k|\ge 5x\right)dx + 5^p p\int_{s_n}^{nM} x^{p-1}\mathbb{P}\left(\sup_{1\le k\le n}|S_k|\ge 5x\right)dx\,.
\end{aligned} \tag{8.12}$$

To handle the first term on the right hand side of (8.12), we first note that

$$5^p p \int_0^{s_n} x^{p-1} \mathbb{P}\left(\sup_{1\le k\le n} |S_k| \ge 5x \right) dx \le 5^p p \int_0^{M} x^{p-1} \mathbb{P}\left(\sup_{1\le k\le n} |S_k| \ge 5x \right) dx + 5^p n^{p/2} \left(\sum_{i=0}^{n-1} |\mathrm{Cov}(X_0, X_i)| \right)^{p/2} .$$

Now by Markov inequality followed by Proposition 1 in [9], we have

$$5^p \int_0^{M} x^{p-1} \mathbb{P}\left(\sup_{1\le k\le n} |S_k| \ge 5x \right) dx \le 4 \times 5^{p-2}(p-2)^{-1} M^{p-2} \left\{ \sum_{i=1}^{n} \mathbb{E}(X_i^2) + 2 \sum_{i=1}^{n-1} \left\| X_i \mathbb{E}_i \left(\sum_{j=i+1}^{n} X_j \right) \right\|_1 \right\} .$$

Hence by stationarity

$$5^p \int_0^{M} x^{p-1} \mathbb{P}\left(\sup_{1\le k\le n} |S_k| \ge 5x \right) dx \le 8 \times 5^{p-2}(p-2)^{-1} M^{p-2} n \sum_{j=0}^{n-1} \|X_0 \mathbb{E}_0(X_j)\|_1 .$$

So, overall,

$$p5^p \int_0^{s_n} x^{p-1} \mathbb{P}\left(\sup_{1\le k\le n} |S_k| \ge x \right) dx \le 8p \times 5^{p-2} n (p-2)^{-1} M^{p-2} \sum_{j=0}^{n-1} \|X_0 \mathbb{E}_0(X_j)\|_1 + 5^p n^{p/2} \left(\sum_{i=0}^{n-1} |\mathrm{Cov}(X_0, X_i)| \right)^{p/2} . \quad (8.13)$$

We handle now the last term on the right hand side of (8.12). With this aim, we use Proposition 8.1 with $q = [x/M]$, $r \in]2p-2, 2p[$ and $\beta \in]r-2, 2p-2[$.

For the first term on the right hand side of Proposition 8.1, we have

$$n^{r/2} \int_{s_n}^{nM} x^{p-1-r} \left(\sum_{i=0}^{q-1} |\mathrm{Cov}(X_0, X_i)| \right)^{r/2} dx \ll n^{p/2} \left(\sum_{i=0}^{n-1} |\mathrm{Cov}(X_0, X_i)| \right)^{p/2} , \quad (8.14)$$

since $s_n^2 \ge n \sum_{i=0}^{n-1} |\mathrm{Cov}(X_0, X_i)|$ and $r > p$.

For the second term on the right hand side of Proposition 8.1, since $r > p$ and $s_n \ge M$,

$$n\|X_1\|_r^r \int_{s_n}^{nM} x^{p-1-r} dx \ll n\|X_1\|_r^r s_n^{p-r} \ll n\|X_1\|_p^p M^{r-p} s_n^{p-r} \ll n\|X_1\|_p^p \ll nM^{p-2}\|X_1\|_2^2 . \quad (8.15)$$

For the third term on the right hand side of Proposition 8.1, we have to give an upper bound for

$$n \int_{s_n}^{nM} x^{p-1} \frac{1}{x^2 q} \sum_{k=q+1}^{2q} \sum_{\ell=q+1}^{n+q} \|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\|_1 \, dx .$$

Write

$$\sum_{k=q+1}^{2q}\sum_{\ell=q+1}^{n+q}\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\|_1=\sum_{k=q+1}^{2q}\sum_{\ell=q+1}^{n}\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\|_1 + \sum_{k=q+1}^{2q}\sum_{\ell=n+1}^{n+q}\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\|_1\,.$$

Here, note that

$$\sum_{k=q+1}^{2q}\sum_{\ell=n+1}^{n+q}\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\|_1\leq\frac{q}{2}\sum_{k=q+1}^{2q}\|\mathbb{E}_0(X_k)\|_2^2+\frac{q}{2}\sum_{\ell=n+1}^{n+q}\|\mathbb{E}_0(X_\ell)\|_2^2\,.$$

