Deviation inequalities for separately Lipschitz functionals of composition of random functions

Jérôme Dedecker,[*] Paul Doukhan[†] and   Xiequan Fan[‡§]

Abstract

We consider a class of non-homogeneous Markov chains, that contains many natural examples. Next, using martingale methods, we establish some deviation and moment inequalities for separately Lipschitz functions of such a chain, under moment conditions on some dominating random variables.

Keywords: Non-homogeneous Markov chains, martingales, exponential inequalities, moment inequalities.

Mathematics Subject Classification (2010): 60G42, 60J05, 60E15.

1 Introduction

Concentration inequalities are essential tools for ensuring the validity of many statistical procedures; let us cite for instance [18] for classification problems, [21] for model selection, and [3] for high dimensional procedures (see also [2] and [33] in a dependent framework).

These inequalities are not easy to prove in a dependent context; up to now this has been done under quite restrictive assumptions, and mainly for bounded functionals of the variables in a stationary/homogeneous context. As a non exhaustive list, let us quote [31], [27], [12], [1], [10] and [11]. Among these references, the case of unbounded functionals has been investigated in [1] for geometrically ergodic Markov chains, and in [10] for iterated random functions satisfying a mean-contraction condition (see condition (2.8) below, with $F_n = F$).

In the paper [10], the authors obtained very precise inequalities for Lipschitz functionals of the chain, by assuming moreover a Lipschitz condition on the function that generates the chain (see condition (2.9) below). However, this last condition is in fact quite restrictive, and does not hold for many natural models satisfying the mean-contraction property.

In the present paper we enlarge the class of Markov chains studied in [10], by considering non-homogeneous Markov chains obtained through composition of random functions (see the recursive mechanism (2.7) below), and by making no extra assumptions than the (uniform) mean-contraction (2.8). As in [10], we shall use the decomposition of the functional of the chain in terms of martingale differences, as first introduced by Yurinskii [32]. This method is well adapted to the non-homogeneous Markov context, because it is intrinsically a non-stationary method. Following this approach, we obtain deviation and moment inequalities for separately Lipschitz functionals of the chain that are driven by the distribution of some dominating random variables.

The present article was conceived within the general framework of non stationary time series, which is now widely discussed in the context global warming [6]. Besides temperatures or

---

[*]Université Paris Descartes, Sorbonne Paris Cité, Laboratoire MAP5 and CNRS UMR 8145, 75016 Paris, France. E-mail: jerome.dedecker@parisdescartes.fr

[†]AGM UMR 8088, Université Paris Seine, UCP site Saint-Martin, 2 Bd. Adolphe Chauvin 95000 Cergy-Pontoise, France, and CIMFAV, Universidad de Valparaiso, Chile. Email: doukhan@u-cergy.fr

[‡]Center for Applied Mathematics, Tianjin University, Tianjin, China. E-mail: fanxiequan@hotmail.com

[§]All authors contributed equally to this work

ozone concentration, most of the real life phenomena present trends and periodicities. A first excellent view of those questions may be found in [5], but this is a linear view of time series analysis. It appears important to consider cases where the dynamic of the models itself is non time-homogeneous. For instance [8] and [4] provide different views for a more relevant dynamical approach including local stationarity and non-periodic features. The present paper aims at developing reasonable concentration and probability inequalities for non-homogeneous Markov chains able to model some of the above features.

Before going into further details, let us give a simple class of examples to which our results apply. We consider a generalized $\mathbb{R}^k$-valued auto-regressive processes

$$X_n = A_n X_{n-1} + B_n, \tag{1.1}$$

where A_n is a random $k \times k$ matrix and B_n is an $\mathbb{R}^k$-valued random variable. Here $\varepsilon_n = (A_n, B_n)$ are independent random variables, and $(\varepsilon_n)_{n\geq 2}$ is independent of the initial random variable X_1. Let $|\cdot|$ be a norm on $\mathbb{R}^k$. Then, the Markov chain X_n satisfies the mean contraction condition (2.8) for the norm $|\cdot|$ as soon as

$$\sup_{n\geq 2}\mathbb{E}[|A_n|] \leq \rho \quad \text{for some } \rho < 1, \tag{1.2}$$

where as usual $|A_n| = \sup_{|x|=1}|A_n x|$.

Model (1.1) contains a lot of natural examples (see for instance [9], Sections 2.2 and 2.3), but does not fit within the framework of [10]; moreover, it has no reason to be mixing in the sense of Rosenblatt [30] without further assumptions on the distribution of $(\varepsilon_n)_{n\geq 2}$. Recall that the chain X_n is non-homegeneous, since we do not assume here that the variables ε_n are iid.

Let now $f : (\mathbb{R}^k)^n \mapsto \mathbb{R}$ be a separately Lipschitz function, such that

$$|f(x_1, x_2, \ldots, x_n) - f(x_1', x_2', \ldots, x_n')| \leq |x_1 - x_1'| + \cdots + |x_n - x_n'|, \tag{1.3}$$

and let also $S_n = f(X_1, \ldots, X_n) - \mathbb{E}[f(X_1, \ldots, X_n)]$.

For simplicity, let us consider the case where the chain starts at $X_1 = 0$. Assuming that $\|A_n\|_p^p := \mathbb{E}[|A_n|^p] < \infty$ and $\|B_n\|_p^p := \mathbb{E}[|B_n|^p] < \infty$ for any $n \geq 2$ and some $p > 1$, we infer from (2.11) (control of the $\mathbb{L}^p$-norm of the dominating variables $H_k(X_{k-1}, \varepsilon_k)$ defined in (2.10)) and Propositions 5.1 and 5.3 that

$$\|S_n\|_p^p \leq C(p,\rho)\left(\sum_{k=3}^n \|X_{k-1}\|_p^2 \|A_k\|_p^2 + \sum_{k=2}^n \|B_k\|_p^2\right)^2 \quad \text{if } p \geq 2,$$

and

$$\|S_n\|_p^p \leq C(p,\rho)\left(\sum_{k=3}^n \|X_{k-1}\|_p^p \|A_k\|_p^p + \sum_{k=2}^n \|B_k\|_p^p\right) \quad \text{if } p \in (1,2),$$

for some constant $C(p,\rho)$ depending only on (p,ρ). These inequalities are satisfactory, because if $X_n = B_n$ for $n \geq 2$ (case $A_n = 0$), we recover for $p \geq 2$ the usual Marcinkiewicz-Zygmund inequalities (see [28]) for $\mathbb{L}_p$-norms of sums of independent random variables, and for $p \in (1,2)$ the usual von Bahr-Esseen inequalities (see [26]). Note that, under the stronger condition than (1.2): $\sup_{n\geq 2}\|A_n\|_p \leq \rho$ ($\mathbb{L}^p$-contraction), and if $\sup_{n\geq 2}\|B_n\|_p < \infty$, we obtain that $\|S_n\|_p = O(\sqrt{n})$ if $p \geq 2$ and $\|S_n\|_p = n^{1/p}$ if $p \in (1,2)$, which is exactly what we could expect for $\mathbb{L}^p$-norms of partial sums in a quasi-stationary regime.

Under more restrictive conditions on $(A_n, B_n)_{n\geq 2}$, one can also obtain some semi-exponential bounds for the deviation of S_n. For the sake of simplicity, let us assume that $\sup_{n\geq 2}\|A_n\|_\infty \leq \rho$ (uniform contraction) and that there exist $\kappa > 0, \alpha \in (0,1)$ such that

$$\sup_{k\geq 2}\mathbb{E}\left[\exp\left\{\kappa|B_k|^{\frac{2\alpha}{1-\alpha}}\right\}\right] < \infty. \tag{1.4}$$

It is then easy to see that dominating variables $H_k(X_{k-1}, \varepsilon_k)$ defined in (2.10) also satisfy the uniform bound (1.4) (for the same α and a different κ, say κ'). Hence, it follows from Proposition 4.2 that

$$\mathbb{P}\left(|S_n| \geq nx\right) \leq C(x) \exp\left\{-Kx^{2\alpha}n^\alpha\right\}, \tag{1.5}$$

for any $x > 0$, where the positive constant K depends only on (ρ, α, κ'), and

$$C(x) = 2 + c(\alpha, \rho, \kappa') \left(\frac{1}{x^{2\alpha}} + \frac{1}{x^2}\right).$$

In particular, we obtain from (1.5) the following moderate deviation behavior: for any $\delta \in (1/2, 1]$, there exist $a > 0$ such that

$$\mathbb{P}\left(|S_n| \geq n^\delta\right) = O\left(\exp\left\{-an^{\alpha(2\delta-1)}\right\}\right). \tag{1.6}$$

Note that, for $\delta = 1$, this is in accordance with the best possible rate for large deviation of sums of martingale differences (see Theorem 2.1 in [15]).

