Stochastics and Dynamics
Vol. 18, No. 2 (2018) 1850017 (38 pages)

DOI: 10.1142/S021949371850017X

Large and moderate deviations for bounded functions of slowly mixing Markov chains

Jérôme Dedecker
Laboratoire MAP5 (UMR CNRS 8145),
Université Paris Descartes,
Sorbonne Paris Cité
45 rue des saints-pères, 75270 Paris cedes 06, France
jerome.dedecker@parisdescartes.fr

Sébastien Gouëzel
Laboratoire Jean Leray,
CNRS UMR 6629, Université de Nantes,
2 rue de la Houssinière, 44322 Nantes, France
sebastien.gouezel@univ-nantes.fr

Florence Merlevède
LAMA (UMR CNRS 8050), Université Paris-Est Marne-la-Vallée
5 boulevard Descartes, 77420 Champs-sur-Marne, France
florence.merlevede@u-pem.fr

Received 14 December 2016
Revised 11 February 2017
Accepted 19 February 2017
Published 21 April 2017

We consider Markov chains which are polynomially mixing, in a weak sense expressed in terms of the space of functions on which the mixing speed is controlled. In this context, we prove polynomial large and moderate deviations inequalities. These inequalities can be applied in various natural situations coming from probability theory or dynamical systems. Finally, we discuss examples from these various settings showing that our inequalities are sharp.

Keywords: Deviation inequalities; concentration inequalities; Markov chains.

AMS Subject Classification: 60E15, 60J05, 37A25

1. Introduction

For stationary α-mixing sequences in the sense of Rosenblatt (see [25]) a Fuk–Nagaev type inequality has been proved by Rio (see Theorem 6.2 in [24]). This deviation inequality is very powerful and gives for instance sharp upper bounds for the deviation of partial sums when the strong mixing coefficients decrease at a polynomial rate. In particular for a bounded observable f of a strictly stationary Markov chain $(Y_i)_{i\in\mathbb{Z}}$ with strong mixing coefficients of order $O(n^{1-p})$ for $p \geq 2$, Rio's inequality gives: for any $x > 0$ and any $r \geq 1$,

$$\mathbb{P}\left(\max_{1\leq k\leq n}\left|\sum_{i=1}^{k}(f(Y_i)-\pi(f))\right| \geq x\right) \leq C\left\{\frac{n}{x^p}+\frac{n^{r/2}}{x^r}+\frac{(n\log n)^{r/2}}{x^r}\mathbf{1}_{p=2}\right\}, \tag{1.1}$$

where C depends on $\|f\|_\infty$, p and r.

However, many stationary processes are not strong mixing in the sense of Rosenblatt. This is the case, for instance, of the iterates of an ergodic measure-preserving transformation. In the recent paper [10], the authors proved, using a weaker version of the α-mixing coefficients, that it is still possible to get the same upper bound as (1.1) but for bounded variation observables and with the restriction $r \in (2(p-1), 2p)$. This last restriction does not affect the asymptotic behavior of the probability of large deviations (that is when $x = ny$ in (1.1) with y fixed) but gives a restriction for the moderate deviation behavior (that is when $x = n^\alpha y$ in (1.1) with $\alpha \in (1/2, 1)$ and y fixed).

The aim of this paper is to obtain upper bounds of the type (1.1) for stationary Markov chains, when the mixing property of the chain is defined through a subclass of bounded observables $\mathcal{B}$, but without restriction on r. In that case, the deviation inequality (see our Theorem 1.1) will be valid for any observable $f \in \mathcal{B}$. Maybe the same kind of inequalities can be proved in a more general (but non α-mixing) context than the Markovian setting, but the proof we give here uses the Markovian property in a crucial way.

Let us now present more precisely the assumptions on the Markov chains and the main results of the paper.

Let $(Y_i)_{i\in\mathbb{Z}}$ be a homogeneous Markov chain on a state space $\mathcal{X}$, with transition operator K, admitting a stationary probability measure π. Let $\|\cdot\|$ be a norm on a vector space $\mathcal{B}$ of functions from $\mathcal{X}$ to $\mathbb{R}$. We always require that the constant function equal to 1 belongs to $\mathcal{B}$. This norm will be used to express mixing conditions on the Markov chain.

We will need this norm to behave well with respect to products, and to control the sup norm, as expressed in the next definition.

Definition 1.1. We say that $\|\cdot\|$ is a Banach algebra norm on bounded functions if, for all f and g in $\mathcal{B}$, one has $\|f\|_\infty \leq \|f\|$ and $\|fg\| \leq \|f\|\|g\|$.

Remark 1.1. If a norm $\|\cdot\|$ satisfies $\|f\|_\infty \leq C\|f\|$ and $\|fg\| \leq C\|f\|\|g\|$ for some constant C, then it is equivalent to a Banach algebra norm on bounded functions, namely $\|f\|' = C\|f\|$.

The main mixing condition we require is that the iterates of functions in $\mathcal{B}$ under the Markov chain converge polynomially to their average. This is expressed in terms of the following two conditions.

Definition 1.2. Let $p > 1$. We say that the condition $\mathbf{H}_1(p)$ is satisfied if there exists a positive constant C_1 such that, for any function $f \in \mathcal{B}$ and any $n \geq 1$,

$$\pi(|K^n(f) - \pi(f)|) \leq \frac{C_1\|f\|}{n^{p-1}}. \qquad \mathbf{H}_1(p)$$

We say that the condition $\mathbf{H}_2$ is satisfied if the space $\mathcal{B}$ is invariant under the iterates K^n of K, uniformly in n, i.e. there exists a positive constant C_2 such that, for any function f in $\mathcal{B}$ and any $n \geq 1$,

$$\|K^n(f)\| \leq C_2\|f\|. \qquad \mathbf{H}_2$$

When both conditions are satisfied, we say that the chain converges polynomially to equilibrium for the norm $\|\cdot\|$ with exponent p, and we denote this condition by $\mathbf{H}(p)$.

Heuristically, partial sums of bounded functions of such a polynomially mixing chain behave like sums of independent random variables with a weak moment of order p (see the paper by Nagaev [21] for precise results in that case). Indeed, if one considers a Harris recurrent Markov chain (see the book [22] for the definition and background) for which the excursion time away from an atom has a weak moment of order p, then the successive excursions are independent and have a weak moment of order p, and the mixing rate behaves like in the definition above. Hence, one expects that one should prove, under $\mathbf{H}(p)$, results that are similar to results for sums of i.i.d. random variables with a weak moment of order p.

In particular, let us consider the question of moderate deviations bounds

$$\mathbb{P}\left(\max_{1\leq k\leq n}\left|\sum_{i=1}^{k}(f(Y_i) - \pi(f))\right| \geq xn^{\alpha}\right),$$

where f belongs to $\mathcal{B}$, $x > 0$. In analogy with the i.i.d. case, one expects that, if $p \geq 2$, then for any $\alpha \in (1/2, 1]$ there should exist positive constants C depending only on p and on $\|f\|$, and $v(x)$ depending only on x, such that

$$\limsup_{n\to\infty} n^{\alpha p-1}\mathbb{P}\left(\max_{1\leq k\leq n}\left|\sum_{i=1}^{k}(f(Y_i) - \pi(f))\right| \geq xn^{\alpha}\right) \leq Cv(x). \qquad (1.2)$$

Our main result ensures that this estimate indeed holds, with bounds that are very similar to the case of the sum of i.i.d. random variables. We also deal with the case $p < 2$, obtaining similar estimates. The proof of Theorem 1.1 below is given in Sec. 2.

Theorem 1.1. *Let $(Y_i)_{i\in\mathbb{Z}}$ be a stationary Markov chain with state space $\mathcal{X}$, transition operator K and stationary measure π. Assume that there exists $p > 1$ such that $\mathbf{H}_1(p)$ holds, for a Banach algebra norm on bounded functions.*

(1) *If $p > 2$ and we assume in addition that $\mathbf{H}_2$ is satisfied then, for any $f \in \mathcal{B}$ and any $x > 0$,*

$$\mathbb{P}\left(\max_{1\le k\le n}\left|\sum_{i=1}^{k}(f(Y_i)-\pi(f))\right|\ge x\right)\le \kappa n x^{-p}+\kappa\exp(-\kappa^{-1}x^2/n), \tag{1.3}$$

where κ is a positive constant depending only on p, $\|f\|$, C_1 and C_2.

(2) *If $p = 2$ and we assume in addition that $\mathbf{H}_2$ is satisfied, then for any $f \in \mathcal{B}$, any $x > 0$ and any $r \in (2, 4)$,*

$$\mathbb{P}\left(\max_{1\le k\le n}\left|\sum_{i=1}^{k}(f(Y_i)-\pi(f))\right|\ge x\right)\le \kappa n x^{-2}+\kappa(n\log n)^{r/2}x^{-r}, \tag{1.4}$$

where κ is a positive constant depending only on $\|f\|$, C_1, C_2 and r.

(3) *If $1 < p < 2$, then for any $f \in \mathcal{B}$ and any $x > 0$,*

$$\mathbb{P}\left(\max_{1\le k\le n}\left|\sum_{i=1}^{k}(f(Y_i)-\pi(f))\right|\ge x\right)\le \kappa n x^{-p}, \tag{1.5}$$

where κ is a positive constant depending only on p, $\|f\|$ and C_1.

As a consequence of this theorem, we obtain that, if $p > 1$, then (1.2) holds with $v(x) = x^{-p}$ for any $\alpha > 1/2$ such that $1/p \le \alpha \le 1$ provided that $\mathbf{H}(p)$ holds.

Remark 1.2. In (1.3), the exponential term $\kappa\exp(-\kappa^{-1}x^2/n)$ is negligible in the regime $x > n^{\alpha}$, for any $\alpha > 1/2$. Hence, the dominating term is $\kappa n x^{-p}$, as expected. However, when x is of the order of $n^{1/2}$, then $\kappa n x^{-p}$ tends to 0, while the probability on the left of (1.3) typically does not, thanks to the central limit theorem. Thus, there has to be a remainder term, given here in exponential form $\kappa\exp(-\kappa^{-1}x^2/n)$. For any $r > 0$, this is for instance bounded by $C_{\kappa,r}n^r/x^{2r}$.

Remark 1.3. In (1.4), the ratio $x^2/(n\log n)$ appearing in the error term is the right one: in this setting there is sometimes a central limit theorem with anomalous scaling $\sqrt{n\log n}$ (see for instance [15]), meaning that the probability on the left of (1.4) does not tend to 0 when x is of the order of $\sqrt{n\log n}$. While (1.3) is completely satisfactory, we expect that the error term in (1.4) can be improved, from $(n\log n)^{r/2}x^{-r}$ with $r \in (2, 4)$ to $(n\log n)^{r/2}x^{-r}$ for any $r > 2$, or even to $\exp(-\kappa^{-1}x^2/(n\log n))$. However, we are not able to prove such a result.

Remark 1.4. The spectral gap property for the operator K on the space $\mathcal{B}$ reads as follows: there exist a positive constant C and a $\rho \in (0, 1)$ such that for any $f \in \mathcal{B}$ and any $n \ge 1$

$$\|K^n f - \pi(f)\| \le C\rho^n\|f\|, \tag{1.6}$$

which obviously implies $\mathbf{H}_2$ and $\mathbf{H}_1(p)$ for any $p > 1$. However, Theorem 1.1 is useless in this case since a sub-Gaussian bound holds. For instance, using Proposition 2 in [23], we get that for any $n \geq 1$, any $f \in \mathcal{B}$ and any $x > 0$,

$$\mathbb{P}\left(\max_{1\leq k\leq n}\left|\sum_{i=1}^{k}(f(Y_i)-\pi(f))\right| \geq x\right) \leq 4\sqrt{e}\exp(-\kappa^{-1}x^2/n), \tag{1.7}$$

where κ is a positive constant depending only on ρ, C and $\|f\|$. Note that the geometrical mixing in not important here: what really matters is that one can control the sup-norm in (1.6). More precisely, the deviation bound (1.7) still holds provided that

$$\sum_{k=1}^{\infty}\frac{\|K^n f-\pi(f)\|_\infty}{\sqrt{k}} < \infty.$$

Alternatively, for Markov chains with a spectral gap in $\mathbb{L}^2$ (ρ-mixing Markov chains), a sub-Gaussian inequality holds for partial sums of bounded observables (see [17], p. 861). This kind of upper bound is also true for geometrically ergodic Markov chains (see [3]).

Let us discuss the relevance of the assumption $\mathbf{H}(p)$ in different contexts. Some possible Banach algebra norms on bounded functions that appear in natural examples of Markov chains are the following:

(1) $\|f\| = \|f\|_\infty$.
(2) $\mathcal{X} = \mathbb{R}$ and $\|f\|$ is the total variation norm of the bounded variation function f, i.e. the sum of $\|f\|_\infty$ and of the total variation of the measure $\mathrm{d}f$, i.e. $\|f\| = \|f\|_\infty + |\mathrm{d}f|$.
(3) If $(\mathcal{X}, d)$ is a metric space, then one can consider the Lipschitz norm

$$\|f\| = \|f\|_\infty + \|f\|_{\mathrm{Lip}} \quad \text{where } \|f\|_{\mathrm{Lip}} := \sup_{y\neq z\in\mathcal{X}}\frac{|f(y)-f(z)|}{d(y,z)},$$

or Hölder norms.
(4) $\mathcal{X} = \mathbb{R}$ and $\|f\| = \|f\|_\infty + \|f'\|_{L^r(\lambda)}$ for $r \geq 1$, when f is absolutely continuous and f' is its almost sure derivative. One can also consider more general Sobolev spaces, in dimension 1 or higher.

Here is a more detailed discussion of some corresponding examples:

(1) When $\mathbf{H}_1(p)$ is satisfied with $\|f\| = \|f\|_\infty$, then the chain is said to be strong mixing in the sense of Rosenblatt with polynomial rate of convergence n^{p-1}, and we write in this case

$$\alpha_n = \sup_{k\geq n}\sup_{\|f\|_\infty\leq 1}\pi(|K^k(f)-\pi(f)|) \leq \frac{C_1}{n^{p-1}}.$$

Note that for this norm, $\mathbf{H}_2$ is trivially satisfied. In this situation, for $p \geq 2$, one can apply the Fuk–Nagaev type inequality in [24], Theorem 6.1, (with $q = cx$

for suitably small c) to deduce (1.1). Hence, (1.2) follows (note, however, that the error term we get in (1.3) is better than the error term from (1.1)).

(2) When $\mathbf{H}(p)$ is satisfied with the total variation norm (i.e. $\mathcal{B}$ is the set of functions of bounded variation), one does not have at our disposal a Fuk–Nagaev type inequality as in the strong mixing case. If $p > 2$, an application of the deviation inequality given in Proposition 5.1 of [10] gives that for any $x > 0$ and any $r \in (2(p-1), 2p)$,

$$\mathbb{P}\left(\max_{1\le k\le n}\left|\sum_{i=1}^{k}(f(Y_i)-\pi(f))\right| \ge x\right) \le C\left\{\frac{n}{x^p} + \frac{n^{r/2}}{x^r} + \frac{(n\log n)^{r/2}}{x^r}\mathbf{1}_{p=2}\right\},$$

where C is a positive constant depending on p, $\|f\|$, C_1 and C_2 but not on n or on x. So, provided that $p < 1/(1-\alpha)$, one can take $2p > r > \frac{2(\alpha p-1)}{2\alpha-1}$ and it follows that (1.2) is satisfied with $v(x) = x^{-p}$. Our Theorem 1.1 above shows that this restriction of α is not necessary, by removing the restriction $r \in (2(p-1), 2p)$ for $p > 2$. Our proof is in the spirit of that given in [10], but we can get a better bound by taking advantage of the Markovian setting.

