

McDiarmid’s martingale for a class of iterated random functions

Jérôme Dedecker* and Xiequan Fan†

Abstract

We consider an \mathcal{X} -valued Markov chain X_1, X_2, \dots, X_n belonging to a class of iterated random functions, which is “one-step contracting” with respect to some distance d on \mathcal{X} . If f is any separately Lipschitz function with respect to d , we use a well known decomposition of $S_n = f(X_1, \dots, X_n) - \mathbb{E}[f(X_1, \dots, X_n)]$ into a sum of martingale differences d_k with respect to the natural filtration \mathcal{F}_k . We show that each difference d_k is bounded by a random variable η_k independent of \mathcal{F}_{k-1} . Using this very strong property, we obtain a large variety of deviation inequalities for S_n , which are governed by the distribution of the η_k ’s. Finally, we give an application of these inequalities to the Wasserstein distance between the empirical measure and the invariant distribution of the chain.

Keywords. Iterated random functions, martingales, exponential inequalities, moment inequalities, Wasserstein distances.

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

1 A class of iterated random functions

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let (\mathcal{X}, d) and (\mathcal{Y}, δ) be two complete separable metric spaces. Let $(\varepsilon_i)_{i \geq 1}$ be a sequence of independent and identically distributed (iid) \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

$$X_n = F(X_{n-1}, \varepsilon_n), \quad \text{for } n \geq 2, \quad (1.1)$$

where $F : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{X}$ is such that

$$\mathbb{E}[d(F(x, \varepsilon_1), F(x', \varepsilon_1))] \leq \rho d(x, x') \quad (1.2)$$

for some $\rho \in [0, 1)$, and

$$d(F(x, y), F(x, y')) \leq C\delta(y, y') \quad (1.3)$$

for some $C > 0$.

This class of Markov chains, that we call “one-step contracting”, is very restrictive, but still contains a lot of pertinent examples. Among them, in the case where \mathcal{X} is a separable Banach space with norm $|\cdot|$, let us cite the functional auto-regressive model

$$X_n = f(X_{n-1}) + g(\xi_n),$$

where $f : \mathcal{X} \rightarrow \mathcal{X}$ and $g : \mathcal{Y} \rightarrow \mathcal{X}$ are such that

$$|f(x) - f(x')| \leq \rho|x - x'| \quad \text{and} \quad |g(y) - g(y')| \leq C\delta(y, y').$$

*Université Paris Descartes, Sorbonne Paris Cité, Laboratoire MAP5 and CNRS UMR 8145, 75016 Paris, France.

†Regularity Team, Inria and MAS Laboratory, Ecole Centrale Paris - Grande Voie des Vignes, 92295 Châtenay-Malabry, France.

We refer to the paper by Diaconis and Freedman [10] for many other interesting examples. Note also that this class of Markov chains contains the iid sequence $X_i = \varepsilon_i$, by taking $\mathcal{Y} = \mathcal{X}$ and $F(x, y) = y$ (note that $\rho = 0$ in that case).

This class possesses the property of exponential forgetting of the starting point: If X_n^x is the chain starting from $X_1 = x$, then one has

$$\mathbb{E}[d(X_n^x, X_n^{x'})] \leq \rho^n d(x, x').$$

Hence it has a unique stationary distribution μ (see for instance Theorem 1 in Diaconis and Freedman [10]), meaning that if X_1 is distributed as μ , then the chain $(X_i)_{i \geq 1}$ is strictly stationary. Moreover, one can easily prove that, if $(X_i)_{i \geq 1}$ is strictly stationary, then, for any $(x_0, y_0) \in \mathcal{X} \times \mathcal{Y}$, and any positive measurable function H ,

$$\mathbb{E}[H(d(X_n, x_0))] \leq \mathbb{E}\left[H\left(\sum_{i=0}^{\infty} \rho^i (d(F(x_0, y_0), x_0) + C\delta(\varepsilon_{i+1}, y_0))\right)\right]. \quad (1.4)$$

Although the one-step contraction is a very restrictive condition, this class of iterated random functions contains a lot of non Harris-recurrent Markov chains. For instance, if $\mathcal{X} = \mathcal{Y} = [0, 1]$ the chain

$$X_n = \frac{1}{2}(X_{n-1} + \varepsilon_n)$$

with X_1 uniformly distributed over $[0, 1]$, and $\varepsilon_i \sim \mathcal{B}(1/2)$ is strictly stationary, but it is not mixing in the sense of Rosenblatt [29].