2 Composition of random functions

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let $(\mathcal{X}, d)$ and $(\mathcal{Y}, \delta)$ be two complete separable metric spaces. Let $(\varepsilon_i)_{i \geq 2}$ be a sequence of independent $\mathcal{Y}$-valued random variables. Let X_1 be a $\mathcal{X}$-valued random variable independent of $(\varepsilon_i)_{i \geq 2}$. We consider the Markov chain $(X_i)_{i \geq 1}$ such that

$$X_n = F_n(X_{n-1}, \varepsilon_n), \quad n \geq 2, \tag{2.7}$$

where $F_n : \mathcal{X} \times \mathcal{Y} \to \mathcal{X}$ is such that

$$\mathbb{E}\left[d\left(F_n(x, \varepsilon_n), F_n(x', \varepsilon_n)\right)\right] \leq \rho \, d(x, x') \tag{2.8}$$

for some constant $\rho \in [0, 1)$ not depending on n.

In the paper [10], the authors studied a class of homogeneous Markov chains (that is, with $F_n = F$ and $(\varepsilon_i)_{i \geq 2}$ a sequence of i.i.d. random variables) satisfying (2.8) and the condition

$$d(F(x, y), F(x, y')) \leq C \, \delta(y, y') \tag{2.9}$$

for some positive constant C. Under this additional constraint, they obtained very precise upper bounds for the deviation of separately Lipschitz functionals of the chain; this is possible, because in that case, the martingales differences M_k from McDiarmid's decomposition are bounded by a function of ε_k, which is then independent of the past σ-field of the chain.

However, condition (2.9) is quite restrictive, and is not satisfied for many natural models (a short list of such models is presented below). In the present paper, we shall not assume that (2.9) is satisfied. In this more general setting, the dominating random variables are

$$H_k(X_{k-1}, \varepsilon_k) \quad \text{where} \quad H_k(x, y) = \int d(F_k(x, y), F_k(x, y'))P_{\varepsilon_k}(dy') \tag{2.10}$$

(see Proposition 3.1 below). The main difference with [10] is that these dominating random variables are no longer independent from the past σ-field of the chain. Hence, the deviations bounds that we obtain are not as precise as in [10], but apply to a much larger class of (non homogeneous) Markov chains.

Remark 2.1. Note that if (2.8) holds for the distance d, then, for any $\alpha \in (0, 1]$, it also hold for the distance $d_\alpha(x, y) = (d(x, y))^\alpha$ with ρ^α instead of ρ. This is elementary, but nevertheless important: it means that we can also obtain concentration inequalities for separately Lipschitz functions with respect to d_α by controlling the behavior of $H_{k,\alpha}(X_{k-1}, \varepsilon_k)$ (whose definition is as in (2.10) for the distance d_α). Note that separately Lipschitz functions with respect to d_α are less and less regular as α approaches 0.

Remark 2.2. Let us quote an error in the paper [10]. The inequality (1.4) of that paper gives an upper bound for the quantity $\mathbb{E}[H(d(X_n, x_0))]$ when H is any increasing function from $\mathbb{R}^+$ to $\mathbb{R}^+$. However this upper bound is not true in general under the assumption (1.2) of [10] (which is similar to our assumption (2.8)), but it holds under the much more restrictive assumption $d(F(x, y), F(x', y)) \leq \rho\, d(x, x')$. The error comes from the fact that the first version of the paper [10] was written under this more restrictive assumption. Note that this wrong inequality was not used at any points in the proofs of the main results in [10], but only in Items 4 of Remarks 3.1 and 3.2 (which are therefore not correct).

2.1 Examples

In this subsection, we give a non exhaustive list of models satisfying condition (2.8), and we show how to control the moments of the dominating variables $H_k(X_{k-1}, \varepsilon_k)$ defined by (2.10). For the sake of simplicity, we shall only deal with the moments of order p of $H_k(X_{k-1}, \varepsilon_k)$, but similar computations may be done for exponential moments. We refer to [13] and [8] for more examples.

- **ARCH-type models**. For which

$$F_n(x, y) = M_{\theta_n}(x, y), \qquad \text{with} \quad M_\theta(x, y) = \sqrt{a^2 x^2 + b^2} \cdot y, \quad \theta = (a, b).$$

  Thus the non stationarity appears simply from changes in the parameter $\theta_n = (a_n, b_n)$. In that case (2.8) is satisfied for $d(x, x') = |x - x'|$ provided $\sup_{n \geq 2} |a_n| \mathbb{E}[|\varepsilon_n|] = \rho$ for some $\rho < 1$.

  For these models

$$H_k(X_{k-1}, \varepsilon_k) = \sqrt{a_k^2 X_{k-1}^2 + b_k^2} \int |\varepsilon_k - y| P_{\varepsilon_k}(dy),$$

  and the moments of order p of $H_k(X_{k-1}, \varepsilon_k)$ satisfy

$$\|H_k(X_{k-1}, \varepsilon_k)\|_p^p \leq 2^{p-1} \mathbb{E}\left[\left(a_k^2 X_{k-1}^2 + b_k^2\right)^{p/2}\right] \|\varepsilon_k\|_p^p, \quad p \geq 1.$$

  Those models are easy to extend in an $\mathbb{R}^k$-valued framework. For instance, one can consider $X_n = A_n(X_{n-1}) \varepsilon_n$, where $A_n(x)$ is a $k \times k$ matrix and ε_n are $\mathbb{R}^k$-valued random variables. Let $|\cdot|$ be a norm on $\mathbb{R}^k$, and $|A| = \sup_{|x|=1} |A_n x|$ be the associated matrix norm. Now, if $|A_n(x) - A_n(x')| \leq a_n |x - x'|$, then the condition (2.8) is satisfied as soon as $\sup_{n \geq 2} a_n \mathbb{E}[|\varepsilon_n|] = \rho$ for some $\rho < 1$.

- **Switching models.** Many analogous models can be provided with a switching, e.g. for the first ARCH-model, such a parametric model is given with $\mathcal{X} = \mathbb{R}$, $\mathcal{Y} = \mathbb{R} \times \{0, 1\}$, a parameter $\theta = (a, b, a', b') \in \mathbb{R}^4$ and

$$M_\theta(x, y) = y_2 \sqrt{a^2 x^2 + b^2} \cdot y_1 + (1 - y_2) \sqrt{a'^2 x^2 + b'^2} \cdot y_1.$$

  Here $(\varepsilon_n)_{n \geq 2}$ is a sequence of independent random variables with values in $\mathbb{R} \times \{0, 1\}$. Using the notation $\varepsilon_n = (\varepsilon_n^{(1)}, \varepsilon_n^{(2)})$, we see that condition (2.8) is satisfied as soon as

$$\sup_{n \geq 2} \left((1 - \mathbb{E}[\varepsilon_n^{(2)}])|a_n| + \mathbb{E}[\varepsilon_n^{(2)}]|a_n'|\right) \mathbb{E}[|\varepsilon_n^{(1)}|] = \rho$$

  for some $\rho < 1$.

  Now, similar computations as for the first example lead to

$$\|H_k(X_{k-1}, \varepsilon_k)\|_p^p \leq 4^{p-1} \mathbb{E}\left[\left(a_k^2 X_{k-1}^2 + b_k^2\right)^{p/2}\right] \left\|\varepsilon_k^{(1)} \varepsilon_k^{(2)}\right\|_p^p$$
$$+ 4^{p-1} \mathbb{E}\left[\left(a_k'^2 X_{k-1}^2 + b_k'^2\right)^{p/2}\right] \left\|\varepsilon_k^{(1)}(1 - \varepsilon_k^{(2)})\right\|_p^p, \quad p \geq 1.$$

4

- **Generalized $\mathbb{R}^k$-valued auto-regressive processes**. We consider here the Model (1.1) presented in the introduction. Recall that A_n is a random $k \times k$ matrix and B_n is an $\mathbb{R}^k$-valued random variable. Here $\varepsilon_n = (A_n, B_n)$ are independent random variables, and $(\varepsilon_n)_{n \geq 2}$ is independent of the initial random variable X_1. Model (1.1) is a composition of random functions as in (2.7), with

$$F_n(x, y) = F(x, y) = y_1 x + y_2 \, .$$

Let $| \cdot |$ be a norm on $\mathbb{R}^k$, and let as usual $|A_n| = \sup_{|x|=1} |A_n x|$. The condition (2.8) is satisfied as soon as (1.2) holds.

For these models

$$H_k(X_{k-1}, \varepsilon_k) \leq |X_{k-1}| \int |A_k - y| P_{A_k}(dy) + \int |B_k - y| P_{B_k}(dy) \, ,$$

and the moments of order p of $H_k(X_{k-1}, \varepsilon_k)$ satisfy

$$\|H_k(X_{k-1}, \varepsilon_k)\|_p^p \leq 4^{p-1} \mathbb{E}\left[|X_{k-1}|^p\right] \mathbb{E}\left[|A_k|^p\right] + 4^{p-1} \mathbb{E}\left[|B_k|^p\right], \quad p \geq 1. \tag{2.11}$$

- **INAR(1) type models.** In this case, let $y = (y_0, y_1, y_2, \ldots, y_p, \ldots) \in \mathcal{Y} = \mathbb{N}^{\mathbb{N}}$ and $\varepsilon = (\varepsilon^{(0)}, \varepsilon^{(1)}, \ldots, \varepsilon^{(p)}, \ldots)$, where $(\varepsilon^{(1)}, \ldots, \varepsilon^{(p)}, \ldots)$ is a sequence of i.i.d. integer valued random variables. The function F is then given by

$$F(x, y) = y_0 + \mathbf{1}_{\{x \neq 0\}} \sum_{k=1}^{x} y_k \, .$$

Here, $(\varepsilon_n)_{n \geq 2}$ is an i.i.d. sequence distributed as ε. It is then easy to see that (2.8) is satisfied provided $\rho = \mathbb{E}[\varepsilon^{(1)}] < 1$.