Let us describe briefly a class of examples satisfying $\mathbf{H}(p)$ with the total variation norm. Let T_γ be a GPM map, as defined in [7], that is an expanding map of $[0,1]$ with a neutral fixed point at 0; the behavior of the map around 0 is described by the parameter $\gamma \in (0,1)$. It is proved in [7] that the Markov chain associated with T_γ satisfies $\mathbf{H}(p)$ for the total variation norm and $p = 1/\gamma$. Hence, Theorem 1.1 applies to BV-functions f of such a chain, and then to the maximum of partial sums $\sum_{i=1}^{k}(f \circ T_\gamma^i - \pi(f))$ on the probability space $([0,1], \pi)$ (see for instance inequality (4.3) in [10]).

(3) Several examples satisfy the assumption $\mathbf{H}(p)$ when $\|\cdot\|$ is the Lipschitz norm or the Hölder norm.

For dynamical systems, there is a combinatorial model, called Young tower, that can be used to model wide classes of systems and for which the assumption $\mathbf{H}(p)$ is directly related to return time estimates to the basis of the tower (this is explicitly written, for instance, in (4.3) of [9]). We refer the interested readers for instance to the introduction of [16], where motivations, examples and definitions are given. Our theorem applies to such examples, and improves the previous upper bounds of the literature such as in Melbourne [18], who obtained, when $\alpha \in (1/2, 1)$ and $p \ge 2$, a rate of order $(\ln n)^{1-p} n^{(p-1)(2\alpha-1)}$ instead of $n^{\alpha p-1}$ in (1.2). Using specific properties of such systems established in [16], we are also able to extend Theorem 1.1 to more general functionals than additive functionals, see Theorem 4.1 in Sec. 4.

Here is another class of examples. Assume that the stationary Markov chain $(X_i)_{i\ge 0}$ with state space $(\mathcal{X}, d)$ may be written as $X_i = F(X_{i-1}, \varepsilon_i)$, where $(\varepsilon_i)_{i\ge 1}$ is i.i.d. and independent of X_0. Let $X_{n,x}$ be the chain starting from $X_0 = x$.

Let Λ_d be the set of Lipschitz functions such that $|f(x) - f(y)| \leq d(x, y)$. Assume that one can prove that

$$\pi\left(\sup_{f\in\Lambda_d} |K^n(f) - \pi(f)|\right) \leq C_1\rho^{n^a}, \tag{1.8}$$

for some $\rho \in (0, 1)$, $C_1 > 0$ and $a > 0$, and that

$$\sup_{n\geq 0} \mathbb{E}(d(X_{n,x}, X_{n,y})) \leq C_2 d(x, y), \tag{1.9}$$

for some $C_2 \geq 1$.

Let c be a concave, non-decreasing and sub-additive function from $\mathbb{R}^+$ to $\mathbb{R}^+$, such that $c(t) \leq C_3|\log(t)|^{-\gamma}$ in a neighborhood of 0, for some $C_3 > 0, \gamma > 0$. Then, $d_c(x, y) = c(d(x, y))$ satisfies the triangle inequality, and one can prove that

$$\pi\left(\sup_{f\in\Lambda_{d_c}} |K^n(f) - \pi(f)|\right) \leq \frac{C_4}{n^{a\gamma}},$$

and that, for any Lipschitz function f in Λ_{d_c},

$$|K^n(f)(x) - K^n(f)(y)| \leq C_2 c(d(x, y)).$$

It follows that $\mathbf{H}_1(p)$ and $\mathbf{H}_2$ are satisfied for $p - 1 = a\gamma$ if we consider the space $\mathcal{B}_c$ with Lipschitz norm $\|f\| = \|f\|_\infty + \|f\|_{\mathrm{Lip}(d_c)}$. Hence, Theorem 1.1 applies to a very large set of continuous bounded observables of the chain.

For instance the inequalities (1.8) and (1.9) will be satisfied with $C_2 = 1$ and $a = 1$ if $\mathbb{E}(d(F(x, \varepsilon_1), F(y, \varepsilon_1))) \leq \rho d(x, y)$ and $\mathbb{E}(d(X_0, x_0)) < \infty$ for some $x_0 \in \mathcal{X}$. Other examples for which $a < 1$ are given in Sec. 4.0.2 of [20].

In addition, concerning the exponent of n, the bound (1.2) is optimal as we shall show in Sec. 3. More precisely, we shall give there three different examples for which the deviation probabilities of Theorem 1.1 are lower bounded by $c\,nx^{-p}$ for some $c > 0$ and x in an appropriate bandwidth. These three examples are: a discrete Markov chain on $\mathbb{N}$ for which $\mathbf{H}(p)$ is satisfied for the sup norm, a class of Young towers with polynomial tails of the return times for which $\mathbf{H}(p)$ is satisfied for a natural Lipschitz norm, and a Harris recurrent Markov chain with state space $[0, 1]$ for which $\mathbf{H}(p)$ is satisfied for both the sup norm and the total variation norm. For each example, the accurate lower bound is given in Propositions 3.1, 3.2 and 3.3 respectively.

To be complete, we give in Sec. 4 an extension of Theorem 1.1 to more general functionals in the specific setting of Young towers (see our Theorem 4.1). Let us emphasize that the proofs of Theorems 1.1 and 4.1 are very different.

2. Upper Bounds for Moderate Deviations

In this section, we prove Theorem 1.1. Cases (3) and (2) follow more or less readily from existing inequalities in the literature, while Case (1) is new.

2.1. *Proof of Item (3) in Theorem 1.1*

Item (3) follows directly from an application of Proposition 4 in [8]. Indeed, let $M = \|f\|_\infty$ and

$$\gamma(k) = \|\mathbb{E}(f(Y_k) \mid Y_0) - \pi(f)\|_1, \quad \text{for } k \geq 0.$$

Proposition 4 in [8] together with stationarity implies that for any integer q in $[1, n]$, and any $x \geq Mq$,

$$\mathbb{P}\left(\max_{1\leq k\leq n}\left|\sum_{i=1}^{k}(f(Y_i)-\pi(f))\right| \geq 4x\right) \leq \frac{4nM}{x^2}\sum_{i=0}^{q-1}\gamma(i) + \frac{2n}{xq}\sum_{i=q+1}^{2q}\gamma(i).$$

Note that if $x \geq nM/2$ the bound is trivial since the probability is equal to zero. It is also trivial if $x \leq M\sqrt{2n}$. Therefore we can always assume that $M \leq x \leq nM$ and select $q = [x/M]$. Combined with the fact that, by $\mathbf{H}_1(p)$,

$$\gamma(k) \leq C_1\|f\|(k+1)^{1-p},$$

this gives

$$\mathbb{P}\left(\max_{1\leq k\leq n}\left|\sum_{i=1}^{k}(f(Y_i)-\pi(f))\right| \geq 4x\right) \leq \left(\frac{4nM^{p-1}C_1\|f\|}{2-p} + 2^p M^{p-1}C_1\|f\|\right) nx^{-p}.$$

This ends the proof of Item (3). □

2.2. *A deviation inequality*

For $r > 2$, the Rosenthal inequality for sums of centered i.i.d. random variables Z_i is the following

$$\mathbb{E}\left(\left|\sum_{i=1}^{N} Z_i\right|^r\right) \leq CN\mathbb{E}(|Z_1|^r) + CN^{r/2}\mathbb{E}(Z_1^2)^{r/2}, \tag{2.1}$$

where C is a positive constant only depending on r. What makes this inequality extremely useful is that the dominating coefficient $N^{r/2}$ is multiplied by an L^2-norm, which is usually mild to control, while the larger L^r-norm only has a coefficient N.

We will use repeatedly a Rosenthal-like inequality for weakly dependent sequences, due to Merlevède and Peligrad [19], in the following form which is well-suited for the applications to moderate deviations we have in mind. Note that, in the following statement, all conditional expectations are of the form $\mathbb{E}(f(Z_i) \mid \mathcal{G}_0)$ for some $i \geq 2$: this means that suitable mixing conditions can be used to control such terms. The other two terms are of Rosenthal-type as in the i.i.d. case, and can thus be controlled using minimal knowledge on Z_1.

Theorem 2.1. *Let Z_i be a strictly stationary sequence of random variables, adapted to a filtration $\mathcal{G}_i$. Write $S_i = \sum_1^i Z_k$. Consider a real number $r > 2$. Then, for all*

N and all x,

$$\mathbb{P}\left(\max_{i\le N}|S_i|\ge x\right)\le C\left\{\frac{N}{x}\|\mathbb{E}(Z_2\mid\mathcal{G}_0)\|_1+\frac{N}{x^r}\mathbb{E}(|Z_1|^r)+\frac{N^{r/2}}{x^r}(\mathbb{E}(Z_1^2))^{r/2}\right.$$

$$\left.+\frac{N}{x^r}\left[\sum_{k=1}^{N}\frac{1}{k^{1+2\delta/r}}\left(\sum_{i=2}^{k}\|\mathbb{E}(Z_i^2\mid\mathcal{G}_0)-\mathbb{E}(Z_i^2)\|_{r/2}\right)^{\delta}\right]^{r/(2\delta)}\right\},$$

where $\delta=\min(1,1/(r-2))\in(0,1]$ and C is a positive constant only depending on r.

Proof. Let $M_i=Z_i-\mathbb{E}(Z_i\mid\mathcal{G}_{i-2})$. Then

$$\max_{i\le N}|S_i|\le\max_{2\le 2j\le N}\left|\sum_{i=1}^{j}M_{2i}\right|+\max_{1\le 2j-1\le N}\left|\sum_{i=1}^{j}M_{2i-1}\right|+\sum_{i=1}^{N}|\mathbb{E}(Z_i\mid\mathcal{G}_{i-2})|. \qquad (2.2)$$

If the maximum of the partial sums S_i is at least x, one of these three terms is at least $x/3$.

First, by Markov inequality and stationarity,

$$\mathbb{P}\left(\sum_{i=1}^{N}|\mathbb{E}(Z_i\mid\mathcal{G}_{i-2})|\ge x/3\right)\le\frac{3}{x}N\|\mathbb{E}(Z_2\mid\mathcal{G}_0)\|_1,$$

giving a term compatible with the statement of the theorem. The other two terms in (2.2) are controlled similarly, let us consider for instance the even indices.

We use first Markov inequality with the exponent r, and then the Rosenthal-like inequality in [19, Theorem 6], giving

$$\mathbb{P}\left(\max_{2\le 2j\le N}\left|\sum_{i=1}^{j}M_{2i}\right|\ge\frac{x}{3}\right)$$

$$\le C\frac{N}{x^r}\left(\|M_1\|_r^r+\left[\sum_{k=1}^{N/2}\frac{1}{k^{1+2\delta/r}}\left\|\mathbb{E}\left(\left(\sum_{i=1}^{k}M_{2i}\right)^2\middle|\mathcal{G}_0\right)\right\|_{r/2}^{\delta}\right]^{r/(2\delta)}\right).$$

Since $\|M_1\|_r^r\le 2^r\mathbb{E}(|Z_1|^r)$, the resulting term is compatible with the statement of the theorem. As M_{2i} is a sequence of martingale differences with respect to $\mathcal{G}_{2i}$, we have

$$\mathbb{E}\left(\left(\sum_{i=1}^{k}M_{2i}\right)^2\middle|\mathcal{G}_0\right)=\sum_{i=1}^{k}\mathbb{E}(M_{2i}^2\mid\mathcal{G}_0)\le\sum_{i=1}^{k}\mathbb{E}(Z_{2i}^2\mid\mathcal{G}_0).$$

Therefore, by stationarity,

$$\left\|\mathbb{E}\left(\left(\sum_{i=1}^{k}M_{2i}\right)^2\middle|\mathcal{G}_0\right)\right\|_{r/2}\le\sum_{i=1}^{k}\|\mathbb{E}(Z_{2i}^2\mid\mathcal{G}_0)-\mathbb{E}(Z_{2i}^2)\|_{r/2}+k\mathbb{E}(Z_1^2).$$

We substitute this estimate into the previous equation. The first term gives a contribution as in the statement of the theorem. On the other hand, the contribution of the second term $k\mathbb{E}(Z_1^2)$ is

$$C\frac{N}{x^r}\mathbb{E}(Z_1^2)^{\delta\cdot r/(2\delta)}\cdot\left[\sum_{k=1}^{N/2}\frac{1}{k^{1+2\delta/r}}k^{\delta}\right]^{r/(2\delta)}$$

$$\leq C\frac{N}{x^r}(\mathbb{E}(Z_1^2))^{r/2}C'N^{r/2-1}=C''\frac{N^{r/2}}{x^r}(\mathbb{E}(Z_1^2))^{r/2},$$

again one of the terms in the statement of the theorem. □

Remark 2.1. Using different Rosenthal inequalities, one can obtain slightly different statements. For instance, using the classical Rosenthal inequality of Burkholder for martingales, one obtains a statement analogous to Theorem 2.1, where the last term in the upper bound is replaced by

$$\frac{N^{r/2}}{x^r}\|\mathbb{E}(Z_2^2\mid\mathcal{G}_0)-\mathbb{E}(Z_2^2)\|_{r/2}^{r/2}. \tag{2.3}$$

This statement uses the decorrelation less strongly than Theorem 2.1: For large i, the quantity $\|\mathbb{E}(Z_i^2\mid\mathcal{G}_0)-\mathbb{E}(Z_i^2)\|_{r/2}$ is likely much smaller than $\|\mathbb{E}(Z_2^2\mid\mathcal{G}_0)-\mathbb{E}(Z_2^2)\|_{r/2}$. Indeed, it turns out that, for the application below, Theorem 2.1 will succeed while an estimate using (2.3) fails (compare for instance (2.10) below to what would be obtained using (2.3)).

2.3. *Proof of Items* (1) *and* (2) *in Theorem* 1.1

Item (2) in Theorem 1.1 follows from an application of Proposition 5.1 in [10] (while the result there applies directly to bounded variation functions, the proof works in the full generality of Theorem 1.1). However, as we shall see it also follows from our proof as a special case.

The strategy of the proof is to apply the Rosenthal bounds of Theorem 2.1 to different parts of $\max_{1\leq k\leq n}|\sum_{i=1}^k(f(Y_i)-\pi(f))|$. To illustrate why this strategy might work, let us recall a way to prove moderate deviation bounds for sums of centered i.i.d. random variables Z_i in L^p. Consider an integer n and a real number $x>0$. Let $X_i=Z_i1_{|Z_i|>n^{1/p}}-\mathbb{E}(Z_i1_{|Z_i|>n^{1/p}})$ and $X_i'=Z_i-X_i$. Then Rosenthal inequality (2.1) (for sums of independent random variables) with the exponent p applied to X_i gives $\mathbb{E}(|\sum X_i|^p)\leq Cn$, while Rosenthal inequality with some exponent $r>p$ applied to X_i' gives $\mathbb{E}(|\sum X_i'|^r)\leq Cn^{r/2}$. Combining these two inequalities, we deduce the moderate deviations bound

$$\mathbb{P}\left(\left|\sum_{i=1}^n Z_i\right|\geq x\right)\leq C\left\{\frac{n}{x^p}+\frac{n^{r/2}}{x^r}\right\}.$$

We will follow the same strategy in our context: split the sum to be estimated in two different parts, and apply a Rosenthal inequality (in our case, Theorem 2.1) to each part, with suitable exponents. Instead of truncating, the splitting will be done by constructing blocks, and separating a conditional average (which is small in L^1, but large in L^p, as X_i above) from the dominating term.

Here is a high level version of the (rather technical) proof to follow. First, we write $\sum_j f(Y_j) - \pi(f)$ as a sum $\sum B_i$, where B_i is a sum of f along a block of length $n^{1/p}$. Then, we write the term B_i as $(B_i - \mathbb{E}(B_i \mid \mathcal{G}^B_{i-2})) + \mathbb{E}(B_i \mid \mathcal{G}^B_{i-2})$, where $\mathcal{G}^B_i$ is the natural filtration along which B_i is measurable. Then $\mathbb{E}(B_i \mid \mathcal{G}^B_{i-2})$ is small in L^1, but possibly large in L^∞. We control the probability of moderate deviations of $\sum \mathbb{E}(B_i \mid \mathcal{G}^B_{i-2})$ by grouping these variables into blocks of size $\asymp x$, then applying Theorem 2.1: all the terms in the upper bound of this theorem can be controlled, in a straightforward albeit tedious way, by using the assumption $\mathbf{H}_1(p)$. Then, to control the probability of moderate deviations of $\sum(B_i - \mathbb{E}(B_i \mid \mathcal{G}^B_{i-2}))$, we consider separately the sums along even and odd indices, use that each such sum is a martingale, and apply an exponential inequality for martingales (here, Freedman inequality). It follows that, to control the probability of moderate deviations, it suffices to control the deviations of the conditional quadratic averages. To handle these, we group them again into blocks of size $\asymp x$ and apply again Theorem 2.1. All the terms in the upper bound of this theorem can also be controlled directly from $\mathbf{H}_1(p)$.