Therefore

$$n\int_{s_n}^{nM}x^{p-1}\frac{1}{x^2q}\sum_{k=q+1}^{2q}\sum_{\ell=n+1}^{n+q}\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\|_1\,dx\leq\frac{n}{2}\int_{s_n}^{nM}x^{p-3}\sum_{k=q+1}^{2q}\|\mathbb{E}_0(X_k)\|_2^2\,dx + \frac{n}{2}\int_{s_n}^{nM}x^{p-3}\sum_{\ell=n+1}^{n+q}\|\mathbb{E}_0(X_\ell)\|_2^2\,dx\,.$$

Now

$$\begin{aligned}n\int_{s_n}^{nM}x^{p-3}\sum_{k=q+1}^{2q}\|\mathbb{E}_0(X_k)\|_2^2\,dx&=n\int_{s_n}^{nM}x^{p-3}\sum_{k=q+1}^{2q}\|\mathbb{E}_0(X_k)\|_2^2\,\mathbf{1}_{q\leq[n/2]}dx\\&\quad+n\int_{s_n}^{nM}x^{p-3}\sum_{k=q+1}^{2q}\|\mathbb{E}_0(X_k)\|_2^2\,\mathbf{1}_{q>[n/2]}dx\\&\leq 2n\int_{s_n}^{nM}x^{p-3}\sum_{k=q+1}^{2q}\|\mathbb{E}_0(X_k)\|_2^2\,\mathbf{1}_{q\leq[n/2]}dx+n^2\int_{s_n}^{nM}x^{p-3}\left\|\mathbb{E}_0(X_{[n/2]+1})\right\|_2^2dx\,,\end{aligned}$$

where we have used the fact that $q\leq n$. Hence

$$\begin{aligned}n\int_{s_n}^{nM}x^{p-3}\sum_{k=q+1}^{2q}\|\mathbb{E}_0(X_k)\|_2^2\,dx&\leq 2n\int_{s_n}^{nM}x^{p-3}\sum_{k=q+1}^{n}\|\mathbb{E}_0(X_k)\|_2^2\,\mathbf{1}_{x\leq kM}dx\\&\quad+n^2\int_{s_n}^{nM}x^{p-3}\left\|\mathbb{E}_0(X_{[n/2]+1})\right\|_2^2dx\\&\leq\frac{2n}{p-2}M^{p-2}\sum_{k=1}^{n}k^{p-2}\|\mathbb{E}_0(X_k)\|_2^2+\frac{n^p}{p-2}M^{p-2}\left\|\mathbb{E}_0(X_{[n/2]+1})\right\|_2^2\,.\end{aligned}$$

Now

$$n^{p-1}\leq 2^{p-1}([n/2]+1)^{p-1}\leq 2^{p-1}(p-1)\sum_{k=1}^{[n/2]+1}k^{p-2}\,,$$

and therefore,

$$n^p\left\|\mathbb{E}_0(X_{[n/2]+1})\right\|_2^2\leq 2^{p-1}(p-1)n\sum_{k=1}^{n}k^{p-2}\|\mathbb{E}_0(X_k)\|_2^2\,.$$

Consequently

$$n\int_{s_n}^{nM} x^{p-3}\sum_{k=q+1}^{2q}\|\mathbb{E}_0(X_k)\|_2^2\,dx \ll nM^{p-2}\sum_{k=1}^{n}k^{p-2}\|\mathbb{E}_0(X_k)\|_2^2\,.$$

On another hand, since $q\le n$,

$$n\int_{s_n}^{nM} x^{p-3}\sum_{\ell=n+1}^{n+q}\|\mathbb{E}_0(X_\ell)\|_2^2\,dx \le \frac{n^pM^{p-2}}{p-2}\|\mathbb{E}_0(X_{n+1})\|_2^2\,.$$

Proceeding as before, we get

$$n\int_{s_n}^{nM} x^{p-3}\sum_{\ell=n+1}^{n+q}\|\mathbb{E}_0(X_\ell)\|_2^2\,dx \ll nM^{p-2}\sum_{k=1}^{n}k^{p-2}\|\mathbb{E}_0(X_k)\|_2^2\,.$$

So, overall,

$$n\int_{s_n}^{nM} x^{p-1}\frac{1}{x^2q}\sum_{k=q+1}^{2q}\sum_{\ell=n+1}^{n+q}\big\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\big\|_1dx \ll nM^{p-2}\sum_{k=1}^{n}k^{p-2}\big\|\mathbb{E}_0(X_k)\big\|_2^2.$$

We handle now the quantity

$$n\int_{s_n}^{nM} x^{p-1}\frac{1}{x^2q}\sum_{k=q+1}^{2q}\sum_{\ell=q+1}^{n}\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\|_1\,dx\,.$$