We shall now give some hints to control the moments of the dominating random variables $H_k(X_{k-1}, \varepsilon_k)$. Let $\tilde{\varepsilon}_k$ be distributed as ε_k and independent of (ε_k, X_{k-1}). We then have that

$$H_k(X_{k-1}, \varepsilon_k) = \mathbb{E}\big[d(F_k(X_{k-1}, \varepsilon_k), F_k(X_{k-1}, \tilde{\varepsilon}_k))\big| X_{k-1}, \varepsilon_k\big].$$

For the INAR(1) model, we have

$$H_k(X_{k-1}, \varepsilon_k) = \mathbb{E}\left[\left|(\varepsilon_k^{(0)} - \tilde{\varepsilon}_k^{(0)}) + \mathbf{1}_{\{X_{k-1} \neq 0\}} \sum_{i=1}^{X_{k-1}} (\varepsilon_k^{(i)} - \tilde{\varepsilon}_k^{(i)})\right| \, \Bigg| X_{k-1}, \varepsilon_k\right]$$

By contraction, we get that

$$\|H_k(X_{k-1}, \varepsilon_k)\|_p^p \leq \left\|(\varepsilon_k^{(0)} - \tilde{\varepsilon}_k^{(0)}) + \mathbf{1}_{\{X_{k-1} \neq 0\}} \sum_{i=1}^{X_{k-1}} (\varepsilon_k^{(i)} - \tilde{\varepsilon}_k^{(i)})\right\|_p^p, \quad p \geq 1.$$

For $p \geq 2$, applying the Marcinkiewicz-Zygmund inequality given in [28], we get that

$$\|H_k(X_{k-1}, \varepsilon_k)\|_p^p \leq 2^{(p-2)/2} \|\varepsilon^{(0)} - \tilde{\varepsilon}^{(0)}\|_p^p + (p-1)^{p/2} 2^{(p-2)/2} \mathbb{E}\left[X_{k-1}^{p/2}\right] \|\varepsilon^{(1)} - \tilde{\varepsilon}^{(1)}\|_p^p \, .$$

For $p \in (1, 2)$, applying the von-Bahr Essen inequality given in [26], we get that

$$\|H_k(X_{k-1}, \varepsilon_k)\|_p^p \leq \|\varepsilon^{(0)} - \tilde{\varepsilon}^{(0)}\|_p^p + 2^{2-p} \mathbb{E}\left[X_{k-1}\right] \|\varepsilon^{(1)} - \tilde{\varepsilon}^{(1)}\|_p^p \, .$$

Note that non-stationary variants of this model can be obtained by considering independent (but non i.i.d.) ε_n's, with the constraint: $\sup_{n \geq 2} \mathbb{E}[\varepsilon_n^{(1)}] < 1$.

- **GLM-Poisson models.** Besides the standard ARCH-models the simplest case is that of Poisson ARCH–models, where $(\varepsilon_n)_{n\geq 2}$ is a sequence of i.i.d. unit Poisson processes. Consider a sequence of functions $f_n : \mathbb{N} \to \mathbb{R}^+$ and set

$$F_n(x, y) = y(f_n(x)),$$

where $y : \mathbb{R}^+ \to \mathbb{N}$ denotes a function. In that case, the condition (2.8) is satisfied if $|f_n(x) - f_n(x')| \leq \rho|x - x'|$ for any $n \geq 2$, any $x, x' \in \mathbb{N}$, and some $\rho < 1$.

For these models,

$$H_k(X_{k-1}, \varepsilon_k) = \mathbb{E}\big[|\varepsilon(f_k(X_{k-1})) - \varepsilon_k(f_k(X_{k-1}))|\big|\varepsilon_k, X_{k-1}\big],$$

where ε is a unit Poisson process independent of (X_{k-1}, ε_k). By contraction, we get that

$$\|H_k(X_{k-1}, \varepsilon_k)\|_p^p \leq 2^{p-1}\|\varepsilon_k(f_k(X_{k-1}))\|_p^p = 2^{p-1}\mathbb{E}\left[Q_p(f_k(X_{k-1}))\right],$$

where $Q_p(t) = \|\varepsilon(t)\|_p^p$. Note that, when p is an integer, Q_p denotes the Stirling polynomial defined through Stirling numbers (see the Lemma A-1 in [13] from [17]).

- **GLM–GARCH Poisson models**. One can give numerous extensions of the previous model. Keeping the same notations, one can consider $f_n : \mathbb{R}^+ \times \mathbb{N} \to \mathbb{R}^+$, and

$$F_n(x, y) = (f_n(x), y(f_n(x))),$$

where $x = (\lambda, z) \in \mathbb{R}^+ \times \mathbb{N}$. Let $|x| = |\lambda| + a|z|$. Then

$$\mathbb{E}[|F_n(x, \varepsilon) - F_n(x', \varepsilon_n)|] \leq (1 + a)|f_n(x) - f_n(x')|,$$

and (2.8) is true provided that $|f_n(x) - f_n(x')| \leq L_n|x - x'|$ and $\sup_{n\geq 2} L_n(1 + a) = \rho$ for some $\rho < 1$.

For these models,

$$H_k(X_{k-1}, \varepsilon_k) = a\mathbb{E}\big[|\varepsilon(f_k(X_{k-1})) - \varepsilon_k(f_k(X_{k-1}))|\big|\varepsilon_k, X_{k-1}\big],$$

where ε is a unit Poisson process independent of (X_{k-1}, ε_k). Hence, the moments of order p of $H_k(X_{k-1}, \varepsilon_k)$ can be controlled exactly as in the previous example.

3   Separately Lipschitz functions of $X_1, \ldots, X_n$

Let $f : \mathcal{X}^n \mapsto \mathbb{R}$ be a separately Lipschitz function, such that

$$|f(x_1, x_2, \ldots, x_n) - f(x_1', x_2', \ldots, x_n')| \leq d(x_1, x_1') + \cdots + d(x_n, x_n'). \tag{3.1}$$

Let

$$S_n := f(X_1, \ldots, X_n) - \mathbb{E}[f(X_1, \ldots, X_n)]. \tag{3.2}$$

We introduce the natural filtration of the chain, that is $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and for all $k \in \mathbb{N}^*$, $\mathcal{F}_k = \sigma(X_1, X_2, \ldots, X_k)$. Define

$$g_k(X_1, \ldots, X_k) = \mathbb{E}[f(X_1, \ldots, X_n)|\mathcal{F}_k] \tag{3.3}$$

and

$$M_k = g_k(X_1, \ldots, X_k) - g_{k-1}(X_1, \ldots, X_{k-1}). \tag{3.4}$$

For all $k \in [1, n - 1]$, let

$$S_k := M_1 + M_2 + \cdots + M_k,$$

and notice that, by the definition of M_k's, the functional S_n introduced in (3.2) satisfies

$$S_n = M_1 + M_2 + \cdots + M_n.$$

Thus S_k is a martingale adapted to the natural filtration $\mathcal{F}_k$. This representation appears in Yurinskii [32] and in p. 33 of the monograph by Milman and Schechtman [24]. In the setting of separately Lipschitz functions of independent random variables (*i.e.* when $X_i = \varepsilon_i$) it has been used by McDiarmid [23] to get an exponential bound on tail probabilities $\mathbb{P}(S_n \geq x), x \geq 0$.

The following Proposition, similar to Proposition 2.1 in [10], collects some interesting properties of the function g_k and of the martingale difference M_k.

Proposition 3.1. *For all $k \in \mathbb{N}$ and any ρ in $[0,1)$, denote $K_k(\rho) = (1 - \rho^{k+1})/(1 - \rho) = 1 + \rho + \cdots + \rho^k$. Let $(X_i)_{i \geq 1}$ be a Markov chain satisfying (2.7) for some functions F_n satisfying (2.8). Let g_k and M_k be defined by (3.3) and (3.4) respectively.*

1. *The function g_k is separately Lipschitz, and satisfies*

$$\left| g_k(x_1, x_2, \ldots, x_k) - g_k(x_1', x_2', \ldots, x_k') \right| \leq d(x_1, x_1') + \cdots + d(x_{k-1}, x_{k-1}') + K_{n-k}(\rho) d(x_k, x_k').$$

2. *Denote by P_{X_1} and P_{ε_k} the distribution of X_1 and the distribution of ε_k respectively. Let G_{X_1} and H_k be two functions defined as follows*

$$G_{X_1}(x) = \int d(x, x') P_{X_1}(dx') \quad \text{and} \quad H_k(x, y) = \int d(F_k(x, y), F_k(x, y')) P_{\varepsilon_k}(dy').$$

Then, the martingale difference M_k satisfies

$$|M_1| \leq K_{n-1}(\rho) G_{X_1}(X_1) \quad \text{and} \quad |M_k| \leq K_{n-k}(\rho) H_k(X_{k-1}, \varepsilon_k), \quad k \in [2, n].$$

Remark 3.1. Assume moreover that F_n satisfies

$$d(F_n(x, y), F_n(x, y')) \leq C(x)\, \delta(y, y') \tag{3.5}$$

for some function $C(x) \geq 0$ not depending on n, and let G_k be the function defined by

$$G_k(y) = \int \delta(y, y') P_{\varepsilon_k}(dy') \quad \text{for all } k \in [2, n].$$

Then $H_k(x, y) \leq C(x) G_k(y)$ and, consequently,

$$|M_k| \leq K_{n-k}(\rho) C(X_{k-1}) G_k(\varepsilon_k) \quad \text{for all } k \in [2, n].$$

Note that (3.5) is a non-uniform version of (2.9), which is satisfied for many examples (for instance the three first examples of Section 2). However, it seems quite difficult to check for INAR or GLM type models, while the moments of the dominating variables $H_k(X_{k-1}, \varepsilon_k)$ are easy to control for such models (see Section 2).