The class of iterated random function satisfying (1.2) has been studied in Section 3.1 of Djellout *et al.* [11] (as a particular case of a general class of Markov chains which are contracting with respect to Wasserstein distances, see their Condition C1). Combining McDiarmid method and a result by Bobkov and Götze [2], Djellout *et al.* [11] proved in their Proposition 3.1 a subgaussian bound for separately Lipschitz functionals of the chain provided

$$\sup_{x \in \mathcal{X}} \mathbb{E}\left[\exp\left(a(d(F(x, \varepsilon_1), F(x, \varepsilon_2)))^2\right)\right] < \infty, \quad (1.5)$$

for some $a > 0$. Because of the supremum in x , this condition is quite delicate to check. However, if (1.3) holds, it is implied by the simple condition

$$\mathbb{E}\left[\exp\left(a(C\delta(\varepsilon_1, \varepsilon_2))^2\right)\right] < \infty.$$

As we shall see in Section 2, this is due to the fact that the martingale differences from McDiarmid's decomposition are bounded by a random variable η_k independent of $\mathcal{F}_{k-1} = \sigma(X_1, \dots, X_{k-1})$. From this simple remark, we can obtain many deviation inequalities for separately Lipschitz functionals of the chain by applying known inequalities for martingales.

A more restrictive class of iterated random function, satisfying (1.3) and the one-step contraction

$$d(F(x, y), F(x', y)) \leq \rho d(x, x'),$$

has been studied by Delyon *et al.* [9] when $\mathcal{X} = \mathbb{R}^\ell$ and $\mathcal{Y} = \mathbb{R}^k$. These authors have proved a moderate deviation principle for additive and Lipschitz functionals of the chain, under a condition on the Laplace transform of the euclidean norm of ε_i .

2 McDiarmid's martingale

2.1 Separately Lipschitz functionals of X_1, \dots, X_n .

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

$$|f(x_1, x_2, \dots, x_n) - f(x'_1, x'_2, \dots, x'_n)| \leq d(x_1, x'_1) + \dots + d(x_n, x'_n). \quad (2.1)$$

Let then

$$S_n := f(X_1, \dots, X_n) - \mathbb{E}[f(X_1, \dots, X_n)]. \quad (2.2)$$

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

$$g_k(X_1, \dots, X_k) = \mathbb{E}[f(X_1, \dots, X_n) | \mathcal{F}_k], \quad (2.3)$$

and

$$d_k = g_k(X_1, \dots, X_k) - g_{k-1}(X_1, \dots, X_{k-1}). \quad (2.4)$$

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

$$S_k := d_1 + d_2 + \dots + d_k,$$

and note that, by definition of the d_k 's, the functional S_n introduced in (2.2) satisfies

$$S_n = d_1 + d_2 + \dots + d_n.$$

Hence S_k is a martingale adapted to the filtration \mathcal{F}_k . This representation was introduced by McDiarmid [21] in the iid case, when $X_i = \varepsilon_i$ (see also Yurinskii [31] in a different context).

The following Proposition collects some interesting properties of the functions g_k and of the martingale differences d_k .

Proposition 2.1. *For $k \in \mathbb{N}$ and ρ in $[0, 1)$, let $K_k(\rho) = (1 - \rho^{k+1})/(1 - \rho) = 1 + \rho + \dots + \rho^k$. Let $(X_i)_{i \geq 1}$ be a Markov chain satisfying (1.1) for some function F satisfying (1.2). Let g_k and d_k be defined by (2.3) and (2.4) respectively.*

1. *The function g_k is separately Lipschitz and such that*

$$|g_k(x_1, x_2, \dots, x_k) - g_k(x'_1, x'_2, \dots, x'_k)| \leq d(x_1, x'_1) + \dots + d(x_{k-1}, x'_{k-1}) + K_{n-k}(\rho)d(x_k, x'_k).$$