We note first that

$$\begin{aligned}n\int_{s_n}^{nM} x^{p-1}\frac{1}{x^2q}&\sum_{k=q+1}^{2q}\sum_{\ell=q+1}^{n}\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\|_1\,dx\\ &\le 2n\int_{s_n}^{nM} x^{p-3}\sum_{k=q+1}^{2q}k^{-1}\sum_{\ell=q+1}^{n}\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\|_1\,dx\,.\end{aligned}$$

Now

$$\begin{aligned}\sum_{k=q+1}^{2q}k^{-1}&\sum_{\ell=q+1}^{n}\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\|_1\\ &=\sum_{k=q+1}^{2q}k^{-1}\sum_{\ell=q+1}^{k}\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\|_1+\sum_{k=q+1}^{2q}k^{-1}\sum_{\ell=k+1}^{n}\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\|_1\\ &\le\sum_{k=q+1}^{2q}\sum_{\ell=q+1}^{k}\ell^{-1}\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\|_1+\sum_{\ell=q+2}^{n}\sum_{k=q+1}^{\ell-1}k^{-1}\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\|_1\,.\end{aligned}$$

Note that

$$\begin{aligned}n\int_{s_n}^{nM} x^{p-3}&\sum_{\ell=q+2}^{n}\sum_{k=q+1}^{\ell-1}k^{-1}\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\|_1\,dx\\ &\le n\sum_{\ell=1}^{n}\sum_{k=1}^{\ell}k^{-1}\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\|_1\int_{s_n}^{nM}x^{p-3}\mathbf{1}_{x\le kM}dx\\ &\ll\frac{nM^{p-2}}{p-2}\sum_{\ell=1}^{n}\sum_{k=1}^{\ell}k^{p-3}\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\|_1\,.\end{aligned}$$

On the other hand

$$
\begin{aligned}
n\int_{s_n}^{nM} x^{p-3}\sum_{k=q+1}^{2q}\sum_{\ell=q+1}^{k} \ell^{-1}\left\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\right\|_1 dx \\
&= n\int_{s_n}^{nM} x^{p-3}\sum_{k=q+1}^{2q}\sum_{\ell=q+1}^{k} \ell^{-1}\left\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\right\|_1 \mathbf{1}_{q\le[n/2]}dx \\
&+ n\int_{s_n}^{nM} x^{p-3}\sum_{k=q+1}^{2q}\sum_{\ell=q+1}^{k} \ell^{-1}\left\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\right\|_1 \mathbf{1}_{q>[n/2]}dx \\
&\le n\int_{s_n}^{nM} x^{p-3}\sum_{k=q+1}^{n}\sum_{\ell=q+1}^{k} \ell^{-1}\left\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\right\|_1 \mathbf{1}_{x\le \ell M}dx \\
&\qquad + n\int_{s_n}^{nM} x^{p-3}\sum_{k=q+1}^{2q}\sum_{\ell=q+1}^{k} \ell^{-1}\left\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\right\|_1 \mathbf{1}_{[n/2]<q\le n}dx\,.
\end{aligned}
$$

Proceeding as before

$$
n\int_{s_n}^{nM} x^{p-3}\sum_{k=q+1}^{n}\sum_{\ell=q+1}^{k} \ell^{-1}\left\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\right\|_1 \mathbf{1}_{x\le \ell M}dx
\ll nM^{p-2}\sum_{\ell=1}^{n}\sum_{k=1}^{\ell} k^{p-3}\left\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\right\|_1
$$

and

$$
n\int_{s_n}^{nM} x^{p-3}\sum_{k=q+1}^{2q}\sum_{\ell=q+1}^{k} \ell^{-1}\left\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\right\|_1 \mathbf{1}_{[n/2]<q\le n}dx
\ll n^p M^{p-2}\left\|\mathbb{E}_0(X_{[n/2]+1})\right\|_2^2 \ll nM^{p-2}\sum_{k=1}^{n} k^{p-2}\left\|\mathbb{E}_0(X_k)\right\|_2^2\,.
$$

So, overall, we obtain the following upper bound

$$
\begin{aligned}
n\int_{s_n}^{nM} x^{p-1}\frac{1}{x^2 q}\sum_{k=q+1}^{2q}\sum_{\ell=q+1}^{n+q} \Big\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\Big\|_1 dx \\
\ll nM^{p-2}\sum_{\ell=1}^{n}\sum_{k=1}^{\ell} k^{p-3}\Big\|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)\Big\|_1 + nM^{p-2}\sum_{k=1}^{n} k^{p-2}\Big\|\mathbb{E}_0(X_k)\Big\|_2^2. \quad (8.16)
\end{aligned}
$$