Proof. The first point will be proved by recurrence in the backward sense. For $k = n$, the result is obvious due to $g_n = f$. Suppose it is true at step k, and let us prove it at step $k-1$. By definition

$$g_{k-1}(X_1, \ldots, X_{k-1}) = \mathbb{E}[g_k(X_1, \ldots, X_k)|\mathcal{F}_{k-1}] = \int g_k(X_k, \ldots, X_{k-1}, F_k(X_{k-1}, y)) P_{\varepsilon_k}(dy).$$

Then it is easy to see that

$$|g_{k-1}(x_1, x_2, \ldots, x_{k-1}) - g_{k-1}(x_1', x_2', \ldots, x_{k-1}')|$$
$$\leq \int \left| g_k(x_1, x_2, \ldots, F_k(x_{k-1}, y)) - g_k(x_1', x_2', \ldots, F_k(x_{k-1}', y)) \right| P_{\varepsilon_k}(dy). \tag{3.6}$$

Now, by assumption and condition (2.8),

$$\int \left| g_k(x_1, x_2, \ldots, F_k(x_{k-1}, y)) - g_k(x_1', x_2', \ldots, F_k(x_{k-1}', y)) \right| P_{\varepsilon_k}(dy)$$

$$\leq d(x_1, x_1') + \cdots + d(x_{k-1}, x_{k-1}') + K_{n-k}(\rho) \int d(F_k(x_{k-1}, y), F_k(x_{k-1}', y)) P_{\varepsilon_k}(dy)$$

$$\leq d(x_1, x_1') + \cdots + (1 + \rho K_{n-k}(\rho)) d(x_{k-1}, x_{k-1}')$$

$$\leq d(x_1, x_1') + \cdots + K_{n-k+1}(\rho) d(x_{k-1}, x_{k-1}'). \tag{3.7}$$

7

The point 1 follows from (3.6) and (3.7).

Next, we prove the point 2. First notice that

$$|M_1| = \left| g_1(X_1) - \int g_1(x) P_{X_1}(dx) \right| \le K_{n-1}(\rho) \int d(X_1, x) P_{X_1}(dx) = K_{n-1}(\rho) G_{X_1}(X_1).$$

Similarly, for all $k \ge 2$,

$$
\begin{aligned}
|M_k| &= \left| g_k(X_1, \cdots, X_k) - \mathbb{E}[g_k(X_1, \cdots, X_k)|\mathcal{F}_{k-1}] \right| \\
&\le \int \left| g_k(X_1, \cdots, F_k(X_{k-1}, \varepsilon_k)) - g_k(X_1, \cdots, F_k(X_{k-1}, y)) \right| P_{\varepsilon_k}(dy) \\
&\le K_{n-k}(\rho) \int d(F_k(X_{k-1}, \varepsilon_k), F_k(X_{k-1}, y)) P_{\varepsilon_k}(dy) = K_{n-k}(\rho) H_k(X_{k-1}, \varepsilon_k).
\end{aligned}
$$

This completes the proof of Proposition 3.1. □

4 Deviation inequalities for the functional S_n

Let $(X_i)_{i \ge 1}$ be a Markov chain satisfying (2.7) for some functions F_n satisfying (2.8). In this section, we apply inequalities for martingales to bound up the deviation of the functional S_n defined by (3.2). Some of these inequalities are direct applications of known inequalities, and some deserve a short proof.

Denote by $S_{2,n} = S_n - M_1$, and let a_n be a sequence of positive numbers. Then, for any $x > 0$,

$$
\begin{aligned}
\mathbb{P}\Big(S_n \ge a_n x\Big) &\le \mathbb{P}\Big(M_1 \ge a_n x/2\Big) + \mathbb{P}\Big(S_{2,n} \ge a_n x/2\Big) \\
&\le \mathbb{P}\left(G_{X_1}(X_1) \ge \frac{a_n x}{2 K_{n-1}(\rho)}\right) + \mathbb{P}\Big(S_{2,n} \ge a_n x/2\Big) =: I_1(a_n, x) + I_2(a_n, x), \quad (4.1)
\end{aligned}
$$

and note that the same bound is valid for $\mathbb{P}(-S_n \ge a_n x)$ by replacing the term $I_2(a_n, x)$ by $\tilde{I}_2(a_n, x) := \mathbb{P}(-S_{2,n} \ge a_n x/2)$.

The term $I_1(a_n, x)$ will be most of the time negligible, and represents the direct influence of the initial distribution of the chain. For instance, when the chain starts from a point $X_1 = x_1$, then $G_{X_1}(X_1) = 0$ and $I_1(a_n, x) = 0$. The main difficulty is to give an upper bound for $I_2(a_n, x)$, which is the purpose of the present paper.

4.1 A first exponential bound

Under a sub-Gaussian type condition, we obtain the following proposition.

Proposition 4.1. *Assume that there exists a positive constant ϵ such that, for any integer $k \ge 2$,*

$$\mathbb{E}\left[\big(H_k(X_{k-1}, \varepsilon_k)\big)^l\right] \le \frac{1}{2} \frac{l! \, \epsilon^{l-2}}{(l-1)^{l/2}} \mathbb{E}\left[\big(H_k(X_{k-1}, \varepsilon_k)\big)^2\right] \text{ for all } l \ge 2. \quad (4.2)$$

Then, for any $x > 0$,

$$\mathbb{P}\Big(\pm S_n \ge x V_n\Big) \le I_1(V_n, x) + \exp\left\{ -\frac{(x/2)^2}{1 + \sqrt{1 + x\epsilon K_{n-2}(\rho)/\sigma_n} + x\epsilon K_{n-2}(\rho)/2\sigma_n} \right\} \quad (4.3)$$

$$\le I_1(V_n, x) + \exp\left\{ -\frac{(x/2)^2}{2\big(1 + x\epsilon K_{n-2}(\rho)/2\sigma_n\big)} \right\}, \quad (4.4)$$

where

$$V_n^2 = \sum_{k=2}^n K_{n-k}^2(\rho) \mathbb{E}\left[\big(H_k(X_{k-1}, \varepsilon_k)\big)^2\right] \text{ and } \sigma_n^2 = \frac{1}{n} V_n^2.$$

8

Remark 4.1. Let us give some comments on Proposition 4.1.

1. Condition (4.2) is in fact a sub-Gaussian condition. On can check that it is satisfied provided

$$\inf_{k \geq 2} \mathbb{E}\big[\big(H_k(X_{k-1}, \varepsilon_k)\big)^2\big] > 0$$

and

$$\sup_{k \geq 2} \mathbb{E}\Big[\exp\Big\{c\,\big(H_k(X_{k-1}, \varepsilon_k)\big)^2\Big\}\Big] < \infty$$

for some positive constant c not depending on k.

2. Assume that

$$\mathbb{E}\Big[\exp\Big\{c\,\sqrt{G_{X_1}(X_1)}\Big\}\Big] < \infty$$

for some positive constant c, and that

$$0 < \liminf_{n \to \infty} \sigma_n \leq \limsup_{n \to \infty} \sigma_n < \infty.$$

Then, it follows from Proposition 4.1 that

$$\mathbb{P}\left(\pm S_n \geq n\right) = O\left(\exp\left\{-C\sqrt{n}\right\}\right) \tag{4.5}$$

for some positive constant C.