2. *Let P_{X_1} be the distribution of X_1 and P_ε be the common distribution of the ε_k 's. Let G_{X_1} and H_ε be the two functions defined by*

$$G_{X_1}(x) = \int d(x, x')P_{X_1}(dx') \quad \text{and} \quad H_\varepsilon(x, y) = \int d(F(x, y), F(x, y'))P_\varepsilon(dy').$$

Then, the martingale difference d_k is such that

$$|d_1| \leq K_{n-1}(\rho)G_{X_1}(X_1) \quad \text{and for } k \in [2, n], \quad |d_k| \leq K_{n-k}(\rho)H_\varepsilon(X_{k-1}, \varepsilon_k).$$

3. *Assume moreover that F satisfies (1.3), and Let G_ε be the function defined by*

$$G_\varepsilon(y) = \int C\delta(y, y')P_\varepsilon(dy').$$

Then $H_\varepsilon(x, y) \leq G_\varepsilon(y)$, and consequently, for $k \in [2, n]$,

$$|d_k| \leq K_{n-k}(\rho)G_\varepsilon(\varepsilon_k).$$

Remark 2.1. Let us comment on the point 3 of Proposition 2.1. The fact that the martingale difference d_k is bounded by the random variable $K_{n-k}(\rho)G_\varepsilon(\varepsilon_k)$ which is *independent of \mathcal{F}_{k-1}* is crucial. It explains why we shall obtain deviations inequalities for S_n under some conditions on the distribution of $G_\varepsilon(\varepsilon_k)$ (typically conditions on the Laplace transform, or moment conditions).

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

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

It follows that

$$\begin{aligned} & |g_{k-1}(x_1, x_2, \dots, x_{k-1}) - g_{k-1}(x'_1, x'_2, \dots, x'_{k-1})| \\ & \leq \int |g_k(x_1, x_2, \dots, F(x_{k-1}, y)) - g_k(x'_1, x'_2, \dots, F(x'_{k-1}, y))| P_\varepsilon(dy). \end{aligned} \quad (2.5)$$

Now, by assumption and condition (1.2),

$$\begin{aligned} & \int |g_k(x_1, x_2, \dots, F(x_{k-1}, y)) - g_k(x'_1, x'_2, \dots, F(x'_{k-1}, y))| P_\varepsilon(dy) \\ & \leq d(x_1, x'_1) + \dots + d(x_{k-1}, x'_{k-1}) + K_{n-k}(\rho) \int d(F(x_{k-1}, y), F(x'_{k-1}, y)) P_\varepsilon(dy) \\ & \leq d(x_1, x'_1) + \dots + (1 + \rho K_{n-k}(\rho)) d(x_{k-1}, x'_{k-1}) \\ & \leq d(x_1, x'_1) + \dots + K_{n-k+1}(\rho) d(x_{k-1}, x'_{k-1}). \end{aligned} \quad (2.6)$$

The point 1 follows from (2.5) and (2.6).

Let us prove the point 2. First note that

$$|d_1| = \left| g_1(X_1) - \int g_1(x) P_{X_1}(dx) \right| \leq K_{n-1}(\rho) \int d(X_1, x) P_{X_1}(dx) = K_{n-1}(\rho) G_{X_1}(X_1).$$

In the same way, for $k \geq 2$,

$$\begin{aligned} |d_k| &= |g_k(X_1, \dots, X_k) - \mathbb{E}[g_k(X_1, \dots, X_k) | \mathcal{F}_{k-1}]| \\ &\leq \int |g_k(X_1, \dots, F(X_{k-1}, \varepsilon_k)) - g_k(X_1, \dots, F(X_{k-1}, y))| P_\varepsilon(dy) \\ &\leq K_{n-k}(\rho) \int d(F(X_{k-1}, \varepsilon_k), F(X_{k-1}, y)) P_\varepsilon(dy) = K_{n-k}(\rho) H_\varepsilon(X_{k-1}, \varepsilon_k). \end{aligned}$$

The point 3 is clear, since if (1.3) is true, then

$$H_\varepsilon(x, y) = \int d(F(x, y), F(x, y')) P_\varepsilon(dy') \leq \int C \delta(y, y') P_\varepsilon(dy') = G_\varepsilon(y).$$