For the fourth term on the right hand side of Proposition 8.1, setting

$$
a(i) = i\|X_0\mathbb{E}_0(X_i)\|_{r/2}^{r/2} + \sum_{j=0}^{i-1}\|\mathbb{E}_0(X_iX_{i+j}) - \mathbb{E}(X_iX_{i+j})\|_{r/2}^{r/2}\,,
$$

we have

$$
n\int_{s_n}^{nM} x^{p-1-r}q^{r/2-1}\sum_{i=1}^{q} i^{r/2-2}a(i)dx \le nM^{1-r/2}\sum_{i=1}^{n} i^{r/2-2}a(i)\int_{s_n}^{nM} x^{p-2-r/2}\mathbf{1}_{x\ge iM}dx\,.
$$

Since $r > 2p-2$ and $s_n \geq M$, it follows that

$$n\int_{s_n}^{nM} x^{p-1-r} q^{r/2-1} \sum_{i=1}^{q} i^{r/2-2} a(i) dx \ll nM^{p-r} \sum_{i=1}^{n} i^{p-3} a(i) \,.$$

We note also that $\beta/2 > r/2-1 > p-2$. Therefore

$$\sum_{i=1}^{n} i^{p-3} \sum_{j=0}^{i-1} \|\mathbb{E}_0(X_i X_{i+j}) - \mathbb{E}(X_i X_{i+j})\|_{r/2}^{r/2}$$
$$= \sum_{i=1}^{n} i^{\beta/2-1+p-2-\beta/2} \sum_{j=0}^{i-1} \|\mathbb{E}_0(X_i X_{i+j}) - \mathbb{E}(X_i X_{i+j})\|_{r/2}^{r/2}$$
$$\leq \sum_{i=1}^{n} i^{\beta/2-1} \sum_{j=0}^{i-1} (j+1)^{p-2-\beta/2} \|\mathbb{E}_0(X_i X_{i+j}) - \mathbb{E}(X_i X_{i+j})\|_{r/2}^{r/2} \,,$$

so that

$$n\int_{s_n}^{nM} x^{p-1-r} q^{r/2-1} \sum_{i=1}^{q} i^{r/2-2} a(i) dx \ll nM^{p-r} \sum_{i=1}^{n} i^{p-2} \|X_0 \mathbb{E}_0(X_i)\|_{r/2}^{r/2}$$
$$+ nM^{p-r} \sum_{i=1}^{n} i^{\beta/2-1} \sum_{j=0}^{i-1} (j+1)^{p-2-\beta/2} \|\mathbb{E}_0(X_i X_{i+j}) - \mathbb{E}(X_i X_{i+j})\|_{r/2}^{r/2} \,. \quad (8.17)$$

Finally, for the fifth term on the right hand side of of Proposition 8.1,

$$n\int_{s_n}^{nM} x^{p-1-r} q^{r-2-\beta/2} \sum_{\ell=0}^{q-1} \sum_{j=q+1}^{n} j^{\beta/2-1} \|\mathbb{E}_0(X_j X_{j+\ell}) - \mathbb{E}(X_j X_{j+\ell})\|_{r/2}^{r/2} dx$$
$$\leq nM^{p-1-r} \int_{s_n}^{nM} q^{p-3-\beta/2} \sum_{\ell=0}^{q-1} \sum_{j=q+1}^{n} j^{\beta/2-1} \|\mathbb{E}_0(X_j X_{j+\ell}) - \mathbb{E}(X_j X_{j+\ell})\|_{r/2}^{r/2} dx \,,$$

since $r > p-1$ and $q < x/M$. Now, since $p-3-\beta/2 < -1$ (indeed $\beta/2 > r/2-1$ and $r/2 > p-1$), we get

$$n\int_{s_n}^{nM} x^{p-1-r} q^{r-2-\beta/2} \sum_{\ell=0}^{q-1} \sum_{j=q+1}^{n} j^{\beta/2-1} \|\mathbb{E}_0(X_j X_{j+\ell}) - \mathbb{E}(X_j X_{j+\ell})\|_{r/2}^{r/2} dx$$
$$\leq n(2M)^{\beta/2+3-p} M^{p-1-r}$$
$$\times \int_{s_n}^{nM} x^{p-3-\beta/2} \sum_{\ell=0}^{q-1} \sum_{j=q+1}^{n} j^{\beta/2-1} \|\mathbb{E}_0(X_j X_{j+\ell}) - \mathbb{E}(X_j X_{j+\ell})\|_{r/2}^{r/2} dx$$
$$\leq n(2M)^{\beta/2+3-p} M^{p-1-r}$$
$$\times \sum_{\ell=0}^{n-1} \sum_{j=\ell+1}^{n} j^{\beta/2-1} \|\mathbb{E}_0(X_j X_{j+\ell}) - \mathbb{E}(X_j X_{j+\ell})\|_{r/2}^{r/2} \int_{s_n}^{nM} x^{p-3-\beta/2} \mathbf{1}_{x \geq (\ell+1)M} dx \,.$$