Proof. By Taylor's expansion of e^x and the fact that $\mathbb{E}[S_{2,n}] = 0$, we have, for all $t \geq 0$,

$$\mathbb{E}\left[\exp\left\{t\frac{S_{2,n}}{\sqrt{n}}\right\}\right] = 1 + \sum_{k=2}^{\infty} \frac{t^k}{k!}\mathbb{E}\left[\left(\frac{S_{2,n}}{\sqrt{n}}\right)^k\right]. \tag{4.6}$$

Using Rio's inequality (see Theorem 2.1 of [28]): for any $p \geq 2$,

$$\left(\mathbb{E}[|S_{2,n}|^p]\right)^{2/p} \leq (p-1) \sum_{i=2}^{n} \left(\mathbb{E}[|M_i|^p]\right)^{2/p}, \tag{4.7}$$

we get, for all $k \geq 2$,

$$\mathbb{E}[|S_{2,n}|^k] \leq (k-1)^{k/2}\left(\sum_{i=2}^{n} \left(\mathbb{E}[|M_i|^k]\right)^{2/k}\right)^{k/2}. \tag{4.8}$$

Hence, by Hölder's inequality, inequality (4.8) implies that, for all $k \geq 2$,

$$\mathbb{E}[|S_{2,n}|^k] \leq (k-1)^{k/2} n^{k/2-1} \sum_{i=2}^{n} \mathbb{E}[|M_i|^k]. \tag{4.9}$$

Applying the last inequality to (4.6), we obtain

$$\mathbb{E}\left[\exp\left\{t\frac{S_{2,n}}{\sqrt{n}}\right\}\right] \leq 1 + \sum_{k=2}^{\infty} \left(\frac{t^k}{k!}(k-1)^{k/2} n^{-1} \sum_{i=2}^{n} \mathbb{E}[|M_i|^k]\right). \tag{4.10}$$

By points 2 of Proposition 3.1 and (4.2), we deduce that, for any integer $i \geq 2$,

$$\begin{aligned}
\mathbb{E}[|M_i|^l] &\leq \mathbb{E}[|K_{n-i}(\rho)H_i(X_{i-1}, \varepsilon_i)|^l] \\
&\leq \frac{1}{2}\frac{l!\,(K_{n-2}(\rho)\epsilon)^{l-2}}{(l-1)^{l/2}}\mathbb{E}[(K_{n-i}(\rho)H_i(X_{i-1}, \varepsilon_i))^2] \qquad \text{for all } l \geq 2.
\end{aligned}$$

Hence condition (4.2) implies that, for all $0 \leq t < (K_{n-2}(\rho)\epsilon)^{-1}$,

$$\mathbb{E}\left[\exp\left\{t\frac{S_{2,n}}{\sqrt{n}}\right\}\right] \leq 1 + \sum_{k=2}^{\infty} \frac{\sigma_n^2}{2}t^k(K_{n-2}(\rho)\epsilon)^{k-2} = 1 + \frac{t^2\sigma_n^2}{2\left(1 - tK_{n-2}(\rho)\epsilon\right)}. \tag{4.11}$$

9

By the inequality $1 + x \le e^x$, it follows that, for all $0 \le t < (K_{n-2}(\rho)\epsilon)^{-1}$,

$$\mathbb{E}\left[\exp\left\{t\frac{S_{2,n}}{\sqrt{n}}\right\}\right] \le \exp\left\{\frac{t^2\sigma_n^2}{2(1 - tK_{n-2}(\rho)\epsilon)}\right\}.$$

Applying Markov's inequality, it is then easy to see that, for all $0 \le t < \sigma_n(K_{n-2}(\rho)\epsilon)^{-1}$ and $x \ge 0$,

$$\mathbb{P}(S_{2,n} \ge xV_n/2) \le \exp\left\{-tx/2\right\}\mathbb{E}\left[\exp\left\{t\frac{S_{2,n}}{V_n}\right\}\right].$$

Hence

$$
\begin{aligned}
\mathbb{P}(S_{2,n} \ge xV_n/2) &\le \inf_{0 \le t < \sigma_n(K_{n-2}(\rho)\epsilon)^{-1}} \exp\left\{-tx/2 + \frac{t^2}{2(1 - t\,K_{n-2}(\rho)\epsilon/\sigma_n)}\right\} \\
&= \exp\left\{-\frac{(x/2)^2}{1 + \sqrt{1 + xK_{n-2}(\rho)\epsilon/\sigma_n} + xK_{n-2}(\rho)\epsilon/2\sigma_n}\right\},
\end{aligned}
$$

which gives (4.3). Using the inequality $\sqrt{1 + xK_{n-2}(\rho)\epsilon/\sigma_n} \le 1 + xK_{n-2}(\rho)\epsilon/2\sigma_n$, we get (4.4) from (4.3). □

4.2 Semi-exponential bounds

In the case where the variables $H_k(X_{k-1}, \varepsilon_k)$ have semi-exponential moments, the following proposition holds.

Proposition 4.2. *Let $\alpha \in (0, 1)$. Assume that there exists a constant C_1 such that, for any integer $k \ge 2$,*

$$\mathbb{E}\left[\exp\left\{\left(H_k(X_{k-1}, \varepsilon_k)\right)^{\frac{2\alpha}{1-\alpha}}\right\}\right] \le C_1. \tag{4.12}$$

Then, for any $x > 0$,

$$\mathbb{P}(\pm S_n \ge nx) \le I_1(n, x) + C(\alpha, x)\exp\left\{-\left(\frac{x}{8K_{n-2}(\rho)}\right)^{2\alpha} n^\alpha\right\}, \tag{4.13}$$

where

$$C(\alpha, x) = 2 + 35C_1\left(\frac{K_{n-2}^{2\alpha}(\rho)}{x^{2\alpha}4^{2-3\alpha}} + \frac{4K_{n-2}^2(\rho)}{x^2}\left(\frac{3(1-\alpha)}{2\alpha}\right)^{\frac{1-\alpha}{\alpha}}\right)$$

depends on n only through the term $K_{n-2}(\rho)$.

Remark 4.2. Let us comment on inequality (4.13). Assume moreover that

$$\mathbb{E}\left[\exp\left\{c\left(G_{X_1}(X_1)\right)^\alpha\right\}\right] \le C_2, \tag{4.14}$$

for two positive constants c, C_2. Then, it follows from (4.13) that

$$\mathbb{P}(\pm S_n \ge n) = O\left(\exp\left\{-Cn^\alpha\right\}\right), \tag{4.15}$$

for some positive constant C. This rate is in accordance with the best possible rate for large deviation of partial sums of martingales differences, as proved in Theorem 2.1 of [15]. For partial sums of independent random variables, the rate (4.15) holds under weaker conditions on exponential moments, see Lanzinger and Stadtmüller [22].

Proof. From point 2 of Proposition 3.1 and condition (4.12), it is easy to see that, for any $k \in [2, n]$,

$$\mathbb{E}\left[\exp\left\{|K_{n-2}^{-1}(\rho)M_k|^{\frac{2\alpha}{1-\alpha}}\right\}\right] \leq C_1. \tag{4.16}$$

Applying Theorem 2.1 of Fan et al. [15] to the martingale sequence $(K_{n-2}^{-1}(\rho)M_k, \mathcal{F}_k)_{k=2,..,n}$, we get, for any $x > 0$,

$$I_2(n, x) \leq C(\alpha, x)\exp\left\{-\left(\frac{x}{8K_{n-2}(\rho)}\right)^{2\alpha}n^\alpha\right\}. \tag{4.17}$$

Combining the inequalities (4.1) and (4.17), we obtain the desired inequality. $\qquad\square$

For the next proposition, let us introduce the random variables

$$L_k(X_{k-1}), \quad \text{where} \quad L_k(x) = \int (H_k(x, y))^2 P_{\varepsilon_k}(dy), \tag{4.18}$$

and note that

$$L_k(X_{k-1}) = \mathbb{E}\left[(H_k(X_{k-1}, \varepsilon_k))^2 \,|X_{k-1}\right].$$

According to Proposition 3.1, for any $k \geq 2$, $\mathbb{E}[M_k^2|\mathcal{F}_{k-1}] \leq K_{n-k}(\rho)L_k(X_{k-1})$.

Proposition 4.3. *Assume (3.5), and let $\alpha \in (0, 1)$. Assume that there exist two constants C_1 and C_2 such that, for any integer $k \geq 2, n \geq 2$,*

$$\mathbb{E}\left[\exp\left\{\left(\frac{1}{n}\sum_{k=2}^n L_k(X_{k-1})\right)^{\frac{\alpha}{1-\alpha}}\right\}\right] \leq C_1 \tag{4.19}$$

and

$$\mathbb{E}\left[\exp\left\{\left(H_k(X_{k-1}, \varepsilon_k)\right)^{\frac{\alpha}{1-\alpha}}\right\}\right] \leq C_2. \tag{4.20}$$

Then, for all $x > 0$,

$$\mathbb{P}\left(\pm S_n \geq nx\right) \leq I_1(n, x) + \exp\left\{-\frac{(xK_{n-2}^{-1}(\rho)/2)^{1+\alpha}}{2\left(1 + xK_{n-2}^{-1}(\rho)/6\right)}n^\alpha\right\}$$
$$+ (C_1 + nC_2)\exp\left\{-\left(xK_{n-2}^{-1}(\rho)/2\right)^\alpha n^\alpha\right\}. \tag{4.21}$$

Remark 4.3. According to Remark 4.2, under the conditions (4.19), (4.20) and (4.14), we have

$$\mathbb{P}\left(\pm S_n \geq n\right) = O\left(\exp\left\{-Cn^\alpha\right\}\right), \tag{4.22}$$

for some positive constant C. This rate is in accordance with the best possible rate for large deviation of partial sums of martingales differences, as proved in Corollary 2.3 of [15]. Note that if $\alpha \in [1/2, 1)$, the condition (4.19) is true provided that

$$\sup_{k \geq 2}\mathbb{E}\left[\exp\left\{\left(L_k(X_{k-1})\right)^{\frac{\alpha}{1-\alpha}}\right\}\right] \leq C_1. \tag{4.23}$$

For $\alpha \in [1/2, 1)$, the two conditions (4.23) and (4.20) are clearly less restrictive than (4.12), so Proposition 4.3 is more precise than Proposition 4.2 in the regime of large deviation. However, it does not allow to control moderate deviations $\mathbb{P}\left(\pm S_n \geq n^\delta\right)$ for δ close to $1/2$, which is possible via Proposition 4.2 (see for instance (1.6)).