The proof of the proposition is now complete. \square

2.2 An important remark

For any $\alpha \in (0, 1)$ define the distances d_α and δ_α on \mathcal{X} and \mathcal{Y} respectively by

$$d_\alpha(x, x') = (d(x, x'))^\alpha \quad \text{and} \quad \delta_\alpha(y, y') = (\delta(y, y'))^\alpha.$$

If F is one-step contracting with respect to a natural distance d (meaning that it satisfies the inequalities (1.2) and (1.3) with $\rho \in [0, 1)$ and $C > 0$ respectively), then for any $\alpha \in (0, 1)$,

$$\mathbb{E}[(d_\alpha(F(x, \varepsilon_1), F(x', \varepsilon_1)))] \leq \rho^\alpha d_\alpha(x, x'), \quad (2.7)$$

and

$$d_\alpha(F(x, y), F(x, y')) \leq C^\alpha \delta_\alpha(y, y'). \quad (2.8)$$

Hence F is also one-step contracting for the distance d_α , with the new constants $\rho^\alpha \in [0, 1)$ and $C^\alpha > 0$. Consequently, Proposition 2.1 applies to the martingale

$$S_n = f(X_1, \dots, X_n) - \mathbb{E}[f(X_1, \dots, X_n)],$$

where f is separately Lipschitz with respect to d_α . The dominating random variables $G_{X_1, \alpha}(X_1)$ and $G_{\varepsilon, \alpha}(\varepsilon_k)$ are then defined by

$$G_{X_1, \alpha}(x) = \int d_\alpha(x, x') P_{X_1}(dx') \quad \text{and} \quad G_{\varepsilon, \alpha}(y) = \int C^\alpha \delta_\alpha(y, y') P_\varepsilon(dy').$$

Hence, all the results of the following section apply to the functional S_n , provided the corresponding conditions on the dominating random variables $G_{X_1, \alpha}(X_1)$ and $G_{\varepsilon, \alpha}(\varepsilon_k)$ are satisfied.

For instance, if $\mathcal{X} = \mathbb{R}^\ell$ and $d(x, y) = \|x - y\|$ is the euclidean distance on \mathbb{R}^ℓ , then one can consider the class of separately Hölder functions f such that

$$|f(x_1, x_2, \dots, x_n) - f(x'_1, x'_2, \dots, x'_n)| \leq \|x_1 - x'_1\|^\alpha + \dots + \|x_n - x'_n\|^\alpha.$$

3 Deviation inequalities for the functional S_n .

Let $(X_i)_{i \geq 1}$ be a Markov chain satisfying (1.1) for some function F satisfying (1.2) and (1.3). In this section, we apply inequalities for martingales to bound up the deviation of the functional S_n defined by (2.2). Some of these inequalities are direct applications of known inequalities, some deserve a short proof and some other are new.

Note that deviation inequalities for Lipschitz functions of dependent sequences have been proved for instance by Rio [26], Collet *et al.* [6], Djellout *et al.* [11], Kontorovich and Ramanan [18], and Chazottes and Gouëzel [5] among others. Except for Djellout *et al.* [11] (who also consider more general Markov chains), the examples studied by these authors are different from the class described in the present paper. For instance, the Markov chains associated to the maps studied by Chazottes and Gouëzel [5] do not in general satisfy the one step contraction property.

The interest of the one step contraction is that, thanks to Proposition 2.1, we shall obtain very precise inequalities, with precise constants depending on the distribution of the dominating random variables $G_{X_1}(X_1)$ and $G_\varepsilon(\varepsilon_k)$.

Let us note that, in the iid case, when $X_i = \varepsilon_i$, the additive functional

$$f(x_1, x_2, \dots, x_n) = \sum_{k=1}^n G_\varepsilon(x_k)$$

is of course separately Lipschitz and satisfies (2.1). Hence, the inequalities of the following section apply to this simple functional, under the usual moment or Laplace conditions on the (non centered) variables $G_\varepsilon(\varepsilon_i)$. This shows that, in the iid case, these inequalities cannot be much improved without additional assumptions on the functional f .