Hence, since $s_n \geq M$,

$$n \int_{s_n}^{nM} x^{p-1-r} q^{r-2-\beta/2} \sum_{\ell=0}^{q-1} \sum_{j=q+1}^{n} j^{\beta/2-1} \|\mathbb{E}_0(X_j X_{j+\ell}) - \mathbb{E}(X_j X_{j+\ell})\|_{r/2}^{r/2} dx$$
$$\ll nM^{p-r} \sum_{\ell=0}^{n-1} (\ell+1)^{p-2-\beta/2} \sum_{j=\ell+1}^{n} j^{\beta/2-1} \|\mathbb{E}_0(X_j X_{j+\ell}) - \mathbb{E}(X_j X_{j+\ell})\|_{r/2}^{r/2}$$
$$\ll nM^{p-r} \sum_{j=1}^{n} j^{\beta/2-1} \sum_{\ell=0}^{j-1} (\ell+1)^{p-2-\beta/2} \|\mathbb{E}_0(X_j X_{j+\ell}) - \mathbb{E}(X_j X_{j+\ell})\|_{r/2}^{r/2} . \quad (8.18)$$

Corollary 8.1 follows from (8.12), (8.13), and Proposition 8.1 combined with the bounds (8.14), (8.15), (8.16), (8.17) and (8.18).

8.2 Proof of Proposition 2.1

The next lemma gives covariance-type inequalities in terms of the variables $b_\ell(i)$ and $b_\ell(i,j)$ of Definition 2.1. It is almost the same as Lemma 35 in [12], the only difference is that in [12], the authors used a slightly different definition of the variables $b_\ell(i,j)$. The proof of the version we give here can be done by following the proof of Lemma 35 in [12], and is therefore omitted.

Lemma 8.1. *Let Z be a $\mathcal{F}_\ell$-measurable real-valued random variable and let h and g be two BV functions (recall that $\|dh\|$ is the variation norm of the measure dh). Let $Z^{(0)} = Z - \mathbb{E}(Z)$, $h^{(0)}(Y_i) = h(Y_i) - \mathbb{E}(h(Y_i))$ and $g^{(0)}(Y_j) = g(Y_j) - \mathbb{E}(g(Y_j))$. Define the random variables $b_\ell(i)$ and $b_\ell(i,j)$ as in Definition 2.1. Then*

1. $\left|\mathbb{E}\left(Z^{(0)} h^{(0)}(Y_i)\right)\right| = |\mathrm{Cov}(Z, h(Y_i))| \leq \|dh\| \, \mathbb{E}\left(|Z| b_\ell(i)\right)$.
2. $\left|\mathbb{E}\left(Z^{(0)} h^{(0)}(Y_i) g^{(0)}(Y_j)\right)\right| \leq \|dh\| \|dg\| \, \mathbb{E}\left(|Z| b_\ell(i,j)\right)$.

We now begin the proof of Proposition 2.1. Note first that if $X_i = h(Y_i) - \mathbb{E}(h(Y_i))$ for some BV function h, then $|X_i| \leq \|dh\|$ almost surely. To prove Proposition 2.1, we apply Corollary 8.1 with $M = \|dh\|$, $r \in]\max(2p-2, 4), 2p[$ and $\beta \in]r-2, 2p-2[$. We have to bound up the second, third, fourth and fifth terms on the right hand side of (8.2). Let us do this in that order.

To control the second term, we note that, by stationarity,

$$\begin{aligned}
\mathbb{E}\,|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)| &= \mathbb{E}\,|\mathbb{E}_{-k}(X_{\ell-k})\mathbb{E}_{-k}(X_0)| \\
&= \mathbb{E}\left(\mathbb{E}_{-k}(X_{\ell-k})\mathbb{E}_{-k}(X_0)\,\mathrm{sign}\left\{\mathbb{E}_{-k}(X_{\ell-k})\mathbb{E}_{-k}(X_0)\right\}\right) \\
&= \mathbb{E}\left(X_{\ell-k}\mathbb{E}_{-k}(X_0)\,\mathrm{sign}\left\{\mathbb{E}_{-k}(X_{\ell-k})\mathbb{E}_{-k}(X_0)\right\}\right) .
\end{aligned}$$