Proof. From point 2 of Proposition 3.1 and condition (4.20), it is easy to see that, for any $k \in [2, n]$,

$$\mathbb{E}\left[\exp\left\{|K_{n-2}^{-1}(\rho)M_k|^{\frac{\alpha}{1-\alpha}}\right\}\right] \leq C_2.$$

11

For any $k \in [2, n]$,

$$
\begin{aligned}
\mathbb{E}\left[\left|K_{n-2}^{-1}(\rho)M_k\right|^2\Big|\mathcal{F}_{k-1}\right] &\leq \mathbb{E}\left[\left(K_{n-2}^{-1}(\rho)K_{n-k}(\rho)H_k(X_{k-1},\varepsilon_k)\right)^2\Big|\mathcal{F}_{k-1}\right] \\
&\leq \mathbb{E}\left[\left(H_k(X_{k-1},\varepsilon_k)\right)^2\Big|\mathcal{F}_{k-1}\right].
\end{aligned}
\tag{4.24}
$$

Thus

$$
\sum_{k=2}^n \mathbb{E}\left[\left|K_{n-2}^{-1}(\rho)M_k\right|^2\Big|\mathcal{F}_{k-1}\right] \leq \sum_{k=2}^n \mathbb{E}\left[\left(H_k(X_{k-1},\varepsilon_k)\right)^2\Big|\mathcal{F}_{k-1}\right] = \sum_{k=2}^n L_k(X_{k-1}).
$$

Using Theorem 2.2 of Fan et al. [15], we have

$$
\mathbb{P}\left(K_{n-2}^{-1}(\rho)S_{2,n} \geq nK_{n-2}^{-1}(\rho)x/2 \ \text{ and } \ \sum_{k=2}^n L_k(X_{k-1}) \leq nv^2\right)
$$
$$
\leq \exp\left\{-\frac{(K_{n-2}^{-1}(\rho)x/2)^2}{2\left(n^{\alpha-1}v^2 + \frac{1}{3}(K_{n-2}^{-1}(\rho)x/2)^{2-\alpha}\right)}n^\alpha\right\} + nC_2\exp\left\{-\left(\frac{x}{2K_{n-2}(\rho)}\right)^\alpha n^\alpha\right\}.
$$

From the last inequality, we deduce that

$$
\begin{aligned}
I_2(n,x) &\leq \exp\left\{-\frac{(K_{n-2}^{-1}(\rho)x/2)^2}{2\left(n^{\alpha-1}v^2 + \frac{1}{3}(K_{n-2}^{-1}(\rho)x/2)^{2-\alpha}\right)}n^\alpha\right\} + nC_2\exp\left\{-\left(\frac{x}{2K_{n-2}(\rho)}\right)^\alpha n^\alpha\right\} \\
&\quad + \mathbb{P}\left(\sum_{k=2}^n L_k(X_{k-1}) > nv^2\right).
\end{aligned}
$$

Using the exponential Markov inequality and the condition (4.19), we get, for all $v > 0$,

$$
\mathbb{P}\left(\sum_{k=2}^n L_k(X_{k-1}) > nv^2\right) \leq C_1\exp\left\{-v^{\frac{2\alpha}{1-\alpha}}\right\}.
$$

Taking $v^2 = (xnK_{n-2}^{-1}(\rho)/2)^{(1-\alpha)/2}$, we have, for all $x > 0$,

$$
I_2(n,x) \leq \exp\left\{-\frac{(xK_{n-2}^{-1}(\rho)/2)^{1+\alpha}}{2\left(1 + xK_{n-2}^{-1}(\rho)/6\right)}n^\alpha\right\} + (C_1 + nC_2)\exp\left\{-\left(xK_{n-2}^{-1}(\rho)/2\right)^\alpha n^\alpha\right\}.
$$

Combining the last inequality and (4.1), we obtain the desired inequality. $\qquad\square$

4.3 Fuk-Nagaev type bound

We now consider the case where the random variables $H_k(X_{k-1},\varepsilon_k), k \geq 2$, have only a weak moment of order $p > 2$. For any real-valued random variable Z and any $p \geq 1$, define the weak moment of order p by

$$
\|Z\|_{w,p}^p = \sup_{x>0} x^p \mathbb{P}(|Z| > x).
\tag{4.25}
$$

The following proposition is a Fuk-Nagaev type inequality (cf. Fuk [19] and Nagaev [25]; see also Fan *et al.* [16] and Rio [29] for martingales).

Proposition 4.4. *Let $p \geq 2$ and $\delta > 0$, and consider the variables $L_k(X_{k-1})$ defined in (4.18). Assume that there exist two constants C_1 and C_2 such that, for any integer $k \geq 2$,*

$$
\left\|\frac{1}{n}\sum_{k=2}^n L_k(X_{k-1})\right\|_{w,p+\delta}^{p+\delta} \leq C_1
\tag{4.26}
$$

and

$$
\|H_k(X_{k-1},\varepsilon_k)\|_{w,p+\delta}^{p+\delta} \leq C_2.
\tag{4.27}
$$

12

Then, for all $x > 0$,

$$\mathbb{P}\left(\pm S_n \geq nx\right) \leq I_1(n, x) + \exp\left\{-\frac{(K_{n-2}^{-1}(\rho)/2)^2}{2\left(n^{-1/(p+\delta)}x^{-1} + \frac{1}{6}K_{n-2}^{-1}(\rho)\right)}(nx)^{\delta/(p+\delta)}\right\} + \frac{C_1 + C_2}{n^{p-1}x^p}. \tag{4.28}$$

Remark 4.4. Let us comment on Proposition 4.4.

1. If there exists a constant C_3 such that

$$\|G_{X_1}(X_1)\|_{w,p-1}^{p-1} \leq C_3, \tag{4.29}$$

   then, for any $x > 0$,

$$I_1(n, x) = \mathbb{P}\left(G_{X_1}(X_1) \geq \frac{nx}{2K_{n-1}(\rho)}\right) \leq (2K_{n-1}(\rho))^{p-1}\frac{C_3}{(nx)^{p-1}}. \tag{4.30}$$

   Thus under conditions (4.26), (4.27) and (4.29), we have

$$\mathbb{P}\left(\pm S_n \geq n\right) = O\left(\frac{1}{n^{p-1}}\right).$$

2. Assume moreover that F_n satisfies (3.5). Then, according to Remark 3.1, Proposition 4.4 remains valid when $H_k(X_{k-1}, \varepsilon_k)$ is replaced by $C(X_{k-1})G_k(\varepsilon_k)$. Since $C(X_{k-1})$ and $G_k(\varepsilon_k)$ are independent, we easily see that

$$L_k(X_{k-1}) \leq (C(X_{k-1}))^2\mathbb{E}\left[(G_k(\varepsilon_k))^2\right]$$

   and

$$\|C(X_{k-1})G_k(\varepsilon_k)\|_{w,p+\delta}^{p+\delta} \leq \mathbb{E}\left[(C(X_{k-1}))^{p+\delta}\right]\|G_k(\varepsilon_k)\|_{w,p+\delta}^{p+\delta}. \tag{4.31}$$

   Thus if

$$\mathbb{E}\left[(C(X_{k-1}))^{p+\delta}\right] \leq C_3 \quad \text{and} \quad \|G_k(\varepsilon_k)\|_{w,p+\delta}^{p+\delta} \leq C_4,$$

   then condition (4.27) is satisfied with $C_2 = C_3C_4$. Of course, the same computations may be done by interchanging $C(X_{k-1})$ and $G_k(\varepsilon_k)$. Hence, if

$$\mathbb{E}\left[(G_k(\varepsilon_k))^{p+\delta}\right] \leq C_3 \quad \text{and} \quad \|C(X_{k-1})\|_{w,p+\delta}^{p+\delta} \leq C_4,$$

   then condition (4.27) still holds with $C_2 = C_3C_4$.

Proof. To prove Proposition 4.4, we need the following inequality whose proof can be found in Fan *et al.* [14] (cf. Corollary 2.3 and Remark 2.1 therein).