Let us now consider the case where we only assume that F satisfies (1.2). Then all the inequalities of this section will be true provided the appropriate conditions of the type $\mathbb{E}[f(G_\varepsilon(\varepsilon))] \leq C$ for some positive measurable function f are replaced by

$$\sup_{k \in [2, n]} \left\| \mathbb{E} \left[f(H_\varepsilon(X_{k-1}, \varepsilon_k)) \middle| X_{k-1} \right] \right\|_\infty \leq C. \quad (3.1)$$

Note that the latter condition is true provided

$$\sup_{x \in \mathcal{X}} \mathbb{E} \left[f(H_\varepsilon(x, \varepsilon_1)) \right] \leq C,$$

which is of the same type as condition (1.5) for the subgaussian bound (with $f(x) = \exp(ax^2)$ in that particular case). Recall that condition (1.5) is due to Djellout *et al.* [11] (see their Proposition 3.1).

For the weak and strong moment bounds on S_n , we shall see in Subsections 3.7, 3.8 and 3.9 that condition (3.1) can be replaced by an appropriate moment condition on $H_\varepsilon(X_{k-1}, \varepsilon_k)$.

To conclude the introduction of this section, let us note that the deviation inequalities of Subsections 3.1 – 3.6 are given for $\mathbb{P}(\pm S_n > x)$, but we shall only prove them for S_n . The proofs of the deviation inequalities for $-S_n$ are exactly the same, the upper bounds of points 2 and 3 of Proposition 2.1 being valid for d_k and $-d_k$.

In all this section, $G_\varepsilon(\varepsilon)$ denotes a random variable distributed as $G_\varepsilon(\varepsilon_k)$.

3.1 Bernstein type bound

Under the conditional Bernstein condition, van de Geer [16] and De La Peña [7] have obtained some tight Bernstein type inequalities for martingales. Applying Proposition 2.1, we obtain the following proposition.

Proposition 3.1. *Assume that there exist some constants $M > 0$, $V_1 \geq 0$ and $V_2 \geq 0$ such that, for any integer $k \geq 2$,*

$$\mathbb{E}\left[\left(G_{X_1}(X_1)\right)^k\right] \leq \frac{k!}{2}V_1M^{k-2} \text{ and } \mathbb{E}\left[\left(G_\varepsilon(\varepsilon)\right)^k\right] \leq \frac{k!}{2}V_2M^{k-2}. \quad (3.2)$$

Let

$$V = V_1\left(K_{n-1}(\rho)\right)^2 + V_2\sum_{k=2}^n\left(K_{n-k}(\rho)\right)^2 \quad \text{and} \quad \delta = MK_{n-1}(\rho).$$

Then, for any $t \in [0, \delta^{-1}]$,

$$\mathbb{E}\left[e^{\pm tS_n}\right] \leq \exp\left(\frac{t^2V}{2(1-t\delta)}\right). \quad (3.3)$$

Consequently, for any $x > 0$,

$$\mathbb{P}(\pm S_n \geq x) \leq \exp\left(\frac{x^2}{V(1 + \sqrt{1 + 2x\delta/V}) + x\delta}\right) \quad (3.4)$$

$$\leq \exp\left(\frac{x^2}{2(V + x\delta)}\right). \quad (3.5)$$

Remark 3.1. Let us comment on condition (3.2).

1. In the iid case, when $X_i = \varepsilon_i$, condition (3.2) is the Bernstein condition

$$\mathbb{E}\left[\left(G_\varepsilon(\varepsilon)\right)^k\right] \leq \frac{k!}{2}VM^{k-2}.$$

In that case the inequalities (3.4) and (3.5) hold with $\rho = 0$.

2. Since $G_\varepsilon(\varepsilon) \leq C\delta(\varepsilon, y_0) + C\mathbb{E}[\delta(\varepsilon, y_0)]$, it follows that

$$\mathbb{E}\left[\left(G_\varepsilon(\varepsilon)\right)^k\right] \leq 2^k\mathbb{E}\left[\left(C\delta(\varepsilon, y_0)\right)^k\right].$$

Hence, the condition

$$\mathbb{E}\left[\left(C\delta(\varepsilon, y_0)\right)^k\right] \leq \frac{k!}{2}A(y_0)B(y_0)^{k-2} \quad (3.6)$$

implies the second condition in (3.2) with $V_2 = 4A(y_0)$ and $M = 2B(y_0)$. In the same way, the condition

$$\mathbb{E}\left[\left(d(X_1, x_0)\right)^k\right] \leq \frac{k!}{2}C(x_0)D(x_0)^{k-2} \quad (3.7)$$

implies the first condition in (3.2) with $V_1 = 4C(x_0)$ and $M = 2D(x_0)$.