Below are the details of the proof.

Proof of Items (1) and (2) in Theorem 1.1. We will use the following notations throughout the proof. Let $f^{(0)} = f - \pi(f)$ and $M = \|f\|_\infty$ and $\mathcal{F}_k = \sigma(Y_i, i \le k)$ and $\mathbb{E}_k(\cdot) = \mathbb{E}(\cdot \mid \mathcal{F}_k)$ and $\mathbb{E}^{(0)}_k(\cdot) = \mathbb{E}(\cdot \mid \mathcal{F}_k) - \mathbb{E}(\cdot)$.

Fix $x > 0$ and an integer n. It suffices to estimate

$$\mathbb{P}\left(\max_{1\le k\le n}\left|\sum_{i=1}^{k}(f(Y_i) - \pi(f))\right| \ge 8x\right). \tag{2.4}$$

Indeed, if one proves the theorem for this quantity, then the original result follows by letting $x' = x/8$, as polynomial bounds involving x' or x are equivalent. From this point on, we concentrate on bounding (2.4).

We first notice that since $\|\max_{1\le k\le n}|\sum_{i=1}^{k} f(Y_i) - \pi(f)|\|_\infty \le 2\|f\|_\infty n$, we can assume that

$$x \le 4^{-1}\|f\|_\infty n, \tag{2.5}$$

otherwise the probability under consideration equals zero. In addition, we can also assume that

$$x \ge 2\|f\|_\infty n^{1/p}, \tag{2.6}$$

otherwise what we have to prove is trivial as soon as κ is greater than or equal to $(16\|f\|_\infty)^p$. So from now on, we assume the two restrictions above on x.

The strategy to prove the desired inequalities is in two steps. First, we split the sum into blocks of size

$$t = [n^{1/p}]$$

as this is the characteristic size when dealing with mixing bounds of exponent p. Then, we write these blocks as sums of a martingale difference and a remainder. For each of these two terms, we will prove the desired estimate on the deviation probability using Theorem 2.1 with a suitable exponent r. While the different sizes of blocks and the filtrations we will introduce all depend on n, we suppress n from the notations for brevity.

Let

$$B_i = \sum_{j=(i-1)t+1}^{it} f^{(0)}(Y_j) \quad \text{and} \quad X_i = \mathbb{E}(B_i \mid \mathcal{F}_{(i-2)t}).$$

Let $n_t = [n/t]$ be the number of size t blocks. The following inequality is then valid:

$$\max_{1\le k\le n}\left|\sum_{i=1}^{k}(f(Y_i)-\pi(f)\right| \le 2t\|f\|_\infty + \max_{1\le j\le n_t}\left|\sum_{i=1}^{j}(B_i - X_i)\right| + \max_{1\le j\le n_t}\left|\sum_{i=1}^{j} X_i\right|.$$

Since $2t\|f\|_\infty \le 2n^{1/p}\|f\|_\infty \le x$, it follows that

$$\mathbb{P}\left(\max_{1\le k\le n}\left|\sum_{i=1}^{k}(f(Y_i)-\pi(f)\right| \ge 8x\right) \le \mathbb{P}\left(\max_{1\le j\le n_t}\left|\sum_{i=1}^{j}(B_i - X_i)\right| \ge 3x\right) + \mathbb{P}\left(\max_{1\le j\le n_t}\left|\sum_{i=1}^{j} X_i\right| \ge 4x\right). \tag{2.7}$$

We will control separately these two terms.

First step. Controlling $\mathbb{P}(\max_{1\le j\le n_t}|\sum_{i=1}^{j} X_i| \ge 4x)$.

Consider some $r \in (2(p-1), 2p)$. We will show that

$$\mathbb{P}\left(\max_{1\le j\le n_t}\left|\sum_{i=1}^{j} X_i\right| \ge 4x\right) \le \begin{cases} \kappa n x^{-p} & \text{if } p>2 \\ \kappa(nx^{-2} + (n\log n)^{r/2}x^{-r}) & \text{if } p=2, \end{cases} \tag{2.8}$$

where κ is a positive constant depending only on p, r, $\|f\|$, C_1 and C_2 but not on x or n.

With this aim, we first let

$$u = \left[\frac{x}{2\|f\|_\infty n^{1/p}}\right]$$

and we notice that, by (2.6), $u \ge 1$. We will regroup the X_i into blocks of length u, which corresponds to blocks of size $tu \asymp x$ for Y_j: this is the time scale where the sum over a block cannot exceed x. By (2.5),

$$n_t \ge \frac{n}{2t} \ge \frac{n}{2n^{1/p}} \ge \frac{4x}{2n^{1/p}\|f\|_\infty} \ge 4u. \tag{2.9}$$

Define

$$U_i = \sum_{j=(i-1)u+1}^{iu} X_j.$$

It is measurable with respect to $\mathcal{G}_i^U = \sigma(Y_i, \ell \leq itu - 2t)$ thanks to the conditional expectation in the definition of X_i. Since $\|X_i\|_\infty \leq 2\|f\|_\infty t$, we have

$$\max_{1\leq j\leq n_t} \left|\sum_{i=1}^{j} X_i\right| \leq 2\|f\|_\infty tu + \max_{1\leq j\leq [n_t/u]} \left|\sum_{i=1}^{j} U_i\right|.$$

Since $2\|f\|_\infty tu \leq x$, it follows that

$$\mathbb{P}\left(\max_{1\leq j\leq n_t} \left|\sum_{i=1}^{j} X_i\right| \geq 4x\right) \leq \mathbb{P}\left(\max_{1\leq j\leq [n_t/u]} \left|\sum_{i=1}^{j} U_i\right| \geq 3x\right),$$

which we will control using Theorem 2.1 applied to $Z_i = U_i$ and $\mathcal{G}_i = \mathcal{G}_i^U$ and $N = [n_t/u]$ and the exponent r. We should thus show that all the terms in the upper bound of this theorem are controlled as in (2.8).

By using $\mathbf{H}_1(p)$,

$$\|\mathbb{E}(U_2 \mid \mathcal{G}_0^U)\|_1 \leq \sum_{i=u+1}^{2u} \sum_{j=(i-1)t+1}^{it} \|\mathbb{E}(f^{(0)}(Y_j) \mid \mathcal{F}_0)\|_1 \leq C_1\|f^{(0)}\| \frac{ut}{(ut)^{p-1}}.$$

Note now that $ut \geq x(8\|f\|_\infty)^{-1}$. Therefore,

$$\frac{[n_t/u]}{x}\|\mathbb{E}(U_2 \mid \mathcal{G}_0^U)\|_1 \leq C_1\|f^{(0)}\| \frac{n_t/u}{x} \frac{ut}{(ut)^{p-1}} \leq C_1\|f^{(0)}\|(8\|f\|_\infty)^{p-1} n x^{-p}.$$

This handles the first term in the upper bound of Theorem 2.1.

To control the term involving $\mathbb{E}(|U_1|^r)$, we recall that U_1 is a sum of u random variables X_i, all bounded in sup norm by $2\|f\|_\infty t$. Any precise inequality for the r norm of a sum will do here. We use for instance Theorem 2.5 in [24] with $p = r/2$. It gives

$$\mathbb{E}(|U_1|^r) \leq (ur)^{r/2} \left(\sum_{i=1}^{u} \|X_1\|_\infty \|\mathbb{E}_0(X_i)\|_{r/2}\right)^{r/2}.$$

Moreover,

$$\begin{aligned}
\|\mathbb{E}_0(X_i)\|_{r/2} &\leq \sum_{j=(i-1)t+1}^{it} \|\mathbb{E}_0(\mathbb{E}_{(i-2)t} f^{(0)}(Y_j))\|_{r/2} \\
&= \sum_{j=(i-1)t+1}^{it} \|\mathbb{E}_{(i-2)t\wedge 0} f^{(0)}(Y_j)\|_{r/2} \\
&\leq \sum_{j=(i-1)t+1}^{it} [\|\mathbb{E}_{(i-2)t\wedge 0} f^{(0)}(Y_j)\|_\infty^{r/2-1} \|\mathbb{E}_{(i-2)t\wedge 0} f^{(0)}(Y_j)\|_1]^{1/(r/2)}.
\end{aligned}$$

The first sup norm is bounded by $2\|f\|_\infty \le 2\|f\|$, while the L^1 norm is bounded by $C_1\|f^{(0)}\|/(t \vee (i-1)t)^{p-1}$, thanks to $\mathbf{H}_1(p)$. Hence, $\|\mathbb{E}_0(X_i)\|_{r/2} \le \frac{Ct}{t^{(p-1)/(r/2)}} \frac{1}{(1\vee(i-1))^{(p-1)/(r/2)}}$, for some constant C. As $r > 2(p-1)$, we get a bound

$$\begin{aligned}\mathbb{E}(|U_1|^r) &\le Cu^{r/2}\left(\sum_{i=1}^{u} t \cdot \frac{t}{t^{(p-1)/(r/2)}} \frac{1}{(1 \vee (i-1))^{(p-1)/(r/2)}}\right)^{r/2} \\ &\le C' \frac{(t^2 u)^{r/2}}{t^{(p-1)}} \cdot u^{r/2-(p-1)}.\end{aligned}$$

Taking into account that $ut \le x/(2\|f\|_\infty)$ and $n_t = [n/t]$, we derive

$$\frac{[n_t/u]}{x^r}\mathbb{E}(|U_1|^r) \le \kappa n x^{-p}.$$

This handles the second term in the upper bound of Theorem 2.1.

Let us now control the term involving $\mathbb{E}(U_1^2)$. We have

$$\begin{aligned}\mathbb{E}(U_1^2) &= \sum_{j=1}^{u}\sum_{\ell=1}^{u} \mathbb{E}(X_j X_\ell) \\ &= \sum_{j=1}^{u}\sum_{\ell=1}^{u} \sum_{k=(j-1)t+1}^{jt} \sum_{m=(\ell-1)t+1}^{\ell t} \mathbb{E}[\mathbb{E}_{(j-2)t}(f^{(0)}(Y_k)) \cdot \mathbb{E}_{(\ell-2)t}(f^{(0)}(Y_m))].\end{aligned}$$

Each such term is equal to $\mathbb{E}[\mathbb{E}_{(j-2)t\wedge(\ell-2)t}(f^{(0)}(Y_k)) \cdot \mathbb{E}_{(j-2)t\wedge(\ell-2)t}(f^{(0)}(Y_m))]$. We bound one of the factors (corresponding to the minimal j or ℓ) by $2\|f\|_\infty$, and use $\mathbf{H}_1(p)$ to bound the other one in terms of the gap size, which is at least $(|\ell - j|+1)t$. Hence,

$$\mathbb{E}(U_1^2) \le 2\|f\|_\infty \sum_{j=1}^{u}\sum_{\ell=1}^{u} t^2 \frac{C_1\|f\|}{t^{p-1}(|j-\ell|+1)^{p-1}} \le \kappa u t^2 \times \frac{1}{t^{p-1}}(1 + (\log n)\mathbf{1}_{p=2})$$

as $p \ge 2$, where κ is a positive constant. This yields

$$\begin{aligned}[n_t/u]^{r/2}\mathbb{E}(U_1^2)^{r/2} &\le \kappa^{r/2}\left(\frac{n}{tu} u t^2 \times \frac{1}{t^{p-1}}\right)^{r/2}\left(1 + (\log n)^{r/2}\mathbf{1}_{p=2}\right) \\ &\le \kappa^{r/2}\left(n t^2 \times \frac{1}{t^p}\right)^{r/2}(1 + (\log n)^{r/2}\mathbf{1}_{p=2}) \\ &\le 2^{rp/2}\kappa^{r/2} n^{r/p}(1 + (\log n)^{r/2}\mathbf{1}_{p=2}).\end{aligned}$$

Using (2.6), we note that as $r \ge p$ we have $x^{r-p} \ge 2^{r-p}\|f\|_\infty^{r-p} n^{r/p-1}$. Hence,

$$\frac{[n_t/u]^{r/2}}{x^r}\mathbb{E}(U_1^2)^{r/2} \le 2^{rp/2}\kappa^{r/2}\left(2^{p-r}\|f\|_\infty^{p-r}\frac{n}{x^p} + \frac{(n\log n)^{r/2}}{x^r}\mathbf{1}_{p=2}\right).$$

This handles the third term in the upper bound of Theorem 2.1.

We analyze now the last term in the upper bound of Theorem 2.1. With this aim, we notice that $\mathcal{G}_0^U = \mathcal{F}_{-2t}$. Therefore, for any $i \geq 2$,

$$\begin{aligned}\|\mathbb{E}(U_i^2 \mid \mathcal{G}_0^U) - \mathbb{E}(U_i^2)\|_{r/2} &\leq \sum_{j=(i-1)u+1}^{iu} \sum_{\ell=(i-1)u+1}^{iu} \|\mathbb{E}(X_j X_\ell \mid \mathcal{F}_{-2t}) - \mathbb{E}(X_j X_\ell)\|_{r/2} \\ &\leq 2 \sum_{j=(i-1)u+3}^{iu+2} \sum_{\ell=j}^{iu+2} \|\mathbb{E}_0^{(0)}(X_j X_\ell)\|_{r/2}.\end{aligned}$$

Fix $j \in [(i-1)u+3, iu+2]$ and $\ell \in [j, iu+2]$. Then $X_j X_\ell$ is a sum of t^2 terms of the form $\mathbb{E}_{(j-2)t}(f^{(0)}(Y_k)) \cdot \mathbb{E}_{(\ell-2)t}(f^{(0)}(Y_m))$ for $k \in [(j-1)t+1, jt]$ and $m \in [(\ell-1)t+1, \ell t]$. For each such term, writing $k' = k-(j-2)t$ and $m' = m-(j-2)t$, we have

$$\begin{aligned}&\|\mathbb{E}_0^{(0)}[\mathbb{E}_{(j-2)t}(f^{(0)}(Y_k)) \cdot \mathbb{E}_{(\ell-2)t}(f^{(0)}(Y_m))]\|_{r/2} \\ &\quad = \|K^{(j-2)t}[(K^{k'} f^{(0)}) \cdot (K^{m'} f^{(0)})] - \pi[(K^{k'} f^{(0)}) \cdot (K^{m'} f^{(0)})]\|_{\pi,r/2}.\end{aligned}$$

Therefore

$$\begin{aligned}&\|\mathbb{E}_0^{(0)}[\mathbb{E}_{(j-2)t}(f^{(0)}(Y_k)) \cdot \mathbb{E}_{(\ell-2)t}(f^{(0)}(Y_m))]\|_{r/2}^{r/2} \\ &\quad \leq (8\|f\|_\infty^2)^{r/2-1} \|K^{(j-2)t}[(K^{k'} f^{(0)}) \cdot (K^{m'} f^{(0)})] - \pi[(K^{k'} f^{(0)}) \cdot (K^{m'} f^{(0)})]\|_{\pi,1}.\end{aligned}$$

Both functions $K^{k'}(f^{(0)})$ and $K^{m'}(f^{(0)})$ belong to $\mathcal{B}$, with a norm bounded by $C_2\|f^{(0)}\|$ thanks to the condition $\mathbf{H}_2$. As $\|\cdot\|$ is a Banach algebra norm, their product also belongs to $\mathcal{B}$. Applying the condition $\mathbf{H}_1(p)$ to this product, we deduce that

$$\|K^{(j-2)t}[(K^{k'} f^{(0)}) \cdot (K^{m'} f^{(0)})] - \pi[(K^{k'} f^{(0)}) \cdot (K^{m'} f^{(0)})]\|_{\pi,1} \leq \frac{C_3}{((j-2)t)^{p-1}},$$

for some constant C_3. Combining these inequalities yields

$$\|\mathbb{E}_0^{(0)}(X_j X_\ell)\|_{r/2} \leq \frac{t^2 C_4}{((j-2)t)^{(p-1)/(r/2)}}.$$