Lemma 4.1. *Assume that $(\xi_i, \mathcal{G}_i)_{i \geq 1}$ are square integrable martingale differences, and let $Z_n = \xi_1 + \cdots + \xi_n$ and $\langle Z\rangle_n = \sum_{k=1}^{n}\mathbb{E}[\xi_k^2|\mathcal{G}_{k-1}]$. Then, for all $x, y, v > 0$,*

$$\mathbb{P}\left(Z_n \geq x \text{ and } \langle Z\rangle_n \leq v^2\right) \leq \exp\left\{-\frac{x^2}{2(v^2 + \frac{1}{3}xy)}\right\} + \mathbb{P}\left(\max_{1 \leq i \leq n}\xi_i > y\right).$$

By Lemma 4.1 and Markov's inequality, it follows that, for all $x, y, v > 0$,

$$\mathbb{P}\left(K_{n-2}^{-1}(\rho)S_{2,n} \geq nK_{n-2}^{-1}(\rho)x/2 \text{ and } \sum_{k=2}^{n}L_k(X_{k-1}) \leq nv^2\right)$$

$$\leq \exp\left\{-\frac{(nK_{n-2}^{-1}(\rho)x/2)^2}{2\left(nv^2 + \frac{1}{6}nK_{n-2}^{-1}(\rho)xy\right)}\right\} + \mathbb{P}\left(\max_{2 \leq i \leq n}K_{n-2}^{-1}(\rho)M_i \geq y\right).$$

It is easy to see that, for all $y > 0$,

$$\mathbb{P}\left(\max_{2 \leq i \leq n} K_{n-2}^{-1}(\rho)M_i \geq y\right) \leq \sum_{i=2}^{n} \mathbb{P}\left(K_{n-2}^{-1}(\rho)M_i \geq y\right) \leq nC_2 y^{-(p+\delta)},$$

and that, for all $v > 0$,

$$\mathbb{P}\left(\sum_{k=2}^{n} L_k(X_{k-1}) > nv^2\right) \leq C_1 v^{-2(p+\delta)}.$$

Thus, for all $x, y, v > 0$,

$$I_2(n, x) \leq \exp\left\{-\frac{(nK_{n-2}^{-1}(\rho)x/2)^2}{2\left(nv^2 + \frac{1}{6}nK_{n-2}^{-1}(\rho)xy\right)}\right\} + \frac{nC_2}{y^{p+\delta}} + \frac{C_1}{v^{2(p+\delta)}}.$$

Taking $y = (nx)^{p/(p+\delta)}$ and $v^{2(p+\delta)} = n^{p-1}x^p$ in the last inequality, we obtain, for all $x > 0$,

$$I_2(n, x) \leq \exp\left\{-\frac{(K_{n-2}^{-1}(\rho)/2)^2}{2\left(n^{-1/(p+\delta)}x^{-1} + \frac{1}{6}K_{n-2}^{-1}(\rho)\right)}(nx)^{\delta/(p+\delta)}\right\} + \frac{C_1 + C_2}{n^{p-1}x^p}.$$

Combining the last inequality and (4.1), we obtain the desired inequality. $\qquad\square$

4.4 von Bahr-Esseen's inequality, weak form

We now consider the case where the variables $G_{X_1}(X_1)$ and $H_k(X_{k-1}, \varepsilon_k)$ have only a weak moment of order $p \in (1, 2)$.

Proposition 4.5. *Let $p \in (1, 2)$. Assume that there exists positive constants $A_k(p)$ such that, for any $k \in [2, n]$,*

$$\|H_k(X_{k-1}, \varepsilon_k)\|_{w,p}^p \leq A_k(p). \tag{4.32}$$

Then, for any $x > 0$,

$$\mathbb{P}(|S_n| \geq x) \leq 2I_1(1, x) + \frac{2^p C_p B_p(n, \rho)}{x^p}, \tag{4.33}$$

where

$$C_p = \frac{4p}{p-1} + \frac{8}{2-p}$$

and

$$B_p(n, \rho) = \sum_{k=2}^{n} (K_{n-k}(\rho))^p A_k(p).$$

Remark 4.5. Let us comment on Proposition 4.5.

1. Contrary to the previous inequalities of Section 4, this inequality is truly non-stationary, in the sense that it is expressed in terms of the weak moments $\|H_k(X_{k-1}, \varepsilon_k)\|_{w,p}^p$, without assuming a uniform bound (in k) on these moments. This will also be the case of the moment inequalities of Section 5.

2. Assume moreover that F_n satisfies (3.5). Then, according to Remark 3.1, Proposition 4.4 remains true if condition (4.32) is replaced by

$$\|C(X_{k-1})G_k(\varepsilon_k)\|_{w,p}^p \leq A_k(p). \tag{4.34}$$

In particular, if either

$$\mathbb{E}\left[(C(X_{k-1}))^p\right] \leq A_{1,k}(p) \quad \text{and} \quad \|G_k(\varepsilon_k)\|_{w,p}^p \leq A_{2,k}(p)$$

or

$$\mathbb{E}\left[(G_k(\varepsilon_k))^p\right] \leq A_{1,k}(p) \quad \text{and} \quad \|C(X_{k-1})\|_{w,p}^p \leq A_{2,k}(p),$$

hold, then condition (4.34) is satisfied with $A_k(p) = A_{1,k}(p)A_{2,k}(p)$.

14

3. Assume that $\|G_{X_1}(X_1)\|_{w,p-1} < \infty$ and that $B_p(n, \rho) = O(n)$, then

$$\mathbb{P}\left(|S_n| \geq n\right) = O\left(\frac{1}{n^{p-1}}\right).$$

Proof. By Proposition 3.3 of Cuny *et al.* [7], we have, for any $x > 0$,

$$\mathbb{P}(|S_{2,n}| \geq x/2) \leq \frac{2^p C_p}{x^p} \sum_{k=2}^{n} \|M_k\|_{w,p}^p. \tag{4.35}$$

From point 2 of Proposition 3.1 and condition (4.32), it follows that, for any $k \in [2, n]$,

$$\|M_k\|_{w,p}^p \leq \|K_{n-k}(\rho) H_k(X_{k-1}, \varepsilon_k)\|_{w,p}^p \leq (K_{n-k}(\rho))^p A_k(p). \tag{4.36}$$

Combining (4.35) and (4.36), we obtain the desired inequality. $\qquad\square$

5   Moment inequalities

In this section, we shall control the $\mathbb{L}_p$-norm of the functional S_n, for $p > 1$. The upper bounds will be expressed in terms of the moments of order p of the dominating random variables $H_k(X_{k-1}, \varepsilon_k)$. Let us emphasize that all the inequalities of these section are completely non-stationary, in the sense that we shall not assume a uniform bound (in k) on $\|H_k(X_{k-1}, \varepsilon_k)\|_p^p$.

5.1   Marcinkiewicz-Zygmund type inequality

We assume in this subsection that the dominating random variables $G_{X_1}(X_1)$ and $H_k(X_{k-1}, \varepsilon_k)$ have a moment of order $p \geq 2$.

Proposition 5.1. *Let $p \geq 2$. Assume that there exist positive constants $A_k(p)$ such that*

$$\mathbb{E}\left[\left(G_{X_1}(X_1)\right)^p\right] \leq A_1(p), \quad \text{and for } k \in [2, n] \quad \mathbb{E}\left[\left(H_k(X_{k-1}, \varepsilon_k)\right)^p\right] \leq A_k(p). \tag{5.37}$$

Then

$$\|S_n\|_p \leq \sqrt{A_p(n, \rho)}, \tag{5.38}$$

where

$$A_p(n, \rho) = \left(K_{n-1}(\rho)\right)^2 \left(A_1(p)\right)^{2/p} + (p-1) \sum_{k=2}^{n} \left(K_{n-k}(\rho)\right)^2 \left(A_k(p)\right)^{2/p}.$$

Remark 5.1. Assume moreover that F_n satisfies (3.5). Then, according to Remark 3.1, inequality (5.38) remains true if the second condition of (5.37) is replaced by

$$\mathbb{E}\left[(C(X_{k-1}))^p\right] \mathbb{E}\left[(G_k(\varepsilon_k))^p\right] \leq A_k(p). \tag{5.39}$$

Proof. Applying Theorem 2.1 of Rio [28], we get

$$\|S_n\|_p^2 \leq \|M_1\|_p^2 + (p-1) \sum_{k=2}^{n} \|M_k\|_p^2.$$

By Proposition 3.1 and condition (5.37), it follows that

$$
\begin{aligned}
\|S_n\|_p^2 &\leq \left(K_{n-1}(\rho)\right)^2 \left(\mathbb{E}\left[\left(G_{X_1}(X_1)\right)^p\right]\right)^{\frac{2}{p}} + (p-1) \sum_{k=2}^{n} \left(K_{n-k}(\rho)\right)^2 \left(\mathbb{E}\left[\left(H_k(X_{k-1}, \varepsilon_k)\right)^p\right]\right)^{\frac{2}{p}} \\
&\leq \left(K_{n-1}(\rho)\right)^2 \left(\mathbb{E}\left[\left(G_{X_1}(X_1)\right)^p\right]\right)^{\frac{2}{p}} + (p-1) \sum_{k=2}^{n} \left(K_{n-k}(\rho)\right)^2 \left(\mathbb{E}\left[\left(H_k(X_{k-1}, \varepsilon_k)\right)^p\right]\right)^{\frac{2}{p}} \\
&\leq A_p(n, \rho),
\end{aligned}
$$

which gives the desired inequality. $\qquad\square$

15

5.2 Rosenthal's inequality

Under the same assumptions as in the previous subsection, one can prove the following Rosenthal-type inequality.