3. Consider the chain with non random starting point $X_1 = x$. Then $G_{X_1}(X_1) = 0$, and the first condition in (3.2) holds with $V_1 = 0$.
4. Let us consider now the case where X_1 is distributed according to the invariant probability measure μ . We shall see that in that case (3.7) follows from (3.6). To avoid to many computations, assume that one can find (x_0, y_0) such that $d(F(x_0, y_0), x_0) = 0$, which is true in many cases. If (3.6) holds, it follows from (1.4) applied to $H(x) = x^k$ that (3.7) holds with $C(x_0) = (1 - \rho)^{-2}A(y_0)$ and $D(x_0) = (1 - \rho)^{-1}B(y_0)$. According to the point 2 of this remark, condition (3.2) is satisfied by taking $M = 2(1 - \rho)^{-1}B(y_0)$, $V_2 = 4A(y_0)$ and $V_1 = 4(1 - \rho)^{-2}A(y_0)$.

Proof. From Proposition 2.1 and condition (3.2), it is easy to see that, for any $t \in [0, \delta^{-1})$,

$$\begin{aligned}
\mathbb{E}[e^{td_1}] &\leq 1 + \sum_{i=2}^{\infty} \frac{t^i}{i!} \mathbb{E}[(d_1)^i] \\
&\leq 1 + \sum_{i=2}^{\infty} \frac{t^i}{i!} \mathbb{E}[|d_1|^i] \\
&\leq 1 + \sum_{i=2}^{\infty} \frac{t^i}{i!} (K_{n-1}(\rho))^i \mathbb{E}\left[(G_{X_1}(X_1))^i\right] \\
&\leq 1 + \sum_{i=2}^{\infty} \frac{t^i}{i!} (K_{n-1}(\rho))^{i!} \frac{V_1 M^{i-2}}{2} = 1 + \frac{t^2 V_1 (K_{n-1}(\rho))^2}{2(1-t\delta)}. \tag{3.8}
\end{aligned}$$

Similarly, for any $k \in [2, n]$,

$$\mathbb{E}[e^{td_k} | \mathcal{F}_{k-1}] \leq 1 + \frac{t^2 V_2 (K_{n-k}(\rho))^2}{2(1-t\delta)}. \tag{3.9}$$

Using the inequality $1+t \leq e^t$, we find that, for any $t \in [0, \delta^{-1})$,

$$\mathbb{E}[e^{td_1}] \leq \exp\left(\frac{t^2 V_1 (K_{n-1}(\rho))^2}{2(1-t\delta)}\right) \tag{3.10}$$

and

$$\mathbb{E}[e^{td_k} | \mathcal{F}_{k-1}] \leq \exp\left(\frac{t^2 V_2 (K_{n-k}(\rho))^2}{2(1-t\delta)}\right). \tag{3.11}$$

By the tower property of conditional expectation, it follows that, for any $t \in [0, \delta^{-1})$,

$$\begin{aligned}
\mathbb{E}[e^{tS_n}] &= \mathbb{E}[\mathbb{E}[e^{tS_n} | \mathcal{F}_{n-1}]] \\
&= \mathbb{E}[e^{tS_{n-1}} \mathbb{E}[e^{td_n} | \mathcal{F}_{n-1}]] \\
&\leq \mathbb{E}[e^{tS_{n-1}}] \exp\left(\frac{t^2 V_2}{2(1-t\delta)}\right) \\
&\leq \exp\left(\frac{t^2 V}{2(1-t\delta)}\right),
\end{aligned}$$

which gives inequality (3.3). Using the exponential Markov inequality, we deduce that, for any $x \geq 0$ and $t \in [0, \delta^{-1})$,

$$\begin{aligned}
\mathbb{P}(S_n \geq x) &\leq \mathbb{E}[e^{t(S_n-x)}] \\
&\leq \exp\left(-tx + \frac{t^2 V}{2(1-t\delta)}\right). \tag{3.12}
\end{aligned}$$