Therefore, we get that for any $i \geq 2$,

$$\|\mathbb{E}(U_i^2 \mid \mathcal{G}_0) - \mathbb{E}(U_i^2)\|_{r/2} \leq C_5 \frac{(ut)^2}{(iut)^{2(p-1)/r}}.$$

As $r > 2(p-1)$, this implies that

$$\sum_{i=2}^{k} \|\mathbb{E}(U_i^2 \mid \mathcal{G}_0) - \mathbb{E}(U_i^2)\|_{r/2} \leq C_6 \frac{(ut)^2}{(ut)^{2(p-1)/r}} k^{1-2(p-1)/r}.$$

Hence

$$[n_t/u]\left[\sum_{k=1}^{[n_t/u]}\frac{1}{k^{1+2\delta/r}}\left(\sum_{i=2}^{k}\|\mathbb{E}(U_i^2\mid\mathcal{G}_0^U)-\mathbb{E}(U_i^2)\|_{r/2}\right)^{\delta}\right]^{r/(2\delta)}$$
$$\le C_6^{r/2}\frac{(ut)^r}{(ut)^{p-1}}\frac{n_t}{u}\left[\sum_{k=1}^{[n_t/u]}\frac{1}{k^{1+2\delta/r}}(k^{1-2(p-1)/r})^{\delta}\right]^{r/(2\delta)}.$$

As $r<2p$, the sum over k is uniformly bounded, independently of n or x. Taking into account that $ut\le x/(2\|f\|_\infty)$, we get that there exists a positive constant κ such that

$$\frac{[n_t/u]}{x^r}\left[\sum_{k=1}^{[n_t/u]}\frac{1}{k^{1+2\delta/r}}\left(\sum_{i=2}^{k}\|\mathbb{E}(U_i^2\mid\mathcal{G}_0^U)-\mathbb{E}(U_i^2)\|_{r/2}\right)^{\delta}\right]^{r/(2\delta)}\le\kappa n x^{-p}. \quad (2.10)$$

This handles the last term in the upper bound of Theorem 2.1. Altogether, this proves (2.8) and concludes the proof of the first step.

Second step. Controlling $\mathbb{P}(\max_{1\le j\le n_t}|\sum_{i=1}^{j}B_i-X_i|\ge 3x)$.

We will prove

$$\mathbb{P}\left(\max_{1\le k\le n_t}\left|\sum_{i=1}^{k}(B_i-X_i)\right|\ge 3x\right)\le\begin{cases}\kappa n x^{-p}+\kappa\exp(-\kappa^{-1}x^2/n) & \text{if } p>2\\ \kappa n x^{-2}+\kappa\exp(-\kappa^{-1}x^2/(n\log n)) & \text{if } p=2,\end{cases} \quad (2.11)$$

where κ is a positive constant depending only on p, $\|f\|$, C_1 and C_2 but not on x or n. Starting from (2.7), this upper bound combined with (2.8) will end the proof of Items 1 and 2 of the theorem.

To prove (2.11), we start by setting

$$d_i=B_i-X_i \quad\text{and}\quad \mathcal{G}_i^B=\mathcal{F}_{it},$$

and we write the following decomposition:

$$\mathbb{P}\left(\max_{1\le k\le n_t}\left|\sum_{i=1}^{k}(B_i-X_i)\right|\ge 3x\right)$$
$$\le\mathbb{P}\left(\max_{1\le 2k\le n_t}\left|\sum_{i=1}^{k}d_{2i}\right|\ge 3x/2\right)+\mathbb{P}\left(\max_{1\le 2k-1\le n_t}\left|\sum_{i=1}^{k}d_{2i-1}\right|\ge 3x/2\right). \quad (2.12)$$

Note that $(d_{2i})_{i\in\mathbb{Z}}$ (resp. $(d_{2i-1})_{i\in\mathbb{Z}}$) is a strictly stationary sequence of martingale differences with respect to the non-decreasing filtration $(\mathcal{G}_{2i}^B)_{i\in\mathbb{Z}}$ (resp.

$(\mathcal{G}^B_{2i-1})_{i\in\mathbb{Z}})$. Therefore, since $\|d_{2i}\|_\infty \le 2t\|f\|_\infty \le 2\|f\|_\infty n^{1/p}$ a.s., by Proposition 2.1 in [13], for any $y > 0$,

$$\mathbb{P}\left(\max_{2\le 2k\le n_t}\left|\sum_{i=1}^{k} d_{2i}\right| \ge 3x/2\right) \le 2\exp\left(\frac{-9x^2}{16y}\right) + 2\exp\left(\frac{-9x}{16\|f\|_\infty n^{1/p}}\right)$$
$$+\,\mathbb{P}\left(\sum_{i=1}^{[n_t/2]} \mathbb{E}(d_{2i}^2 \mid \mathcal{G}^B_{2(i-1)}) \ge y\right). \tag{2.13}$$

Note now that

$$\mathbb{E}(d_{2i}^2 \mid \mathcal{G}^B_{2(i-1)}) \le \mathbb{E}(B_{2i}^2 \mid \mathcal{G}^B_{2(i-1)}).$$

Moreover, by stationarity, we infer that

$$\sum_{i=1}^{[n_t/2]} \mathbb{E}(B_{2i}^2) \le 2n\|f\|_\infty \sum_{k=0}^{t-1} \|\mathbb{E}_0(f^{(0)}(Y_k)\|_1.$$

Therefore, by $\mathbf{H}_1(p)$, there exists a positive constant κ depending only on p, $\|f\|$ and C_1 such that

$$\sum_{i=1}^{[n_t/2]} \mathbb{E}(B_{2i}^2) \le \kappa n(1 + (\log n)\mathbf{1}_{p=2})).$$

Selecting

$$y = \begin{cases} \max(2\kappa n, 16xn^{1/p}\|f\|_\infty) & \text{if } p > 2 \\ \max(2\kappa n\log n, 16x(n\log n)^{1/2}\|f\|_\infty) & \text{if } p = 2, \end{cases}$$

and starting from (2.13), we get that, for any $r \ge 1$,

$$\mathbb{P}\left(\max_{2\le 2k\le n_t}\left|\sum_{i=1}^{k} d_{2i}\right| \ge 3x/2\right)$$
$$\le c\frac{n}{x^p} + c\exp\left(-c'\frac{x^2}{n + n\log n\mathbf{1}_{p=2}}\right)$$
$$+\,\mathbb{P}\left(\sum_{i=1}^{[n_t/2]} (\mathbb{E}(B_{2i}^2 \mid \mathcal{G}^B_{2(i-1)}) - \mathbb{E}(B_{2i}^2)) \ge y/2\right), \tag{2.14}$$

where c and c' are positive constants.

Let us prove now that

$$\mathbb{P}\left(\left|\sum_{i=1}^{[n_t/2]} (\mathbb{E}(B_{2i}^2 \mid \mathcal{G}^B_{2(i-1)}) - \mathbb{E}(B_{2i}^2))\right| \ge y/2\right) \le cnx^{-p}, \tag{2.15}$$

where c is a positive constant depending only on p, $\|f\|$, C_1 and C_2 but not on x or n. A similar bound will hold for odd indices. Hence, starting from (2.12) and considering the inequality (2.14), this upper bound will lead to (2.11) and then will end the proof of Items 1 and 2 of the theorem.

It remains then to prove (2.15). With this aim, we do again blocks of size u with as before $u = [\frac{x}{2\|f\|_\infty n^{1/p}}]$. Let

$$W_i = \mathbb{E}(B_{2i}^2 \mid \mathcal{G}_{2(i-1)}^B) - \mathbb{E}(B_{2i}^2), \quad V_i = \sum_{k=(i-1)u+1}^{iu} W_k$$

and $\mathcal{G}_i^V = \mathcal{G}_{2(iu-1)}^B = \mathcal{F}_{2(iu-1)t}$. Setting $n_u = [\frac{[n_t/2]}{u}]$ (note that, by (2.9), $n_u \geq 1$), we have

$$\left|\sum_{i=1}^{[n_t/2]} (\mathbb{E}(B_{2i}^2 \mid \mathcal{G}_{2(i-1)}^B) - \mathbb{E}(B_{2i}^2))\right| = \left|\sum_{i=1}^{[n_t/2]} W_i\right| \leq \left|\sum_{i=1}^{n_u} V_i\right| + 8ut^2\|f\|_\infty^2.$$

Note that

$$8ut^2\|f\|_\infty^2 \leq 4xn^{1/p}\|f\|_\infty \leq y/4.$$

Therefore

$$\mathbb{P}\left(\left|\sum_{i=1}^{[n_t/2]} (\mathbb{E}(B_{2i}^2 \mid \mathcal{G}_{2(i-1)}^B) - \mathbb{E}(B_{2i}^2))\right| \geq y/2\right) \leq \mathbb{P}\left(\left|\sum_{i=1}^{n_u} V_i\right| \geq y/4\right).$$

To prove (2.15), it suffices to show that

$$\mathbb{P}\left(\left|\sum_{i=1}^{n_u} V_i\right| \geq y/4\right) \leq cnx^{-p}. \tag{2.16}$$

We will show this inequality by applying Theorem 2.1 to $Z_i = V_i$ and $\mathcal{G}_i = \mathcal{G}_i^V$ and $N = n_u$ and some fixed $r \in (2p-2, 2p)$. We should thus show that all the terms in the upper bound of this theorem are controlled as in (2.16).

We start with the first term involving $\|\mathbb{E}(V_2 \mid \mathcal{G}_0^V)\|_1$. Since $y \geq 16xn^{1/p}\|f\|_\infty$, we have

$$\begin{aligned}
\frac{n_u}{y}\|\mathbb{E}(V_2 \mid \mathcal{G}_0^V)\|_1 &\leq \frac{n_u}{xn^{1/p}\|f\|_\infty} \sum_{k=u+1}^{2u} \|\mathbb{E}(W_k \mid \mathcal{F}_{-2t})\|_1 \\
&\leq \frac{n_u}{xn^{1/p}\|f\|_\infty} \sum_{k=u+1}^{2u} \|\mathbb{E}(B_{2k}^2 \mid \mathcal{F}_{-2t}) - \mathbb{E}(B_{2k}^2)\|_1 \\
&\leq \frac{2n_u}{xn^{1/p}\|f\|_\infty} \sum_{k=u+1}^{2u} \sum_{j=(2k-1)t+1}^{2kt} \sum_{\ell=j}^{2kt} \|\mathbb{E}^{(0)}(f^{(0)}(Y_j)f^{(0)}(Y_\ell)|\mathcal{F}_{-2t})\|_1.
\end{aligned}$$

Using $\mathbf{H}(p)$, we infer that there exists a positive constant c depending on C_1, C_2 and $\|f^{(0)}\|$, such that this quantity is bounded by

$$c\frac{n}{xn^{1/p}tu\|f\|_\infty}ut^2\frac{1}{(ut)^{p-1}} \le c\frac{n}{x\|f\|_\infty}\frac{1}{(ut)^{p-1}} \le 8^{p-1}c\|f\|_\infty^{p-2}nx^{-p},$$

thanks to the inequality $ut \ge x(8\|f\|_\infty)^{-1}$. This handles the first term in the upper bound of Theorem 2.1.

We turn to the second term, involving $\mathbb{E}(|V_1|^r)$. By stationarity and Theorem 2.5 in [24], we have

$$\begin{aligned}\mathbb{E}(|V_1|^r)^{2/r} &= \left\|\sum_{k=1}^{u} W_k\right\|_r^2 \le ur\sum_{k=0}^{u-1}\|W_0\mathbb{E}(W_k \mid \mathcal{F}_{-2t})\|_{r/2} \\ &\le ur\|W_0^2\|_{r/2} + ur\|W_0\|_\infty\sum_{k=1}^{u-1}\|\mathbb{E}(W_k \mid \mathcal{F}_{-2t})\|_{r/2} \\ &\le ur\|W_0^2\|_{r/2} + ur\cdot 16t^2\|f\|_\infty^2\sum_{k=1}^{u-1}\|\mathbb{E}(W_k \mid \mathcal{F}_{-2t})\|_{r/2}.\end{aligned}$$

Using $\mathbf{H}(p)$, we infer that there exists a positive constant c_4 depending on C_1, C_2, $\|f\|$ and r such that

$$\begin{aligned}\|W_0^2\|_{r/2} = \|W_0\|_r^2 &= \|\mathbb{E}_0(B_2^2) - \mathbb{E}(B_2^2)\|_r^2 \\ &\le \left(\sum_{j=t+1}^{2t}\sum_{i=t+1}^{2t}\|\mathbb{E}_0^{(0)}(f^{(0)}(Y_j)f^{(0)}(Y_i))\|_r\right)^2 \\ &\le c_4\left(t^2\frac{1}{t^{(p-1)/r}}\right)^2 \le c_4\frac{u}{u^{2(p-1)/r}}\cdot\frac{t^4}{t^{2(p-1)/r}},\end{aligned}$$

as $r > 2(p-1)$. On the other hand, using again $\mathbf{H}(p)$, we get that there exists a positive constant c_5 such that for any $k \ge 1$,

$$\begin{aligned}\|\mathbb{E}(W_k \mid \mathcal{F}_{-2t})\|_{r/2} &\le \sum_{j=(2k-1)t+1}^{2kt}\sum_{i=(2k-1)t+1}^{2kt}\|\mathbb{E}_{-2t}^{(0)}(f^{(0)}(Y_j)f^{(0)}(Y_i))\|_{r/2} \\ &\le c_5t^2\frac{1}{(kt)^{2(p-1)/r}}.\end{aligned}$$

The sum of these quantities over k from 1 to $u-1$ is bounded by $c_6\frac{t^2}{t^{2(p-1)/r}}\cdot\frac{u}{u^{2(p-1)/r}}$, as $r > 2(p-1)$. We infer that there exists a positive constant c_7 such that

$$n_u\mathbb{E}(|V_1|^r) \le c_7n_u(u^2t^4)^{r/2}\frac{1}{(ut)^{p-1}} \le c_7nt^r(ut)^{r-p}.$$

Hence, using the fact that $y \ge 16xn^{1/p}\|f\|_\infty$ and $ut \le x(2\|f\|_\infty)^{-1}$ and $t \le n^{1/p}$, we get that

$$\frac{n_u}{y^r}\mathbb{E}(|V_1|^r) \le 8^{-r}c_7(2\|f\|_\infty)^{p-2r}\frac{n}{x^p}.$$

This handles the second term in the upper bound of Theorem 2.1.