Hence, applying Lemma 8.1, we get that, for any $\ell \geq k \geq 0$,

$$\mathbb{E}\,|\mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell)| \leq \|dh\| \mathbb{E}\left(|\mathbb{E}_{-k}(X_0)|\, b_{-k}(\ell-k)\right) \leq \|dh\| \mathbb{E}\left(|X_0| b_{-k}(\ell-k)\right) .$$

Let then

$$T_0(n) = \sum_{\ell=1}^{n} \sum_{k=0}^{\ell} (k+1)^{p-3} b_{-k}(\ell-k) ,$$

and note that $T_0(n)$ is a positive random variable which is $\mathcal{F}_0$-measurable and such that

$$\mathbb{E}(T_0(n)) \leq \sum_{\ell=1}^{n} \sum_{k=0}^{\ell} (k+1)^{p-3} \beta_{1,Y}(\ell) \leq C \sum_{\ell=1}^{n} \ell^{p-2} \beta_{1,Y}(\ell) \,,$$

for some positive constant C. Moreover

$$n\|dh\|^{p-2} \sum_{\ell=0}^{n} \sum_{k=0}^{\ell} (k+1)^{p-3} \left\| \mathbb{E}_0(X_k)\mathbb{E}_0(X_\ell) \right\|_1 \ll n\|dh\|^{p-1} \mathbb{E}\left(|X_0| \left(1 + T_0(n)\right) \right) . \quad (8.19)$$

To control the third term, we note that

$$\|\mathbb{E}_0(X_k)\|_2^2 = \mathbb{E}\left(\mathbb{E}_{-k}(X_0) X_0 \right) .$$

Hence, applying Lemma 8.1, we get that

$$\|\mathbb{E}_0(X_k)\|_2^2 \leq \|dh\| \mathbb{E}\left(|X_0| b_{-k}(0) \right) .$$

Let then

$$U_0(n) = \sum_{k=1}^{n} k^{p-2} b_{-k}(0) \,,$$

and note that $U_0(n)$ is a positive random variable which is $\mathcal{F}_0$-measurable and such that

$$\mathbb{E}(U_0(n)) \leq \sum_{k=1}^{n} k^{p-2} \beta_{1,Y}(k) \,.$$

Moreover

$$n\|dh\|^{p-2} \sum_{k=1}^{n} k^{p-2} \left\| \mathbb{E}_0(X_k) \right\|_2^2 \leq n\|dh\|^{p-1} \mathbb{E}\left(|X_0| U_0(n) \right) . \quad (8.20)$$

To control the fourth term, let first $Z = |X_0|^{r/2} |\mathbb{E}_0(X_i)|^{r/2-1} \operatorname{sign}\{\mathbb{E}_0(X_i)\}$. Then

$$\|X_0 \mathbb{E}_0(X_i)\|_{r/2}^{r/2} = \mathbb{E}\left(Z X_i \right) = \mathbb{E}\left((Z - \mathbb{E}(Z)) X_i \right) .$$

Applying Lemma 8.1, it follows that

$$\|X_0 \mathbb{E}_0(X_i)\|_{r/2}^{r/2} \leq \|dh\| \mathbb{E}(|Z| b_0(i)) \leq \|dh\|^{r-1} \mathbb{E}(|X_0| b_0(i)) \,.$$

Let then

$$V_0(n) = \sum_{i=1}^{n} i^{p-2} b_0(i) \,,$$

and note that $V_0(n)$ is a positive random variable which is $\mathcal{F}_0$-measurable and such that $\mathbb{E}(V_0(n)) \leq \sum_{i=1}^{n} i^{p-2} \beta_{2,Y}(i)$. Moreover

$$n\|dh\|^{p-r} \sum_{i=1}^{n} i^{p-2} \|X_0 \mathbb{E}_0(X_i)\|_{r/2}^{r/2} \leq n\|dh\|^{p-1} \mathbb{E}\left(|X_0| V_0(n) \right) . \quad (8.21)$$

To control the fifth term, note that, since $r/2-1\geq 1$,

$$\begin{aligned}
&\|\mathbb{E}_0(X_iX_{j+i})-\mathbb{E}(X_iX_{j+i})\|_{r/2}^{r/2}\\
&\quad=\mathbb{E}\left(|\mathbb{E}_0(X_iX_{j+i})-\mathbb{E}(X_iX_{j+i})|^{r/2-1}\,|\mathbb{E}_0(X_iX_{j+i})-\mathbb{E}(X_iX_{j+i})|\right)\\
&\qquad\leq 2^{r/2-2}\mathbb{E}\left(|X_iX_{j+i}|^{r/2-1}\,|\mathbb{E}_0(X_iX_{j+i})-\mathbb{E}(X_iX_{j+i})|\right)\\
&\qquad\qquad+2^{r/2-2}\,|\mathbb{E}(X_iX_{j+i})|^{r/2-1}\,\mathbb{E}\left(|\mathbb{E}_0(X_iX_{j+i})-\mathbb{E}(X_iX_{j+i})|\right)\,.
\end{aligned}$$