The minimum is reached at

$$t = t(x) := \frac{2x/V}{2x\delta/V + 1 + \sqrt{1 + 2x\delta/V}}.$$

Substituting $t = t(x)$ in (3.12), we obtain the desired inequalities

$$\begin{aligned}
\mathbb{P}(S_n \geq x) &\leq \exp\left(\frac{x^2}{V(1 + \sqrt{1 + 2x\delta/V}) + x\delta}\right) \\
&\leq \exp\left(\frac{x^2}{2(V + x\delta)}\right),
\end{aligned}$$

where the last line follows from the inequality $\sqrt{1 + 2x\delta/V} \leq 1 + x\delta/V$. \square

3.2 Cramér type bound

If the Laplace transform of the dominating random variables $G_{X_1}(X_1)$ and $G_\varepsilon(\varepsilon_k)$ satisfy the Cramér condition, we obtain the following proposition similar to that of Liu and Watbled [20] under the conditional Cramér condition. For the optimal convergence speed of martingales under the Cramér condition, we refer to Lesigne of Volný [19] and Fan *et al.* [13].

Proposition 3.2. *Assume that there exist some constants $a > 0$, $K_1 \geq 1$ and $K_2 \geq 1$ such that*

$$\mathbb{E}\left[\exp\left(aG_{X_1}(X_1)\right)\right] \leq K_1 \quad \text{and} \quad \mathbb{E}\left[\exp\left(aCG_\varepsilon(\varepsilon)\right)\right] \leq K_2. \quad (3.13)$$

Let

$$K = \frac{2}{e^2} \left(K_1 + K_2 \sum_{i=2}^n \left(\frac{K_{n-i}(\rho)}{K_{n-1}(\rho)} \right)^2 \right) \quad \text{and} \quad \delta = \frac{a}{K_{n-1}(\rho)}.$$

Then, for any $t \in [0, \delta)$,

$$\mathbb{E}[e^{\pm tS_n}] \leq \exp\left(\frac{t^2 K \delta^{-2}}{1 - t\delta^{-1}}\right).$$

Consequently, for any $x > 0$,

$$\mathbb{P}(\pm S_n \geq x) \leq \exp\left(\frac{(x\delta)^2}{2K(1 + \sqrt{1 + x\delta/K}) + x\delta}\right) \quad (3.14)$$

$$\leq \exp\left(\frac{(x\delta)^2}{4K + 2x\delta}\right). \quad (3.15)$$

Remark 3.2. Let us comment on condition (3.13).

1. In the iid case, when $X_i = \varepsilon_i$, the condition (3.13) writes simply

$$\mathbb{E}\left[\exp\left(aG_\varepsilon(\varepsilon)\right)\right] \leq K.$$

In that case the inequalities (3.14) and (3.15) hold with $\rho = 0$.

2. Since $G_\varepsilon(\varepsilon) \leq C\delta(\varepsilon, y_0) + C\mathbb{E}(\delta(\varepsilon, y_0))$ the condition

$$\mathbb{E}\left[\exp\left(aC\delta(\varepsilon, y_0)\right)\right] \leq A(y_0) \quad (3.16)$$

implies the second condition in (3.13) with $K_2 = A(y_0) \exp(aC\mathbb{E}[\delta(\xi, y_0)]) \leq A(y_0)^2$. In the same way, the condition

$$\mathbb{E}\left[\exp\left(ad(X_1, x_0)\right)\right] \leq B(x_0) \quad (3.17)$$

implies the first condition in (3.13) with $K_1 = B(x_0) \exp(a\mathbb{E}[d(X_1, x_0)]) \leq B(x_0)^2$.