We turn to the third term, involving $\mathbb{E}(V_1^2)$. By stationarity, we have

$$\mathbb{E}(V_1^2) = \left\| \sum_{k=u+1}^{2u} W_k \right\|_2^2 = u\|W_1\|_2^2 + 2\sum_{k=1}^{u-1}\sum_{\ell=1}^{u-k} \mathrm{cov}(W_0, W_\ell).$$

But, by using $\mathbf{H}(p)$, we infer that there exists a positive constant c_1 such that

$$\|W_1\|_2 = \|\mathbb{E}_0^{(0)}(B_2^2)\|_2 \le \sum_{j=t+1}^{2t}\sum_{\ell=t+1}^{2t} \|\mathbb{E}_0^{(0)}(f^{(0)}(Y_j)f^{(0)}(Y_\ell))\|_2 \le c_1 \frac{t^2}{t^{(p-1)/2}}.$$

On the other hand, using again $\mathbf{H}(p)$, we get that there exists a positive constant c_2 such that for any $\ell \ge 1$,

$$\begin{aligned}
|\mathrm{cov}(W_0, W_\ell)| &\le \|B_0^2\|_\infty \|\mathbb{E}_{-2t}(B_{2\ell}^2) - \mathbb{E}(B_{2\ell}^2)\|_1 \\
&\le (8t\|f\|)^2 \sum_{j=(2\ell-1)t+1}^{2\ell t} \sum_{i=(2\ell-1)t+1}^{2\ell t} \|\mathbb{E}_0^{(0)}(f^{(0)}(Y_j)f^{(0)}(Y_i))\|_1 \\
&\le c_2 \frac{t^4}{(\ell t)^{p-1}}.
\end{aligned}$$

So, overall, there exists a positive constant c_3 such that

$$\mathbb{E}(V_1^2) \le c_3 u \frac{t^4}{t^{p-1}}(1 + (\log n)\mathbf{1}_{p=2}).$$

This upper bound implies that

$$(n_u \mathbb{E}(V_1^2))^{r/2} \le (2^p c_3)^{r/2} n^{2r/p}(1 + (\log n)^{r/2}\mathbf{1}_{p=2}).$$

Next using the fact that $y \ge 16xn^{1/p}\|f\|_\infty$ if $p > 2$ and $y \ge 16x(n\log n)^{1/2}\|f\|_\infty$ if $p = 2$, we get

$$\frac{n_u^{r/2}}{y^r}\mathbb{E}(V_1^2)^{r/2} \le (2^p c_3)^{r/2} \frac{n^{r/p}}{16^r x^r \|f\|_\infty^r}.$$

By (2.6) and since $r \ge p$, we have $x^{r-p} \ge (2\|f\|_\infty)^{r-p} n^{r/p-1}$. Therefore,

$$\frac{n_u^{r/2}}{y^r}\mathbb{E}(V_1^2)^{r/2} \le (2^{p-6} c_3)^{r/2} (2\|f\|_\infty)^{p-2r} \frac{n}{x^p}.$$

This handles the third term in the upper bound of Theorem 2.1.

Finally, we turn to the last term, involving $\|\mathbb{E}(V_i^2 \mid \mathcal{G}_0^V) - \mathbb{E}(V_i^2)\|_{r/2}$. For any $i \ge 2$, we have

$$\begin{aligned}
&\|\mathbb{E}(V_i^2 \mid \mathcal{G}_0^V) - \mathbb{E}(V_i^2)\|_{r/2} \\
&\quad \le \sum_{\ell=(i-1)u+1}^{iu} \sum_{m=(i-1)u+1}^{iu} \|\mathbb{E}_{-2t}^{(0)}(W_\ell W_m)\|_{r/2}
\end{aligned}$$

$$\leq \sum_{\ell=(i-1)u+1}^{iu} \sum_{m=(i-1)u+1}^{iu} \|\mathbb{E}_{-2t}^{(0)}[\mathbb{E}(B_{2\ell}^2 \mid \mathcal{G}_{2(\ell-1)}^B)\mathbb{E}(B_{2m}^2 \mid \mathcal{G}_{2(m-1)}^B)]\|_{r/2}$$

$$+ 2 \sum_{\ell=(i-1)u+1}^{iu} \sum_{m=(i-1)u+1}^{iu} \mathbb{E}(B_{2\ell}^2) \cdot \|\mathbb{E}_{-2t}^{(0)}(B_{2m}^2)\|_{r/2},$$

where this expansion is obtained from the definition $W_i = \mathbb{E}(B_{2i}^2 \mid \mathcal{G}_{2(i-1)}^B) - \mathbb{E}(B_{2i}^2)$ by expanding the product $W_\ell W_m$, using the fact that $\mathbb{E}_{-2t}^{(0)}$ is linear and vanishes on the constant $\mathbb{E}(B_{2\ell}^2) \cdot \mathbb{E}(B_{2m}^2)$.

For any $m \geq \ell \geq 1$,

$$\begin{aligned}
&\|\mathbb{E}_{-2t}^{(0)}[\mathbb{E}(B_{2\ell}^2 \mid \mathcal{G}_{2(\ell-1)}^B) \cdot \mathbb{E}(B_{2m}^2 \mid \mathcal{G}_{2(m-1)}^B)]\|_{r/2} \\
&\quad = \|\mathbb{E}_{-2t}^{(0)}[\mathbb{E}(B_{2\ell}^2 \mid \mathcal{G}_{2(\ell-1)}^B) \cdot \mathbb{E}(B_{2m}^2 \mid \mathcal{G}_{2(\ell-1)}^B)]\|_{r/2} \\
&\quad \leq \sum_{a,a'=(2\ell-1)t+1}^{2\ell t} \sum_{b,b'=(2m-1)t+1}^{2mt} \|\mathbb{E}_{-2t}^{(0)}[\mathbb{E}_{2(\ell-1)t}(f^{(0)}(Y_a)f^{(0)}(Y_{a'})) \\
&\qquad \times \mathbb{E}_{2(\ell-1)t}(f^{(0)}(Y_b)f^{(0)}(Y_{b'}))]\|_{r/2} \\
&\quad \leq 2t^2 \sup_{a',b' \geq 0} \sum_{a=(2\ell+1)t+1}^{2(\ell+1)t} \sum_{b=(2m+1)t+1}^{2(m+1)t} \|\mathbb{E}_0^{(0)}[\mathbb{E}_{2\ell t}(f^{(0)}(Y_a)f^{(0)}(Y_{a+a'})) \\
&\qquad \times \mathbb{E}_{2\ell t}(f^{(0)}(Y_b)f^{(0)}(Y_{b+b'}))]\|_{r/2},
\end{aligned}$$

where we have used stationarity. But

$$\mathbb{E}_{2\ell t}(f^{(0)}(Y_a)f^{(0)}(Y_{a+a'})) = \mathbb{E}_{2\ell t}((f^{(0)}K^{a'}f^{(0)})(Y_a)) = (K^{a-2\ell t}(f^{(0)}K^{a'}f^{(0)}))(Y_{2\ell t}).$$

Hence,

$$\begin{aligned}
&\mathbb{E}_0[\mathbb{E}_{2\ell t}(f^{(0)}(Y_a)f^{(0)}(Y_{a+a'})) \cdot \mathbb{E}_{2\ell t}(f^{(0)}(Y_b)f^{(0)}(Y_{b+b'}))] \\
&\quad = (K^{2\ell t}[K^{a-2\ell t}(f^{(0)}K^{a'}f^{(0)}) \cdot K^{b-2\ell t}(f^{(0)}K^{b'}f^{(0)})])(Y_0).
\end{aligned}$$

Therefore, thanks to $\mathbf{H}(p)$, we infer that there exists a positive constant c_8 such that for any $m \geq \ell \geq (i-1)u + 1$,

$$\|\mathbb{E}_{-2t}^{(0)}[\mathbb{E}(B_{2\ell}^2 \mid \mathcal{G}_{2(\ell-1)}^B) \cdot \mathbb{E}(B_{2m}^2 \mid \mathcal{G}_{2(m-1)}^B)]\|_{r/2} \leq c_8 \frac{t^4}{((i-1)tu)^{2(p-1)/r}}.$$

On the other hand, using again $\mathbf{H}(p)$, we infer that there exists a positive constant c_9 such that for any $\ell, m \geq (i-1)u + 1$,

$$\mathbb{E}(B_{2\ell}^2) \cdot \|\mathbb{E}_{-2t}^{(0)}(B_{2m}^2)\|_{r/2} \leq c_9 t^2 \cdot \frac{t^2}{((i-1)tu)^{2(p-1)/r}}.$$

So, overall, as $r > 2(p-1)$, there exists a positive constant c_{10} such that

$$\sum_{i=1}^{k} \|\mathbb{E}(V_i^2 \mid \mathcal{G}_0^V) - \mathbb{E}(V_i^2)\|_{r/2} \leq c_{10} \frac{u^2 t^4}{(tu)^{2(p-1)/r}} \frac{k}{k^{2(p-1)/r}}.$$

Therefore, as in addition $r < 2p$, there exists a positive constant c_{11} such that

$$\left[\sum_{k=1}^{n_u} \frac{1}{k^{1+2\delta/r}} \left(\sum_{i=1}^{k} \|\mathbb{E}(V_i^2 \mid \mathcal{G}_0^V) - \mathbb{E}(V_i^2)\|_{r/2}\right)^{\delta}\right]^{r/(2\delta)} \leq c_{11} \frac{(ut^2)^r}{(tu)^{p-1}}.$$

Using the fact that $y \geq 16xn^{1/p}\|f\|_\infty$, $ut \leq x(2\|f\|_\infty)^{-1}$ and $t \leq n^{1/p}$, this implies that

$$\frac{n_u}{y^r} \left[\sum_{k=1}^{n_u} \frac{1}{k^{1+2\delta/r}} \left(\sum_{i=1}^{k} \|\mathbb{E}(V_i^2 \mid \mathcal{G}_0^V) - \mathbb{E}(V_i^2)\|_{r/2}\right)^{\delta}\right]^{r/(2\delta)} \leq 8^{-r} c_{11} (2\|f\|_\infty)^{p-2r} \frac{n}{x^p}.$$

This handles the last term in the upper bound of Theorem 2.1. Altogether, this proves (2.16). This concludes the second step, and therefore the proof of Items 1 and 2 of the theorem. □

3. Lower Bounds in Moderate Deviations: Three Examples

In this section, we exhibit several examples of Markov chains satisfying $\mathbf{H}(p)$ (for different norms) for which one can prove a lower bound for the deviation probability of some particular observables. This shows that the upper bounds given in Theorem 1.1 cannot be essentially improved.

3.1. *Discrete Markov chains*

Let $p > 1$. We consider a simple renewal type Markov chain on $\mathbb{N}$, jumping from 0 to $n > 0$ with probability $p_{0,n} := 1/(\zeta(p+1)n^{p+1})$ and from $n > 0$ to $n-1$ with probability 1. This Markov chain has an invariant probability measure π given by $\pi\{n\} = \sum_{i \geq n} d/i^{p+1}$ for $n > 0$ and $\pi\{0\} = \pi\{1\}$, where $d > 0$ is chosen so that π is of mass 1.

This Markov chain satisfies $\mathbf{H}(p)$ for the norm $\|f\| = \|f\|_\infty$. Indeed, in this case,

$$\pi\left(\sup_{\|f\|_\infty \leq 1} |K^n(f) - \pi(f)|\right) \leq C_1 \sum_{j \geq n} \sum_{k \geq j+1} p_{0,k} \leq C_2 n^{1-p}$$

(see [6] or Chap. 30 in [4] for more details).

Define a function f by $f(n) = \pi\{0\} - \mathbf{1}_{n=0}$. Its average under π vanishes.

Proposition 3.1. *Let $(Y_i)_{i \in \mathbb{N}}$ be a stationary Markov chain with transition kernel described above, for some $p > 1$. There exists $\kappa > 0$ such that, for any $n \in \mathbb{N}^*$ and any $x \in [\kappa n^{1/p}, \kappa^{-1} n]$,*

$$\mathbb{P}\left(\sum_{i=0}^{n-1} f(Y_i) \geq x\right) \geq \kappa^{-1} \frac{n}{x^p}.$$

This paragraph is devoted to the proof of this proposition. Since we are looking for lower bound, it suffices to consider trajectories starting from 0. Denote by $\tau_0, \tau_1, \dots$ the lengths of the successive excursions outside of 0. This is a sequence of i.i.d. random variables with a weak moment of order p, namely: $\mathbb{P}(\tau_0 > n \mid Y_0 = 0) = \sum_{i \geq n} 1/(\zeta(p+1) i^{p+1})$. We first consider the case $p > 2$, and indicate then the modifications to be done when $p = 2$ and when $p \in (1, 2)$.

First, we study the probability that the lengths of excursions differ much from their average.

Lemma 3.1. *Assume $p > 2$. There exists $C_1 > 0$ such that, for any $n \geq 1$ and any $x \geq n^{1/p}$, one has*

$$\mathbb{P}\left(\sum_{i=0}^{n-1} \tau_i \geq n\mathbb{E}(\tau) + x\right) \geq C_1^{-1} \frac{n}{x^p}.$$

Proof. Write $\bar{\tau}_i = \tau_i - \mathbb{E}(\tau_i)$. There exists $\sigma^2 > 0$ such that $\sum_{i=0}^{n-1} \bar{\tau}_i / \sqrt{n}$ converges to $\mathcal{N}(0, \sigma^2)$. It follows that, for $x \in [n^{1/p}, n^{1/2}]$, the left-hand side in the statement of the lemma converges to a quantity which is bounded from below by $\mathbb{P}(\mathcal{N}(0, \sigma^2) \geq 1) > 0$, while the right-hand side is bounded from above by C_1^{-1}. Taking C_1 large enough, the conclusion of the lemma follows for x in this range.

Let us now assume $x \geq \sqrt{n}$. For $i < n$, let

$$A_i = \{\bar{\tau}_i \geq 3x\} \cap \left\{\sum_{j=0}^{i-1} \bar{\tau}_j \leq x\right\} \cap \left\{\sum_{j=i+1}^{n-1} \bar{\tau}_j \leq x\right\}.$$

This decomposition is the intersection of three independent sets. The first one has probability at least c/x^p as τ has polynomial tails of order p, while the measure of the other ones is bounded from below thanks to the central limit theorem for $\bar{\tau}$, as we assume $x \geq \sqrt{n}$. Hence, for some constant c_1, we obtain

$$\mathbb{P}(A_i) \geq c_1 / x^p.$$

Moreover, $A_i \cap A_j$ is contained in $\{\bar{\tau}_i \geq 3x\} \cap \{\bar{\tau}_j \geq 3x\}$. By independence, this set has probability at most c_2 / x^{2p} for some $c_2 > 0$.

On the set $\bigcup A_i$, one has $\sum_{i=0}^{n-1} \tau_i \geq n\mathbb{E}(\tau) + x$ by construction. To conclude, we should bound from below the measure of this set. We have

$$\mathbb{P}\left(\bigcup A_i\right) \geq \sum_{i=0}^{n-1} \mathbb{P}(A_i) - \sum_{i \neq j = 0}^{n-1} \mathbb{P}(A_i \cap A_j) \geq c_1 \frac{n}{x^p} - c_2 \frac{n^2}{x^{2p}}.$$

If n is large enough, one has $c_2 n^2 / x^{2p} \leq c_1 n / (2x^p)$ when $x \geq \sqrt{n}$. Therefore, we get $\mathbb{P}(\bigcup A_i) \geq (c_1/2) n / x^p$, proving the desired result. As the estimate is trivial for bounded n, the result follows. □

Proof of Proposition 3.1 for $p > 2$. Fix some $n \in \mathbb{N}$. Let N denote the number of visits to 0 of the Markov chain Y_i starting from 0, strictly before time n. Then, given the definition of f, one has

$$\sum_{i=0}^{n-1} f(Y_i) = n\pi\{0\} - N.$$

Therefore, for any $x \geq 0$,

$$\left\{\sum_{i=0}^{n-1} f(Y_i) \geq x\right\} = \{N \leq n\pi\{0\} - x\} = \left\{\sum_{j=0}^{[n\pi\{0\}-x]-1} \tau_j \geq n\right\}.$$

Let $m = [n\pi\{0\} - x]$. It is positive when $x \leq \kappa^{-1}n$, if κ is large enough. We write n as $m\mathbb{E}(\tau) + y$ for some y. As $\mathbb{E}(\tau) = 1/\pi\{0\}$ by Kac formula, we have

$$y = n - [n\pi\{0\} - x]/\pi\{0\} \geq x/\pi\{0\}.$$

If $x \geq \kappa n^{1/p}$ with large enough κ, then $y \geq n^{1/p}$. Hence, we can apply Lemma 3.1 to obtain

$$\mathbb{P}_0\left(\sum_{i=0}^{n-1} f(Y_i) \geq x\right) \geq C_1^{-1}\frac{m}{y^p} \geq C_2^{-1}\frac{n}{x^p}.$$

We obtain the same lower bound for the random walk started from π, with an additional multiplicative factor $\pi\{0\}$. □

Proof of Proposition 3.1 for $p = 2$. In this case, $\sum_{j=0}^{n-1} \bar{\tau}_j/\sqrt{n\log n}$ converges to a Gaussian (see for instance [12]). Following the proof of Lemma 3.1, one deduces first that this lemma holds trivially for any $x \in [n^{1/p}, \sqrt{n\log n}]$, and also that it holds for any $x \geq \sqrt{n\log n}$. It follows then from the same proof as in the $p > 2$ case that the proposition holds for all $x \in [\kappa n^{1/p}, \kappa^{-1}n]$. □

Proof of Proposition 3.1 for $p < 2$. In this case, $\sum_{j=0}^{n-1} \bar{\tau}_j/n^{1/p}$ converges to a stable law (which is totally asymmetric of index p, see [12]). Hence, Lemma 3.1 holds for any $x \geq n^{1/p}$. It follows then from the same proof as in the $p > 2$ case that the proposition holds for all $x \in [\kappa n^{1/p}, \kappa^{-1}n]$. □

3.2. *Young towers*

As we will not need specifics of Young towers, we refer the reader to [16] for the precise definitions, recalling below only what we need for the current argument (and that of Sec. 4). A Young tower is a dynamical system T preserving a probability measure π, on a metric space Z, together with a subset Z_0 (the basis of the tower) for which the successive returns to Z_0 create some form of decorrelation. Thus, an important feature of the Young tower is the return time τ from Z_0 to itself, and in particular its integrability properties.