Proposition 5.2. *Let $p \geq 2$, and consider the variables $L_k(X_{k-1})$ defined in (4.18). If (5.37) holds, then there exists a constant C_p depending only on p such that*

$$\|S_n\|_p^p \leq C_p \left(\mathbb{E}\left[\left(\left(K_{n-1}(\rho)\right)^2 A_1(2) + \sum_{k=2}^n \left(K_{n-k}(\rho)\right)^2 L_k(X_{k-1}) \right)^{p/2} \right] + \sum_{k=1}^n \left(K_{n-k}(\rho)\right)^p A_k(p) \right). \tag{5.40}$$

Remark 5.2. Assume that F_n satisfies (3.5). Then, according to Remark 3.1, it follows from the proof of Proposition 5.2 that inequality (5.40) remains true if the second condition of (5.37) is replaced by (5.39).

Proof. From point 2 of Proposition 3.1, it is easy to see that

$$\sum_{k=1}^n \mathbb{E}[M_k^2 | \mathcal{F}_{k-1}] \leq \left(K_{n-1}(\rho)\right)^2 \mathbb{E}[\left(G_{X_1}(X_1)\right)^2] + \sum_{k=2}^n \left(K_{n-k}(\rho)\right)^2 L_k(X_{k-1}),$$

and that

$$\sum_{k=1}^n \mathbb{E}[|M_k|^p] \leq \left(K_{n-1}(\rho)\right)^p \mathbb{E}[\left(G_{X_1}(X_1)\right)^p] + \sum_{k=2}^n \left(K_{n-k}(\rho)\right)^p \mathbb{E}\left[\left(H_k(X_{k-1}, \varepsilon_k)\right)^p \right]$$

$$\leq \left(K_{n-1}(\rho)\right)^p A_1(p) + \sum_{k=2}^n \left(K_{n-k}(\rho)\right)^p A_k(p).$$

The desired inequality is then a direct consequence of Rosenthal's inequality for martingales (see for instance Theorem 2.12 of Hall and Heyde [20]). $\qquad\square$

5.3 von Bahr-Esseen's inequality

In this subsection, we assume that the dominating random variables $G_{X_1}(X_1)$ and $H_k(X_{k-1}, \varepsilon_k)$ have a moment of order $p \in (1, 2]$.

Proposition 5.3. *Let $p \in (1, 2]$, and assume that (5.37) holds. Then*

$$\|S_n\|_p^p \leq A_p(n, \rho), \tag{5.41}$$

where

$$A_p(n, \rho) = \left(K_{n-1}(\rho)\right)^p A_1(p) + 2^{2-p} \sum_{k=2}^n \left(K_{n-k}(\rho)\right)^p A_k(p). \tag{5.42}$$

Remark 5.3. Assume that F_n satisfies (3.5). Then, according to Remark 3.1, it follows from the proof of Proposition 5.3 that inequality (5.41) remains true if the second condition of (5.37) is replaced by (5.39)

Proof. Using an improvement of the von Bahr-Esseen inequality (see inequality (1.11) in Pinelis [26]), we have

$$\|S_n\|_p^p \leq \|M_1\|_p^p + \tilde{C}_p \sum_{k=2}^n \|M_k\|_p^p,$$

where the constant $\tilde{C}_p$ is described in Proposition 1.8 of Pinelis [26], and is such that $\tilde{C}_p \leq 2^{2-p}$ for any $p \in [1, 2]$. By Proposition 3.1, it follows that

$$\|S_n\|_p^p \leq \left(K_{n-1}(\rho)\right)^p \mathbb{E}\left[\left(G_{X_1}(X_1)\right)^p \right] + \tilde{C}_p \sum_{k=2}^n \left(K_{n-k}(\rho)\right)^p \mathbb{E}\left[\left(H_k(X_{k-1}, \varepsilon_k)\right)^p \right]$$

$$\leq \left(K_{n-1}(\rho)\right)^p A_1(p) + \tilde{C}_p \sum_{k=2}^n \left(K_{n-k}(\rho)\right)^p A_k(p),$$

which is the desired inequality. $\qquad\square$

References

[1] Adamczak, R., 2008. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.* 13: 1000–1034.

[2] Alquier, P., Doukhan, P., 2011. Sparsity considerations for dependent variables. *Electron. J. Statist.* 5: 750–774.

[3] Bickel, P. J., Ritov, Y., Tsybakov, A., 2009. Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* 37: 1705–1732.

[4] Bardet, J.-M., Doukhan, P., 2018. Non-parametric estimation of time varying AR(1)–processes with local stationarity and periodicity. *Electron. J. Statist.* 12: 2323–2354.

[5] Brockwell, P.J., Davis, R.A., 1987. Time series: theory and methods. Springer Science & Business Media.

[6] Cox, P.M., Betts, R.A., Jones, C.D., Spall, S.A., Totterdell, I.J., 2000. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature* 408: 184–187.

[7] Cuny, C., Dedecker, J., Merlevède, F., 2017. Large and moderate deviations for the left random walk on $GL_d(\mathbb{R})$. *ALEA, Lat. Am. J. Probab. Math. Stat.* 14: 503–527.

[8] Dahlhaus, R., Richter, S., Wu, W.B., 2019. Towards a general theory for non-linear locally stationary processes. *Bernoulli* 25(2): 1013-1044.

[9] Diaconis, P., Freedman, D., 1999. Iterated random functions. *SIAM Rev.* 41: 45–76.

[10] Dedecker, J., Fan, X., 2015. Deviation inequalities for separately Lipschitz functionals of iterated random functions. *Stochastic Process. Appl.* 125(1): 60–90.

[11] Dedecker, J., Gouëzel, S., 2015. Subgaussian concentration inequalities for geometrically ergodic Markov chains. *Electron. Commun. Probab.* 20: 12 pp.

[12] Djellout, H., Guillin, A., Wu, L., 2004. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Probab.* 32: 2702–2732.

[13] Doukhan, P., Fokianos, K., Tjøstheim, D., 2012. On weak dependence conditions for Poisson autoregressions. *Statist. Probab. Lett.* 82: 942–948.

[14] Fan, X., Grama, I., Liu, Q., 2012. Hoeffding's inequality for supermartingales. *Stochastic Process. Appl.* 122(10): 3545–3559.

[15] Fan, X., Grama, I., Liu, Q., 2012. Large deviation exponential inequalities for supermartingales. *Electron. Commun. Probab.* 17(59): 1–8.

[16] Fan, X., Grama, I., Liu, Q., 2017. Deviation inequalities for martingales with applications. *J. Math. Anal. Appl.* 448(1): 538–566.

[17] Ferland, R., Latour, A., Oraichi, D., 2006. Integer-valued GARCH processes. *J. Time Series Anal.* 27: 923–942.

[18] Freund, Y., Mansour, Y., Schapire, R.E., 2004. Generalization bounds for averaged classifiers. *Ann. Statist.* 32: 1698–1722.

[19] Fuk, D.Kh., 1973. Some probabilitic inequalties for martingales. *Sib. Math. J.* 14: 185–193.

[20] Hall, P., Heyde, C.C., 1980. *Martingale Limit Theory and its Applications.* Academic Press, New York.

[21] Massart, P., 2003. *Concentration Inequalities and Model Selection.* Ecole d'Eté de Probabilités de Saint-Flour XXXIII. Springer-Verlag.

[22] Lanzinger, H., Stadtmüller, U., 2000. Maxima of increments of partial sums for certain subexponential distributions. *Stochastic Process. Appl.* 86(2): 307–322.

[23] McDiarmid, C., 1989. On the method of bounded differences. *Surveys of combinatorics. London Math. Soc. Lectures Notes Ser.* 141: 148–188.

[24] Milman, V.D., Schechtman, G., 1986. *Asymptotic Theory of Finite Dimensional Normed Spaces: Isometric inequalities in Riemaniann manifolds*. Lecture Notes in Mathematics 1200, Springer-Verlag.

[25] Nagaev, S.V., 1979. Large deviations of sums of independent radom variables. *Ann. Probab.* 7: 745–789.

[26] Pinelis, I., 2015. Best possible bounds of the von Bahr-Esseen type. *Ann. Funct. Anal.* 6: 1–29.

[27] Rio, E., 2000. Inégalités de Hoeffding pour les fonctions Lipschitziennes de suites dépendantes. *C. R. Acad. Sci. Paris Sér. I Math.* 330: 905–908.

[28] Rio, E., 2009. Moment inequalities for sums of dependent random variables under projective condition. *J. Theoret. Probab.* 22: 146–163.

[29] Rio, E., 2017. About the constant in the Fuk-Nagaev inequalities. *Electron. Commun. Probab.* 22(28): 12 pp.

[30] Rosenblatt, M., 1956. A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. USA* 42: 43–47.

[31] Samson, P.M., 2000. Concentration of measure inequalities for Markov chains an ϕ-mixing processes. *Ann. Probab.* 28: 416–461

[32] Yurinskii, V.V., 1974. Exponential bounds for large deviations. *Theory Probab. Appl.* 19: 154–155.

[33] Zhang, D., Wu, W.B., 2017. Gaussian approximation for high dimensional time series. *Ann. Statist.* 45(5): 1895–1919.