Now

$$\begin{aligned}
&\mathbb{E}\left(|X_iX_{j+i}|^{r/2-1}\,|\mathbb{E}_0(X_iX_{j+i})-\mathbb{E}(X_iX_{j+i})|\right)\\
&\quad\leq\mathbb{E}\left(|X_i|^{r-2}\,|\mathbb{E}_0(X_iX_{j+i})-\mathbb{E}(X_iX_{j+i})|\right)+\mathbb{E}\left(|X_{j+i}|^{r-2}\,|\mathbb{E}_0(X_iX_{j+i})-\mathbb{E}(X_iX_{j+i})|\right)\,.
\end{aligned}$$

Let $Z=\mathbb{E}_{-i}(|X_0|^{r-2})\,\mathrm{sign}\{\mathbb{E}_{-i}(X_0X_j)-\mathbb{E}(X_0X_j)\}$. Notice that

$$\begin{aligned}
&\mathbb{E}\left(|X_i|^{r-2}\,|\mathbb{E}_0(X_iX_{j+i})-\mathbb{E}(X_iX_{j+i})|\right)=\mathbb{E}\left(|X_0|^{r-2}\,|\mathbb{E}_{-i}(X_0X_j)-\mathbb{E}(X_0X_j)|\right)\\
&\quad=\mathbb{E}\left(\mathbb{E}_{-i}(|X_0|^{r-2})\,|\mathbb{E}_{-i}(X_0X_j)-\mathbb{E}(X_0X_j)|\right)=\mathbb{E}\left((Z-\mathbb{E}(Z))X_0X_j)\right)\,.
\end{aligned}$$

Applying Lemma 8.1, it follows that

$$\begin{aligned}
&\mathbb{E}\left(|X_i|^{r-2}\,|\mathbb{E}_0(X_iX_{j+i})-\mathbb{E}(X_iX_{j+i})|\right)\leq\|dh\|^2\mathbb{E}\left(|Z|b_{-i}(0,j)\right)\\
&\quad\leq\|dh\|^2\mathbb{E}\left(|X_0|^{r-2}b_{-i}(0,j)\right)\leq\|dh\|^{r-1}\mathbb{E}\left(|X_0|b_{-i}(0,j)\right)\,.
\end{aligned}$$

Similarly we get

$$\mathbb{E}\left(|X_{i+j}|^{r-2}\,|\mathbb{E}_0(X_iX_{j+i})-\mathbb{E}(X_iX_{j+i})|\right)\leq\|dh\|^{r-1}\mathbb{E}\left(|X_0|b_{-i-j}(-j,0)\right)\,.$$

Let then

$$W_0(n)=\sum_{i=1}^{n}i^{\beta/2-1}\sum_{j=0}^{i-1}(j+1)^{p-2-\beta/2}\left(b_{-i}(0,j)+b_{-i-j}(-j,0)\right)\,,$$

and note that $W_0(n)$ is a positive random variable which is $\mathcal{F}_0$-measurable and such that

$$\begin{aligned}
\mathbb{E}(W_0(n))&=\sum_{i=1}^{n}i^{\beta/2-1}\sum_{j=0}^{i-1}(j+1)^{p-2-\beta/2}\left(\mathbb{E}(b_{-i}(0,j))+\mathbb{E}(b_{-i-j}(-j,0))\right)\\
&\leq 2\sum_{i=1}^{n}i^{\beta/2-1}\beta_{2,Y}(i)\sum_{j=0}^{i-1}(j+1)^{p-2-\beta/2}\leq C\sum_{i=1}^{n}i^{p-2}\beta_{2,Y}(i)\,,
\end{aligned}$$

for some positive constant C. Moreover

$$\begin{aligned}
&n\|dh\|^{p-r}\sum_{i=1}^{n}i^{\beta/2-1}\sum_{j=0}^{i-1}(j+1)^{p-2-\beta/2}\|\mathbb{E}_0(X_iX_{j+i})-\mathbb{E}(X_iX_{j+i})\|_{r/2}^{r/2}\\
&\quad\ll n\|dh\|^{p-1}\mathbb{E}\left(|X_0|W_0(n)\right)+n\|dh\|^{p-1}\mathbb{E}\left(|X_0|\right)\mathbb{E}\left(W_0(n)\right)\,. \qquad (8.22)
\end{aligned}$$