Starting from any $z \in Z$, there is a canonical way to choose at random a point among the preimages of z under T. This defines a Markov chain Y_n for which π

is stationary, and which is dual to the dynamics (in the sense that $Y_0, \dots, Y_{n-1}$ is distributed like $T^{n-1}z, \dots, z$ when z is picked according to π).

Consider now a Young tower $T : Z \to Z$ with invariant measure π for which the return time τ to the basis Z_0 of the tower satisfies $\pi\{\tau = n\} \sim c/n^{p+1}$ on Z_0, for some $p > 1$. In perfect analogy with the previous paragraph, we define a function f by $f = \pi(Z_0) - \mathbf{1}_{Z_0}$. Its average under π vanishes. The corresponding Markov chain satisfies $\mathbf{H}(p)$ for the Hölder norm on the tower, see for instance [16] and references therein.

Starting from Y_0 distributed according to π, we can consider $Y_0, T(Y_0), \dots, T^{n-1}(Y_0)$, or the dual Markov chain $Y_0, \dots, Y_{n-1}$. Then $Y_0, \dots, Y_{n-1}$ is distributed as $T^{n-1}(Y_0), \dots, Y_0$, as explained at the beginning of Sec. 4. It follows that moderate deviations controls for one process or the other are equivalent. We will state the lower bound statement for the Markov chain, but we will prove it using the dynamical time direction.

Proposition 3.2. *In this context, assume* $p > 1$. *There exists* $\kappa > 0$ *such that, for any* $n \in \mathbb{N}^*$ *and any* $x \in [\kappa n^{1/p}, \kappa^{-1} n]$,

$$\mathbb{P}\left(\sum_{i=0}^{n-1} f(Y_i) \geq x\right) \geq \kappa^{-1} \frac{n}{x^p}.$$

Proof. We work using the dynamical time direction. Starting from a point in the basis Z_0 of the tower, let $\tau_0, \tau_1, \dots$ denote the lengths of the successive excursions out of Z_0. The proof will be the same as for Proposition 3.1 (notice that the statement is exactly the same). The only difference is that the successive returns to the basis are not independent, which means that the proof of Lemma 3.1 has to be amended. We only give the proof for $p > 2$, as the other cases are virtually identical.

Let $T_0 : Z_0 \to Z_0$ be the map induced by T on the basis. It preserves the probability π_0 induced by π on Z_0. By definition, T_0 is a Gibbs–Markov map with onto branches, i.e. there is a partition α_0 of Z_0 into positive measure subsets, such that T_0 maps bijectively each $a \in \alpha_0$ to Z_0, with the following bounded distortion property. A length k cylinder is a set of the form $[a_0, \dots, a_{k-1}] = \bigcap_{i<k} T_0^{-i} a_i$ for some $a_0, \dots, a_{k-1} \in \alpha_0$. Then there exists a constant C such that, for any $k > 0$, for any length k cylinder A and for any measurable set B,

$$C^{-1}\pi_0(A)\pi_0(B) \leq \pi_0(A \cap T_0^{-k} B) \leq C\pi_0(A)\pi_0(B). \tag{3.1}$$

(See for instance the last line in Sec. 1 of [2].) This estimate readily extends if A is a union of length k cylinders.

We can now prove the analogue of Lemma 3.1 in our situation. Let $\bar{\tau}_i = \tau_i - \mathbb{E}(\tau_i)$. Define

$$A_i = \{\bar{\tau}_i \geq 3x\} \cap \left\{\sum_{j=0}^{i-1} \bar{\tau}_j \leq x\right\} \cap \left\{\sum_{j=i+1}^{n-1} \bar{\tau}_j \leq x\right\} = A_i^1 \cap A_i^2 \cap A_i^3.$$

We should show that, if $x \geq \sqrt{n}$, then $\pi_0(A_i) \geq c_1/x^p$ for some $c_1 > 0$ independent of i or n, and that $\pi_0(A_i \cap A_j) \leq c_2/x^{2p}$ for $i < j$. Then, the proof of Lemma 3.1 applies. In this lemma, the inequality $\mathbb{P}(A_i) \geq c_1/x^p$ follows from independence and the fact that $\mathbb{P}(A_i^2) \geq c$ and $\mathbb{P}(A_i^3) \geq c$ and $\mathbb{P}(A_i^1) \geq c/x^p$. In our context, these three inequalities still hold (the first two ones follow from the fact that the Birkhoff sums of τ satisfy the central limit theorem or converge to a stable law, see [1] and [2], and the last one from the assumptions on the tails of τ), but independence fails. It will be replaced by (3.1). Let us give the details. Recall that $\bar{\tau}_i = \tau(T_0^i z) - \pi(\tau) := \bar{\tau}(T_0^i z)$. Define

$$B_1 = \{y : \bar{\tau}(y) \geq 3x\}, \quad B_2 = \left\{ y : \sum_{j=0}^{i-1} \bar{\tau}(T_0^j y) \leq x \right\} \quad \text{and}$$

$$B_3 = \left\{ y : \sum_{j=0}^{n-1-(i+1)} \bar{\tau}(T_0^j y) \leq x \right\}.$$

We have $A_i^1 = T_0^{-i}(B_1)$, $A_i^2 = B_2$ and $A_i^3 = T_0^{-(i+1)} B_3$. Therefore,

$$\pi_0(A_i) = \pi_0(B_2 \cap T_0^{-i}(B_1 \cap T_0^{-1}(B_3))).$$

Applying inequality (3.1) with $k = i$, $A = B_2$ and $B = B_1 \cap T_0^{-1}(B_3)$ (which is possible since B_2 is a union of length i cylinders since τ is constant on elements of α_0), we get $\pi_0(A_i) \geq C^{-1} \pi_0(B_2) \pi_0(B_1 \cap T_0^{-1}(B_3))$. Next, applying again (3.1) this time with $k = 1$, $A = B_1$ and $B = B_3$ (which is possible since B_1 is a union of length 1 cylinders), we have $\pi_0(B_1 \cap T_0^{-1}(B_3)) \geq C^{-1} \pi_0(B_1) \pi_0(B_3)$. So overall,

$$\pi_0(A_i) \geq C^{-2} \pi_0(A_i^2) \pi_0(A_i^1) \pi_0(A_i^3).$$

This inequality replaces the independence assumption and implies that $\pi_0(A_i) \geq c_1/x^p$. The inequality $\pi_0(A_i \cap A_j) \leq c_2/x^{2p}$ is proved in the same way, using the upper bound in (3.1). $\square$

3.3. *Harris Markov chains with state space* $[0, 1]$

Let $a = p - 1$ with $p > 1$. Let λ denote the Lebesgue measure on $[0, 1]$. Define the probability laws ν and π by

$$\nu = (1 + a) x^a \lambda \quad \text{and} \quad \pi = a x^{a-1} \lambda.$$

We define now a strictly stationary Markov chain by specifying its transition probabilities $K(x, A)$ as follows:

$$K(x, A) = (1 - x) \delta_x(A) + x \nu(A),$$

where δ_x denotes the Dirac measure. Then π is the unique invariant probability measure of the chain with transition probabilities $K(x, \cdot)$. Let $(Y_i)_{i \in \mathbb{Z}}$ be the stationary Markov chain on $[0, 1]$ with transition probabilities $K(x, \cdot)$ and law π. For

$\gamma > 0$, we set

$$c_{a,\gamma} = \frac{a}{a+\gamma}, \quad X_i = f_\gamma(Y_i) - \mathbb{E}(f_\gamma(Y_i)) := Y_i^\gamma - c_{a,\gamma} \quad \text{and} \quad S_n = \sum_{i=0}^{n-1} X_i.$$

Denote by

$$\beta_n := \frac{1}{2}\pi\left(\sup_{\|f\|_\infty \le 1} |K^n(f) - \pi(f)|\right),$$

and set $T(x) = 1 - x$. According to Lemma 2 in [11],

$$\beta_n \le 3\,\mathbb{E}_\pi(T^{[n/2]}).$$

Note now that for any $b > -1$,

$$\int_0^1 (1-x)^k x^b \,\mathrm{d}x = k^{-(b+1)} \int_0^k (1 - x/k)^k x^b \,\mathrm{d}x.$$

Since for any $x \in [0,1]$, $\log(1-x) \le -x$, it follows that

$$\int_0^1 (1-x)^k x^b \,\mathrm{d}x \le k^{-(b+1)} \int_0^k e^{-x} x^b \,\mathrm{d}x \le k^{-(b+1)}\Gamma(b+1), \tag{3.2}$$

implying that

$$\mathbb{E}_\pi(T^k) \le a\Gamma(a)k^{-a}.$$

Therefore

$$\sup_{\|f\|_\infty \le 1} \pi(|K^n(f) - \pi(f)|) \le 2\beta_n \le Cn^{-a},$$

which shows that the condition $\mathbf{H}_1(p)$ is satisfied for the two norms $\|f\|_\infty$ and $\|f\|_{BV}$. For the norm $\|f\|_\infty$, the condition $\mathbf{H}_2$ is trivially satisfied with $C_2 = 1$. Hence, Theorem 1.1 applies to $(f_\gamma(Y_i))_{i\in\mathbb{Z}}$. We shall verify that the condition $\mathbf{H}_2$ also holds for the norm $\|f\|_{BV} = \|f\|_\infty + |\mathrm{d}f|$ at the end of this section. Concerning the lower bound, the following proposition holds:

Proposition 3.3. *Let $(Y_i)_{i\in\mathbb{N}}$ be a stationary Markov chain with transition kernel described above. Assume $p > 1$ and $\gamma > 0$. There exists $\kappa > 0$ such that, for any $n \in \mathbb{N}^*$ and any $x \in [\kappa n^{1/p}, \kappa^{-1}n]$,*

$$\mathbb{P}\left(\max_{1\le k\le n}\left|\sum_{i=0}^{k-1}(Y_i^\gamma - \mathbb{E}(Y_i^\gamma))\right| \ge x\right) \ge \kappa^{-1}\frac{n}{x^p}.$$

Proof. We first define a sequence $(T_k)_{k\ge 0}$ of stopping time as follows:

$$T_0 = \inf\{i > 0 : Y_i \ne Y_{i-1}\} \quad \text{and} \quad T_k = \inf\{i > T_{k-1} : Y_i \ne Y_{i-1}\} \quad \text{for } k > 0.$$

Let $\tau_k = T_{k+1} - T_k$. The r.v.'s $(Y_{T_k}, \tau_k)_{k\ge 0}$ are i.i.d., Y_{T_k} has law ν and the conditional distribution of τ_k given $Y_{T_k} = y$ is the geometric distribution $\mathcal{G}(1-y)$. We have in particular that τ_0 is integrable. The key inequality for proving the lower bound is the following:

$$\mathbb{P}\left(\max_{0\le k\le n-1}\tau_k|X_{T_k}|\ge 24x\right)\le 9\mathbb{P}\left(\max_{1\le k\le [n\mathbb{E}(\tau_1)]+1}|S_k|\ge x\right) + 3\mathbb{P}(T_n\ge 2[n\mathbb{E}(\tau_1)]+1). \tag{3.3}$$

Before proving it, let us show how it will entail the lower bound.

Using the fact that the r.v.'s $(Y_{T_k},\tau_k)_{k\ge 0}$ are i.i.d., Y_{T_k} has law ν and the conditional distribution of τ_k given $Y_{T_k}=y$ is the geometric distribution $\mathcal{G}(1-y)$, straightforward computations imply that for $x\ge\kappa n^{1/p}$ with κ large enough,

$$\mathbb{P}\left(\max_{0\le k\le n-1}\tau_k|X_{T_k}|\ge 24x\right)\ge C_{p,\gamma}\frac{n}{x^p}, \tag{3.4}$$

where

$$C_{p,\gamma}=\frac{1}{4}\left(\frac{c_{a,\gamma}\eta}{48}\right)^p p\Gamma(p), \quad \text{with } \eta = 1-(c_{a,\gamma}/2)^{1/\gamma}.$$

On the other hand,

$$\mathbb{P}(T_n\ge 2[n\mathbb{E}(\tau_1)]+1)\le\mathbb{P}\left(T_0+\sum_{i=0}^{n-1}(\tau_i-\mathbb{E}(\tau_i))\ge[n\mathbb{E}(\tau_1)]\right).$$

Since $\mathbb{E}(\tau_1)\ge 1$, this gives

$$\mathbb{P}(T_n\ge 2[n\mathbb{E}(\tau_1)]+1)\le\mathbb{P}(T_0\ge n/2)+\mathbb{P}\left(\sum_{i=0}^{n-1}(\tau_i-\mathbb{E}(\tau_i))\ge n/2\right).$$

Since $\mathbb{P}(T_0\ge n/2)\le\int_0^1(1-x)^{n/2}d\pi(x)$, according to (3.2)

$$\mathbb{P}(T_0\ge n/2)\le 2^a a n^{-a}\Gamma(a).$$

Assume from now that $p\ge 2$. Since the $(\tau_k)_{k\ge 0}$ are i.i.d., the Fuk–Nagaev inequality for independent random variables (see for instance Theorem B.3 and its proof in [24]) gives that, for any $u>0$ and any $v_n^2(u)\ge\sum_{i=0}^{n-1}\mathbb{E}((\tau_i\wedge u)^2)$,

$$\mathbb{P}\left(\sum_{i=0}^{n-1}(\tau_i-\mathbb{E}(\tau_i))\ge n/2\right)\le n\mathbb{P}(\tau_1\ge u)+\exp\left(-\frac{n}{4u}\log\left(1+\frac{nu}{2v_n^2(u)}\right)\right). \tag{3.5}$$

We shall apply this inequality with the following choice of u:

$$u=\frac{n}{8(p-1)}.$$

The selection of $v_n^2(u)$ will be different if $p>2$ or if $p=2$. Assume first that $p>2$. In this case, we take $v_n^2(u)=n\mathbb{E}(\tau_1^2)$. Since Y_{T_k} has law ν and the conditional distribution of τ_k given $Y_{T_k}=y$ is the geometric distribution $\mathcal{G}(1-y)$, simple computations give

$$\mathbb{E}(\tau_1^2)=\frac{p^2}{(p-1)(p-2)}:=c_p \quad \text{and then } v_n^2(u)=c_p n.$$

On the other hand, if $p = 2$, we first note that

$$\mathbb{E}((\tau_1 \wedge u)^2) = \mathbb{E}(\tau_1^2 \mathbf{1}_{\tau_1 \le u}) + u^2 \mathbb{P}(\tau_1 \ge u) \le \sum_{\ell=0}^{[u-1]} (2\ell+1)\mathbb{P}(\tau_1 \ge \ell) + u^2 \mathbb{P}(\tau_1 \ge u).$$

Now (3.2) implies that $\mathbb{P}(\tau_1 \ge \ell) \le 2\ell^{-2}$. Therefore, if $n \ge 8$,

$$\mathbb{E}((\tau_1 \wedge u)^2) \le \log(u) + 5 \le 3\log(n).$$

So, in case $p = 2$, we take $v_n^2(u) = 3n \log n$.

If $p > 2$, then (3.5) together with the fact that, by (3.2), $\mathbb{P}(\tau_1 \ge \ell) \le p\Gamma(p)\ell^{-p}$ imply that

$$\mathbb{P}\left(\sum_{i=0}^{n-1} (\tau_i - \mathbb{E}(\tau_i)) \ge n/2\right) \le p\Gamma(p) \times (8(p-1))^p n^{-p+1} \\ + (16(p-1)c_p)^{2(p-1)} n^{-2(p-1)}. \tag{3.6}$$

So, overall, starting from (3.3) and taking into account (3.4) and (3.6), we get that for κ large enough

$$\mathbb{P}\left(\max_{1 \le k \le [n\mathbb{E}(\tau_1)]+1} |S_k| \ge x\right) \ge 9^{-1} p\Gamma(p) \left\{4^{-1}\left(\frac{c_{a,\gamma}\eta}{48}\right)^p n x^{-p} - 6(8(p-1))^p n^{-p+1}\right\} \\ - 3^{-1}(16(p-1)c_p)^{2(p-1)} n^{-2(p-1)}.$$

Since $n^{-p} \le (x\kappa)^{-p}$ and $\mathbb{E}(\tau_1) = \frac{p}{p-1} \le 2$, it follows that for κ large enough

$$\mathbb{P}\left(\max_{1 \le k \le 2n+1} |S_k| \ge x\right) \ge \frac{2n}{\kappa x^p},$$

giving the lower bound when $p > 2$.

We turn now to the case when $p = 2$, we derive this time

$$\mathbb{P}\left(\sum_{i=0}^{n-1} (\tau_i - \mathbb{E}(\tau_i)) \ge n/2\right) \le 2 \times 8^2 n^{-1} + (3 \times 16)^2 (\log n)^2 n^{-2}.$$

Proceeding as before, the lower bound follows.

We end the proof by considering the case $1 < p < 2$. Let u be a positive real and set $\bar\tau_i = (\tau_i \wedge u)$. Note that

$$\sum_{i=0}^{n-1} (\tau_i - \mathbb{E}(\tau_i)) = \sum_{i=0}^{n-1} (\bar\tau_i - \mathbb{E}(\bar\tau_i)) + \sum_{i=0}^{n-1} ((\tau_i - u)_+ - \mathbb{E}((\tau_i - u)_+)) \\ \le \sum_{i=0}^{n-1} (\bar\tau_i - \mathbb{E}(\bar\tau_i)) + \sum_{i=0}^{n-1} (\tau_i - u)_+,$$

which implies that

$$\mathbb{P}\left(\sum_{i=0}^{n-1} (\tau_i - \mathbb{E}(\tau_i)) \ge n/2\right) \le \mathbb{P}((\bar\tau_i - \mathbb{E}(\bar\tau_i)) \ge n/2) + \sum_{i=0}^{n-1} \mathbb{P}(\tau_i \ge u).$$

Next, by Markov inequality, we get that for any $u > 0$,

$$\mathbb{P}\left(\sum_{i=0}^{n-1}(\tau_i - \mathbb{E}(\tau_i)) \geq n/2\right) \leq 4n^{-1}\mathbb{E}((\tau_1 \wedge u)^2) + n\mathbb{P}(\tau_1 \geq u)$$

$$\leq 4n^{-1}\mathbb{E}(\tau_1^2 \mathbf{1}_{\tau_1 \leq u}) + 4n^{-1}u^2\mathbb{P}(\tau_1 \geq u) + n\mathbb{P}(\tau_1 \geq u).$$

We have

$$\mathbb{E}(\tau_1^2 \mathbf{1}_{\tau_1 \leq u}) \leq 2\int_0^u t\mathbb{P}(\tau_1 \geq t)\,\mathrm{d}t \leq 1 + 2p\Gamma(p)\int_1^u \frac{t}{[t]^p}\,\mathrm{d}t \leq 1 + \frac{2^{p+1}p\Gamma(p)}{2-p}u^{2-p}.$$

Therefore, choosing $u = n$, we get overall that, in the case $1 < p < 2$,

$$\mathbb{P}\left(\sum_{i=0}^{n-1}(\tau_i - \mathbb{E}(\tau_i)) \geq n/2\right) \leq 4n^{-1} + n^{1-p}p\Gamma(p)\left(5 + \frac{2^{p+3}}{2-p}\right).$$

Proceeding as before, the lower bound follows.

To end the proof of the lower bound, it remains to prove inequality (3.3). With this aim, setting

$$Z_0 = T_0 X_0 \quad \text{and} \quad Z_k = \tau_{k-1}X_{T_{k-1}} \quad \text{for } k \geq 1$$

we note that

$$\max_{0\leq k\leq n-1} \tau_k|X_{T_k}| \leq \max_{0\leq k\leq n}|Z_k|.$$

But for any $k \geq 1$, $Z_k = \sum_{i=0}^{k} Z_i - \sum_{i=0}^{k-1} Z_i$. Therefore

$$\max_{0\leq k\leq n}|Z_k| \leq 2\max_{0\leq k\leq n}\left|\sum_{i=0}^{k} Z_i\right|.$$

The above considerations imply that

$$\mathbb{P}\left(\max_{0\leq k\leq n-1}\tau_k|X_{T_k}| \geq 24x\right) \leq \mathbb{P}\left(\max_{0\leq k\leq n}\left|\sum_{i=0}^{k} Z_i\right| \geq 12x\right).$$

$(Z_k)_{k\geq 0}$ being a sequence of independent random variables, Etemadi's inequality entails that

$$\mathbb{P}\left(\max_{0\leq k\leq n}\left|\sum_{i=0}^{k} Z_i\right| \geq 12x\right) \leq 3\mathbb{P}\left(\left|\sum_{i=0}^{n} Z_i\right| \geq 4x\right).$$

Note now that

$$\sum_{i=0}^{n} Z_i = \sum_{k=0}^{T_0-1} X_0 + \sum_{i=1}^{n}(T_i - T_{i-1})X_{T_{i-1}} = \sum_{k=0}^{T_0-1} X_k + \sum_{i=1}^{n}\sum_{j=T_{i-1}}^{T_i-1} X_j = \sum_{k=0}^{T_n-1} X_k.$$

Therefore

$$\mathbb{P}\left(\left|\sum_{i=0}^{n} Z_i\right| \geq 4x\right) \leq \mathbb{P}\left(\left|\sum_{i=0}^{[n\mathbb{E}(\tau_1)]-1} X_i\right| \geq 2x\right) + \mathbb{P}(|S_{T_n} - S_{[n\mathbb{E}(\tau_1)]}| \geq 2x).$$

Inequality (3.3) follows from all the considerations above, together with the fact that

$$\mathbb{P}(|S_{T_n} - S_{[n\mathbb{E}(\tau_1)]}| \geq 2x) \leq 2\mathbb{P}\left(\max_{1\leq k\leq [n\mathbb{E}(\tau_1)]+1} |S_k| \geq x\right) + \mathbb{P}(T_n \geq 2[n\mathbb{E}(\tau_1)] + 1).$$

The proof is complete. □

To complete this section, it remains to show that the transition operator K of the Markov chain satisfies condition $\mathbf{H}_2$ for the semi norm $|\mathrm{d}f|$. With this aim, we first note that

$$K(f)(x) = (1-x)f(x) + x\nu(f).$$

So iterating, we get for any positive integer n,

$$K^n f(x) = (1-x)^n f(x) + \sum_{k=0}^{n-1} x(1-x)^k \nu(K^{n-1-k}(f)).$$

Therefore, we infer that

$$\begin{aligned}K^n f(x) = (1-x)^n (f(x) - \nu(f)) + \nu(K^{n-1}(f))\\ + \sum_{k=1}^{n-1} (1-x)^{n-k} (\nu(K^{k-1}(f)) - \nu(K^k(f))).\end{aligned}$$

It follows that

$$|\mathrm{d}K^n(f)| \leq 3|\mathrm{d}f| + \sum_{k=1}^{n-1} |\nu(K^{k-1}(f)) - \nu(K^k(f))|. \tag{3.7}$$

Setting $g_0 = f - f(0)$, note now that, for any positive integer k,

$$|\nu(K^{k-1}(f)) - \nu(K^k(f))| = |\nu(K^{k-1}(g_0)) - \nu(K^k(g_0))|.$$

Therefore

$$\sum_{k=1}^{n-1} |\nu(K^{k-1}(f)) - \nu(K^k(f))| \leq \sum_{k=1}^{n-1} \int_0^1 |K^{k-1}(g_0)(x) - K^k(g_0)(x)| \,\mathrm{d}\nu(x).$$

But $\sup_{x\in[0,1]} |g_0(x)| \leq |\mathrm{d}f|$. Hence

$$\begin{aligned}\int_0^1 |K^{k-1}(g_0)(x) - K^k(g_0)(x)| \,\mathrm{d}\nu(x) &= \int_0^1 |(\delta_x K^{k-1} - \delta_x K^k)(g_0)| \,\mathrm{d}\nu(x)\\ &\leq |\mathrm{d}f| \int_0^1 |\delta_x K^{k-1} - \delta_x K^k| \,\mathrm{d}\nu(x).\end{aligned} \tag{3.8}$$

From (3.7) and (3.8), to complete the proof of the fact that K satisfies $\mathbf{H}_2$, it remains to show that

$$\sum_{k\geq 1} \int_0^1 |\delta_x K^{k-1} - \delta_x K^k| \,\mathrm{d}\nu(x) < \infty. \tag{3.9}$$

Set $T(x) = 1 - x$. According to the computations leading to the first inequality on p. 76 of [11], we have, for any integer $k \geq 2$

$$|\delta_x K^{k-1} - \delta_x K^k| \leq 2(T(x))^{k-1} + \sum_{i=1}^{k-1} (1 - T(x))(T(x))^{i-1} |\nu K^{k-1-i} - \nu K^{k-i}|,$$

implying that

$$\int_0^1 |\delta_x K^{k-1} - \delta_x K^k| \, \mathrm{d}\nu(x) \leq 2\mathbb{E}_\nu(T^{k-1}) + \sum_{i=1}^{k-1} \mathbb{E}_\nu((1-T)T^{i-1}) |\nu K^{k-1-i} - \nu K^{k-i}|. \tag{3.10}$$

But, by taking into account (3.2), we get

$$\mathbb{E}_\nu(T^k) = (1+a) \int_0^1 (1-x)^k x^a \, \mathrm{d}x \leq k^{-(a+1)}(a+1)\Gamma(a+1) \tag{3.11}$$

and, for any integer $i \geq 2$,

$$\mathbb{E}_\nu((1-T)T^{i-1}) = (1+a) \int_0^1 (1-x)^{i-1} x^{a+1} \, \mathrm{d}x \leq (i-1)^{-(a+2)}(a+1)\Gamma(a+2). \tag{3.12}$$

We need now to give an upper bound of $|\nu K^j - \nu K^{j+1}|$ for any nonnegative integer j. With this aim, we first notice that

$$K^j(f) - K^{j+1}(f) = sK^j(f) - s\nu(K^j(f)),$$

where $s(x) = x$. Therefore setting $\mu = \frac{s(x)}{\nu(s)}\nu$, we have

$$\nu K^j - \nu K^{j+1} = \nu(s)(\mu K^j - \nu K^j).$$

Taking into account the relation (9.11) in [24], this gives

$$(\nu(s))^{-1}(\nu K^j - \nu K^{j+1}) = \sum_{\ell=1}^{j} a_\ell \nu Q^{j-\ell} + \mu Q^j - \nu Q^j, \tag{3.13}$$

where

$$Q(x, A) = K(x, A) - s(x)\nu(A) = T(x)\delta_x(A) \quad \text{and} \quad a_\ell = \mu K^{\ell-1}(s) - \nu K^{\ell-1}(s).$$

If we can prove that for any positive integer ℓ, a_ℓ is nonnegative, the relation (3.13) will imply that the signed measures $\nu K^j - \nu K^{j+1}$ of null mass can be rewritten as the differences of two positive measures with finite mass (the second one being equal to νQ^j), and therefore we will have

$$|\nu K^j - \nu K^{j+1}| \leq 2\nu(s)\nu Q^j(\mathbf{1}) = 2\nu(s)\mathbb{E}_\nu(T^j). \tag{3.14}$$

Hence, starting from (3.10) and taking into account (3.11), (3.12) and (3.14), we will get that for any integer $k \geq 2$,

$$\int_0^1 |\delta_x K^{k-1} - \delta_x K^k| \, \mathrm{d}\nu(x) \leq C_a \left(\frac{1}{k^{a+1}} + \sum_{i=1}^{k-1} \frac{1}{i^{a+2}} \times \frac{1}{(k-i)^{a+1}} \right) \leq \tilde{C}_a \frac{1}{k^{a+1}}, \tag{3.15}$$

provided that one can prove that, for any positive integer ℓ, a_ℓ is nonnegative. This can be proved by using (3.13) and the arguments developed in the proof of Lemma 9.3 in [24]. We complete the proof by noticing that (3.15) implies (3.9) since $a > 0$.

4. Concentration for Maps that Can be Modeled by Young Towers

In this section, in the specific setting of Young towers, we extend Theorem 1.1 to more general functionals. We refer to Sec. 3.2 for a brief description of the dynamical systems called "Young towers", and we keep the same notations as in this subsection.

Let T be a Young tower, and let τ be the return time to the basis. As already mentioned in Sec. 3.2, the decorrelation properties of the Markov chain associated with T are related to the return time τ. Namely, if τ has a weak moment of order $p > 1$, then the Markov chain satisfies $\mathbf{H}(p)$ for this p and the Hölder norm on the tower.

Proving quantitative estimates for the Markov chain or the dynamics is equivalent. In this section, we shall for simplicity formulate the results for the dynamics. Indeed the estimates of [16] that we shall use below are formulated in this context.

The class of functionals for which we will prove moderate deviations is the class of separately Lipschitz functions: these are the functions $\mathcal{K} = \mathcal{K}(z_0, \ldots, z_{n-1})$ such that, for all $i \in [0, n-1]$, there exists a constant L_i (the Lipschitz constant of $\mathcal{K}$ for the ith variable) with

$$|\mathcal{K}(z_0, \ldots, z_{i-1}, z_i, z_{i+1}, \ldots, z_{n-1}) - \mathcal{K}(z_0, \ldots, z_{i-1}, z_i', z_{i+1}, \ldots, z_{n-1})| \leq L_i d(z_i, z_i')$$

for all points $z_0, \ldots, z_{n-1}, z_i'$. We will write $\mathbb{E}\mathcal{K}$ for the average of $\mathcal{K}$ with respect to the natural measure along trajectories coming from the dynamics, i.e.

$$\mathbb{E}\mathcal{K} = \int \mathcal{K}(z, Tz, \ldots, T^{n-1}z) \, \mathrm{d}\pi(z).$$

The article [16] proves optimal moment estimates for $\mathcal{K} - \mathbb{E}\mathcal{K}$. We can prove moderate deviations for this quantity, extending in this context the results of Theorem 1.1 to more general functionals than additive functionals.

Theorem 4.1. *Consider a Young tower $T : Z \to Z$, for which the return time τ to the basis has a weak moment of order $p > 1$. Let $\mathcal{K}$ be a separately Lipschitz function, with Lipschitz constants L_i. Then*

- *If $p > 2$, then for all $x > 0$ one has*

$$\pi\{z : |\mathcal{K}(z, \ldots, T^{n-1}z) - \mathbb{E}\mathcal{K}| > x\} \leq \kappa \frac{\sum_{i=0}^{n-1} L_i^p}{x^p} + \kappa \exp\left(-\kappa^{-1} \frac{x^2}{\sum L_i^2}\right). \tag{4.1}$$

- *If $p = 2$, then for all $x > 0$ one has*

$$\begin{aligned} &\pi\{z : |\mathcal{K}(z, \ldots, T^{n-1}z) - \mathbb{E}\mathcal{K}| > x\} \\ &\quad \leq \kappa \frac{\sum_{i=0}^{n-1} L_i^2}{x^2} + \kappa \exp\left(-\kappa^{-1} \frac{x^2}{(\sum L_i^2) \cdot (1 + \log(\sum L_i) - \log(\sum L_i^2)^{1/2})}\right). \end{aligned} \tag{4.2}$$

- *If $p < 2$, then for all $x > 0$ one has*

$$\pi\{z : |\mathcal{K}(z, \ldots, T^{n-1}z) - \mathbb{E}\mathcal{K}| > x\} \leq \kappa \frac{\sum_{i=0}^{n-1} L_i^p}{x^p}. \tag{4.3}$$

In all these statements, κ is a positive constant that does not depend on $\mathcal{K}$ or n.