To conclude the proof, let $B_0(n,p)=T_0(n)+U_0(n)+V_0(n)+W_0(n)$, and note that $W_0(n)$ is a positive random variable which is $\mathcal{F}_0$-measurable and such that $\mathbb{E}\left(B_0(n,p)\right)\leq$

$\kappa \sum_{k=1}^{n} k^{p-2}\beta_{2,Y}(k)$, for some positive constant κ. From (8.2), (8.19), (8.20), (8.21) and (8.22), and since $X_0 = h(Y_0) - \mathbb{E}(h(Y_0))$, we infer that

$$\mathbb{E}\left(\sup_{1\leq k\leq n}|S_k|^p\right) \ll n^{p/2}\left(\sum_{i=0}^{n-1}|\mathrm{Cov}(X_0,X_i)|\right)^{p/2} + n\|dh\|^{p-1}\mathbb{E}\left(|h(Y_0)|B_0(n,p)\right)$$
$$+ n\|dh\|^{p-1}\mathbb{E}\left(|h(Y_0)|\right)\sum_{k=0}^{n}(k+1)^{p-2}\beta_{2,Y}(k)\,.$$

Let then $A_0(n,p) = \kappa^{-1}B_0(n,p)$, in such a way that $\mathbb{E}\left(A_0(n,p)\right) \leq \sum_{k=1}^{n} k^{p-2}\beta_{2,Y}(k)$. The random variable $A_0(n,p)$ satisfies the statement of Proposition 2.1, and the proof is complete.

References

[1] H. C. P. Berbee, Random walks with stationary increments and renewal theory, Cent. Math. Tracts, Amsterdam, 1979.

[2] R. C. Bradley (1986), Basic properties of strong mixing conditions, *Dependence in probability and statistics. A survey of recent results. Oberwolfach, 1985.* E. Eberlein and M. S. Taquu editors, Birkäuser, 165-192.

[3] J. H. Bramble and S. R. Hilbert (1970), Estimation of linear functionals on Sobolev spaces with application to Fourier transforms and spline interpolation, *SIAM J. Numer. Anal.* **7** 112-124.

[4] J. Bretagnolle and C. Huber (1979), Estimation des densités: risque minimax, *Z. Wahrsch. Verw. Gebiete* **47** 119-137.

[5] J. Dedecker, H. Dehling and M. S. Taqqu (2015), weak convergence of the empirical process of intermittent maps in $\mathbb{L}^2$ under long-range dependence, *Stoch. Dyn.* **15** 29 pp.

[6] J. Dedecker, S. Gouëzel and F. Merlevède (2012), The almost sure invariance principle for unbounded functions of expanding maps, *ALEA Lat. Am. J. Probab. Math. Stat.* **9** 141-163.

[7] J. Dedecker and C. Prieur (2005), New dependence coefficients. Examples and applications to statistics, *Probab. Theory Related Fields* **132** 203-236.

[8] J. Dedecker and C. Prieur (2007), An empirical central limit theorem for dependent sequences, *Stochastic Process. Appl.* **117** 121-142.

[9] J. Dedecker and E. Rio (2000), On the functional central limit theorem for stationary processes, *Ann. Inst. Henri Poincaré Probab. Stat.* **36** 1-34.

[10] R. A. DeVore and G. G. Lorentz, Constructive approximation, Springer-Verlag, Berlin Heidelberg New-York, 1993.

[11] C. Liverani, B. Saussol and S. Vaienti (1999), A probabilistic approach to intermittency, *Ergodic Theory Dynam. Systems* **19** 671-685.

[12] F. Merlevède and M. Peligrad (2013), Rosenthal-type inequalities for the maximum of partial sums of stationary processes and examples, *Ann. Probab.* **41** 914-960.

[13] E. Rio (2000), Théorie asymptotique des processus aléatoires faiblement dépendants, *Mathématiques et Applications* **31**, Springer-Verlag, Berlin.

[14] M. Rosenblatt (1956), A central limit theorem and a strong mixing condition, *Proc. Nat. Acad. Sci. U. S. A.* **42** 43-47.

[15] M. Thaler (1980), Estimates of the invariant densities of endomorphisms with indifferent fixed points. *Israel J. Math.* **37**, 303–314.

[16] G. Viennet (1997), Inequalities for absolutely regular sequences: application to density estimation, *Probab. Theory Related Fields* **107** 467-492.