The case $p < 2$ is already proved in Theorem 1.9 of [16] and is included only for completeness. The logarithms in the $p = 2$ case are not surprising: this expression is homogeneous in the L_i (i.e. if one multiplies all the L_i by a constant, then the contribution of the logarithms does not vary), and it reduces to a multiple of $\log n$ when all the L_i are equal to 1. The same expression appears in the moment control when $p = 2$ in [16, Theorem 1.9].

To prove this theorem, we use the following deviation inequality for martingales (see [14], Corollary 3′, in which we keep separately the term corresponding to excess probabilities, as in Corollary 3 of the same paper).

Proposition 4.1. *Let $d_1, \ldots, d_k$ be a martingale difference sequence with respect to the nondecreasing σ-fields $\mathcal{F}_0, \ldots, \mathcal{F}_k$. Let $p \geq 2$. Set $\beta = p/(p+2)$ and $c_p^* = (1-\beta)^2/(2e^p)$. Then, for all $x > 0$,*

$$\begin{aligned} \mathbb{P}\left(\max_{1 \leq j \leq k} \left|\sum_{i=1}^{j} d_i\right| \geq x\right) &\leq \sum_{i=1}^{k} \mathbb{P}(|d_i| \geq \beta x) + \frac{2}{\beta^p x^p} \sum_{i=1}^{k} \|\mathbb{E}(|d_i|^p \mathbf{1}_{|d_i| \leq \beta x} \mid \mathcal{F}_{i-1})\|_\infty \\ &\quad + 2 \exp\left(-c_p^* \frac{x^2}{\sum \|\mathbb{E}(d_i^2 \mid \mathcal{F}_{i-1})\|_\infty}\right). \end{aligned}$$

As $\sum_{i=j}^{k} d_i = \sum_{i=1}^{k} d_i - \sum_{i=1}^{j-1} d_k$, a similar result follows for reverse martingale difference sequences, by applying the previous result to the martingale d_{k-i}:

Corollary 4.1. *Let $d_1, \ldots, d_k$ be a reverse martingale difference sequence w.r.t. the nonincreasing σ-fields $\mathcal{F}_1, \ldots, \mathcal{F}_{k+1}$ (so $\mathbb{E}(d_i \mid \mathcal{F}_{i+1}) = 0$ and d_i is $\mathcal{F}_i$-measurable).*

Let $p \geq 2$. *Set* $\tilde{\beta} = p/(p+2)$ *and* $\tilde{c}_p^* = (1-\beta)^2/(8e^p)$. *Then, for all* $x > 0$,

$$\mathbb{P}\left(\max_{1\leq j\leq k}\left|\sum_{i=1}^{j} d_i\right| \geq x\right) \leq \sum_{i=1}^{k}\mathbb{P}(|d_i| \geq \tilde{\beta}x) + \frac{2^{p+1}}{\tilde{\beta}^p x^p}\sum_{1}^{n}\|\mathbb{E}(|d_i|^p \mathbf{1}_{|d_i|\leq\tilde{\beta}x} \mid \mathcal{F}_{i+1})\|_\infty$$
$$+ 4\exp\left(-\tilde{c}_p^*\frac{x^2}{\sum\|\mathbb{E}(d_i^2 \mid \mathcal{F}_{i+1})\|_\infty}\right).$$

We will use the following consequence for reverse martingales having a conditional weak moment of order p, as follows (the same corollary holds as well for martingales). This is a finer version of Corollary 3′ in [14], replacing the strong norm there with a weak norm.

Corollary 4.2. *Let* $d_1, \ldots, d_k$ *be a reverse martingale difference sequence w.r.t. the nonincreasing* σ*-fields* $\mathcal{F}_1, \ldots, \mathcal{F}_{k+1}$ *(so* $\mathbb{E}(d_i \mid \mathcal{F}_{i+1}) = 0$ *and* d_i *is* $\mathcal{F}_i$*-measurable). Let* $p \geq 2$. *Assume that, for all* i, d_i *has a conditional weak moment of order* p *bounded by a constant* M_i, *i.e.* $\mathbb{P}(|d_i| \geq x \mid \mathcal{F}_{i+1}) \leq M_i^p/x^p$. *Then there exists a constant* C_p *only depending on* p *such that, for all* $x > 0$,

$$\mathbb{P}\left(\max_{1\leq j\leq k}\left|\sum_{i=1}^{j} d_i\right| \geq x\right) \leq \frac{C_p}{x^p}\sum_{i=1}^{k} M_i^p + 4\exp\left(-C_p^{-1}\frac{x^2}{\sum\|\mathbb{E}(d_i^2 \mid \mathcal{F}_{i+1})\|_\infty}\right).$$

Proof. We apply Corollary 4.1 with any $q > p$, for instance $q = p+1$. Since $\mathbb{P}(|d_i| \geq \tilde{\beta}x) \leq M_i^p/(\tilde{\beta}x)^p$, the first term in the upper bound of this lemma is bounded as desired. The last term is also bounded as desired. It remains to handle the terms involving $x^{-q}\|\mathbb{E}(|d_i|^q \mathbf{1}_{|d_i|\leq\tilde{\beta}x} \mid \mathcal{F}_{i+1})\|_\infty$. We have

$$\begin{aligned}
x^{-q}\mathbb{E}(|d_i|^q \mathbf{1}_{|d_i|\leq\tilde{\beta}x} \mid \mathcal{F}_{i+1}) &\leq x^{-q} q\int_{u=0}^{\tilde{\beta}x} u^{q-1}\mathbb{P}(|d_i| \geq u \mid \mathcal{F}_{i+1})\,\mathrm{d}u \\
&\leq x^{-q} q M_i^p \int_{u=0}^{\tilde{\beta}x} u^{q-1}u^{-p}\,\mathrm{d}u = x^{-q} q M_i^p\frac{(\tilde{\beta}x)^{q-p}}{q-p} \\
&\leq CM_i^p/x^p.
\end{aligned}$$

Summing these terms over i gives a bound as in the statement of the corollary. □

We can now start the proof of Theorem 4.1. Assume that τ has a weak moment of order $p \geq 2$. Starting from a separately Lipschitz function $\mathcal{K}$, Chazottes and Gouëzel consider in [5] a sequence $(d_k)_{k\geq 0}$ of reverse martingale differences with respect to the filtration $\mathcal{F}_k$ of functions depending only on coordinates $x_k, x_{k+1}, \ldots$, given by

$$d_k = \mathbb{E}(\mathcal{K} \mid \mathcal{F}_k) - \mathbb{E}(\mathcal{K} \mid \mathcal{F}_{k+1}).$$

On p. 869 in [5], it is proved that, if $p > 2$, then

$$\mathbb{E}(d_k^2 \mid \mathcal{F}_{k+1}) \leq \sum_{j\leq k} c_{k-j}^{(0)} L_j^2,$$

where $c_k^{(0)}$ denotes a generic summable sequence that does not depend on $\mathcal{K}$ or n. Therefore,

$$\sum_k \|\mathbb{E}(d_k^2 \mid \mathcal{F}_{k+1})\|_\infty \leq C \sum L_j^2. \tag{4.4}$$

Moreover, if $p = 2$, Gouëzel and Melbourne (see Sec. 4.2 in [16]) show that

$$\sum_k \|\mathbb{E}(d_k^2 \mid \mathcal{F}_{k+1})\|_\infty \leq C \left(\sum L_i^2\right) \cdot \left[1 + \log\left(\sum L_i\right) - \log\left(\sum L_i^2\right)^{1/2}\right]. \tag{4.5}$$

Now we use the following modification of [5, Lemma 6.2]:

Lemma 4.1. *For all $t > 0$ and all integer k,*

$$\mathbb{P}(|d_k| \geq t \mid \mathcal{F}_{k+1}) \leq Ct^{-p} \sum_{j=0}^{k} L_j^p c_{k-j}^{(0)} + Ct^{-p} \sup_{h>0} \left(h^{-1} \sum_{j=k-h+1}^{k} L_j\right)^p.$$

Proof. We just follow the lines of the proof of Lemma 6.2 in [5] up to (6.1). Note that this paper requires the condition $p > 2$ (for the validity of (4.8) there), but Lemma 4.2 in [16] replaces this inequality for $p = 2$.

For the first sum we have as in [5]

$$\sum_{A_1(z_\alpha) \geq t/2} g(z_\alpha) \leq Ct^{-p} \sum_{j \leq k} L_j^p c_{k-j}^{(0)}.$$

On the other hand, if h denotes the smallest ℓ such that $\sum_{j=k-\ell+1}^{k} L_j \geq t/2$, then

$$\sum_{A_2(z_\alpha) \geq t/2} g(z_\alpha) \leq C\pi(\tau \geq h) \leq Ch^{-p} \leq Ct^{-p} \sup_{h>0} \left(h^{-1} \sum_{j=k-h+1}^{k} L_j\right)^p. \qquad \square$$

Proof of Theorem 4.1 when $p \geq 2$. We apply Corollary 4.2 to d_k, with

$$M_k^p = C \sum_{j=0}^{k} L_j^p c_{k-j}^{(0)} + C \sup_{h>0} \left(h^{-1} \sum_{j=k-h+1}^{k} L_j\right)^p \tag{4.6}$$

thanks to Lemma 4.1. As $c_k^{(0)}$ is summable, the sum over k of the first term is bounded by $C' \sum L_j^p$. An application of the Hardy–Littlewood maximal inequality in ℓ^p gives

$$\sum_{k \geq 0} \sup_{h>0} \left(h^{-1} \sum_{j=k-h+1}^{k} L_j\right)^p \leq C \sum_j L_j^p.$$

Hence, the sum over k of the second term in (4.6) is also bounded by $C \sum_j L_j^p$. This shows that the first term in Corollary 4.2 gives rise to a bound $C \sum L_i^p / x^p$.

Finally, the second term in Lemma 4.2 gives rise to the exponential error term in the statement of the theorem, thanks to (4.4) when $p > 2$ and to (4.5) when $p = 2$. $\square$

Remark 4.1. Assume that $p > 2$ and for any i, $L_i \le 1$. In this case, integrating inequality (4.1) leads to

$$\|\mathcal{K} - \mathbb{E}\mathcal{K}\|_{\pi,2(p-1)}^{2(p-1)} \le Cn^{p-1},$$

where C is a positive constant not depending on $\mathcal{K}$ nor n. However, in the case of general L_i, we do not recover for this moment the bound $C(\sum L_i^2)^{p-1}$ proved in [16, Theorem 1.9] (consider for instance the case $L_0 = 1$ and $L_1 = \cdots = L_{n-1} = 1/\sqrt{n}$). This moment bound, combined with Markov inequality, gives

$$\pi\{|\mathcal{K} - \mathbb{E}\mathcal{K}| > x\} \le \kappa \frac{\left(\sum L_i^2\right)^{p-1}}{x^{2p-2}}.$$

For the case where all L_i are of the order of 1, this bound is worse than the bound of Theorem 4.1. However, surprisingly, it can be better when the L_i vary a lot, for instance when $L_0 = 1$ and $L_1 = \cdots = L_{n-1} = 1/\sqrt{n}$, and $x = n^{1/4}$.

References

1. J. Aaronson and M. Denker, A local limit theorem for stationary processes in the domain of attraction of a normal distribution, *Asymptotic Methods in Probability and Statistics with Applications* (St. Petersburg, 1998), Stat. Ind. Technol. (Birkhäuser, 2001), pp. 215–223.
2. J. Aaronson and M. Denker, Local limit theorems for partial sums of stationary sequences generated by Gibbs–Markov maps, *Stoch. Dyn.* **1** (2001) 193–237.
3. R. Adamczak, A tail inequality for suprema of unbounded empirical processes with applications to Markov chains, *Electron. J. Probab.* **13** (2008) 1000–1034.
4. R. C. Bradley, *Introduction to Strong Mixing Conditions*, Vol. 3 (Kendrick Press, 2007).
5. J.-R. Chazottes and S. Gouëzel, Optimal concentration inequalities for dynamical systems, *Comm. Math. Phys.* **316** (2012) 843–889.
6. Y. A. Davydov, Mixing conditions for Markov chains, *Teor. Verojatnost. i Primenen.* **18** (1973) 321–338.
7. J. Dedecker, S. Gouëzel and F. Merlevède, Some almost sure results for unbounded functions of intermittent maps and their associated Markov chains, *Ann. Inst. Henri Poincaré Probab. Stat.* **46** (2010) 796–821.
8. J. Dedecker and F. Merlevède, Convergence rates in the law of large numbers for Banach-valued dependent variables, *Teor. Veroyatn. Primen.* **52** (2007) 562–587 [translated in *Theory Probab. Appl.* **52** (2007) 416–438].
9. J. Dedecker and F. Merlevède, Moment bounds for dependent sequences in smooth Banach spaces, *Stochastic Process. Appl.* **125** (2015) 3401–3429.
10. J. Dedecker and F. Merlevède, A deviation bound for α-dependent sequences with applications to intermittent maps, *Stoch. Dyn.* **17** (2017) 1750005.
11. P. Doukhan, P. Massart and E. Rio, The functional central limit theorem for strongly mixing processes, *Ann. Inst. H. Poincaré Probab. Statist.* **30** (1994) 63–82.

12. W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. II (John Wiley & Sons, 1966).
13. D. A. Freedman, On tail probabilities for martingales, *Ann. Probab.* **3** (1975) 100–118.
14. D. H. Fuk, Certain probabilistic inequalities for martingales, *Sibirsk. Mat. Ž.* **14** (1973) 185–193.
15. S. Gouëzel, Central limit theorem and stable laws for intermittent maps, *Probab. Theory Relat. Fields* **128** (2004) 82–122.
16. S. Gouëzel and I. Melbourne, Moment bounds and concentration inequalities for slowly mixing dynamical systems, *Electron. J. Probab.* **19** (2014) No. 93, 30 pages.
17. P. Lezaud, Chernoff-type bound for finite Markov chains, *Ann. Appl. Probab.* **8** (1998) 849–867.
18. I. Melbourne, Large and moderate deviations for slowly mixing dynamical systems, *Proc. Amer. Math. Soc.* **137** (2009) 1735–1741.
19. F. Merlevède and M. Peligrad, Rosenthal-type inequalities for the maximum of partial sums of stationary processes and examples, *Ann. Probab.* **41** (2013) 914–960.
20. F. Merlevède, M. Peligrad and E. Rio, A Bernstein type inequality and moderate deviations for weakly dependent sequences, *Probab. Theory Relat. Fields* **151** (2011) 435–474.
21. S. V. Nagaev, Large deviations of sums of independent random variables, *Ann. Probab.* **7** (1979) 745–789.
22. E. Nummelin, *General Irreducible Markov Chains and Nonnegative Operators*, Cambridge Tracts in Mathematics, Vol. 83 (Cambridge Univ. Press, 1984).
23. M. Peligrad, S. Utev and W. B. Wu, A maximal $\mathbb{L}^p$-inequality for stationary sequences and its applications, *Proc. Amer. Math. Soc.* **135** (2007) 541–550.
24. E. Rio, *Théorie Asymptotique des Processus Aléatoires Faiblement Dépendants*, Mathématiques et Applications, Vol. 31 (Springer-Verlag, 2000).
25. M. Rosenblatt, A central limit theorem and a strong mixing condition, *Proc. Nat. Acad. Sci. USA* **42** (1956) 43–47.