3. Consider the chain with non random starting point $X_1 = x$. Then $G_{X_1}(X_1) = 0$, and the first condition in (3.13) holds with $K_1 = 1$.
4. Let us consider now the case where X_1 is distributed according to the invariant probability measure μ . We shall see that in that case (3.17) follows from (3.16). Indeed, if (3.16) holds, it follows from (1.4) applied to $H(x) = \exp(ax)$ that

$$\mathbb{E}\left[\exp\left(ad(X_1, x_0)\right)\right] \leq \exp\left(\frac{a}{1-\rho}d(F(x_0, y_0), x_0)\right) \prod_{i=0}^{\infty} (A(y_0))^{\rho^i}.$$

Hence

$$\mathbb{E}\left[\exp\left(ad(X_1, x_0)\right)\right] \leq \exp\left(\frac{a}{1-\rho}d(F(x_0, y_0), x_0)\right) (A(y_0))^{1/(1-\rho)}$$

and (3.17) is true with

$$B(x_0) = (A(y_0))^{1/(1-\rho)} \exp\left(\frac{a}{1-\rho} d(F(x_0, y_0), x_0)\right).$$

According to the point 2 of this remark, condition (3.13) is satisfied by taking $K_2 = (A(y_0))^2$ and $K_1 = (B(x_0))^2$. In particular, if (3.16) holds, and if we can find (x_0, y_0) such that $d(F(x_0, y_0), x_0) = 0$, then one can take $K_1 = (A(y_0))^{2/(1-\rho)}$.

Proof. Let $\delta = a/K_{n-1}(\rho)$. Since $\mathbb{E}[d_1] = 0$, it is easy to see that, for any $t \in [0, \delta)$,

$$\begin{aligned} \mathbb{E}[e^{td_1}] &= 1 + \sum_{i=2}^{\infty} \frac{t^i}{i!} \mathbb{E}[(d_1)^i] \\ &\leq 1 + \sum_{i=2}^{\infty} \left(\frac{t}{\delta}\right)^i \mathbb{E}\left[\frac{1}{i!} |\delta d_1|^i\right]. \end{aligned} \quad (3.18)$$

Here, let us note that, for $t \geq 0$,

$$\begin{aligned} \frac{t^i}{i!} e^{-t} &\leq \frac{i^i}{i!} e^{-i} \\ &\leq 2e^{-2}, \quad \text{for } i \geq 2, \end{aligned} \quad (3.19)$$

where the last line follows from the fact that $i^i e^{-i}/i!$ is decreasing in i . Note that the equality in (3.19) is reached at $t = i = 2$. Using (3.19), Proposition 2.1 and condition (3.13), we have

$$\begin{aligned} \mathbb{E}\left[\frac{1}{i!} |\delta d_1|^i\right] &\leq 2e^{-2} \mathbb{E}[e^{\delta |d_1|}] \\ &\leq 2e^{-2} \mathbb{E}\left[\exp\left(aG_{X_1}(X_1)\right)\right] \\ &\leq 2e^{-2} K_1. \end{aligned} \quad (3.20)$$

Combining the inequalities (3.18) and (3.20) together, we obtain, for any $t \in [0, \delta)$,

$$\mathbb{E}[e^{td_1}] \leq 1 + \sum_{n=2}^{\infty} \frac{2}{e^2} \left(\frac{t}{\delta}\right)^n K_1 = 1 + \frac{2}{e^2} \frac{t^2 K_1 \delta^{-2}}{1 - t\delta^{-1}} \leq \exp\left(\frac{2}{e^2} \frac{t^2 K_1 \delta^{-2}}{1 - t\delta^{-1}}\right). \quad (3.21)$$

Similarly, since $K_{n-i}(\rho)/K_{n-1}(\rho) \leq 1$ for all $i \in [2, n]$, we have, for any $t \in [0, \delta)$,

$$\mathbb{E}[e^{td_i} | \mathcal{F}_{i-1}] \leq \exp\left(\frac{2}{e^2} \frac{t^2 K_2 \delta^{-2}}{1 - t\delta^{-1}} \left(\frac{K_{n-i}(\rho)}{K_{n-1}(\rho)}\right)^2\right). \quad (3.22)$$

By the tower property of conditional expectation, it follows that, for any $t \in [0, \delta)$,

$$\begin{aligned} \mathbb{E}[e^{tS_n}] &= \mathbb{E}[\mathbb{E}[e^{tS_n} | \mathcal{F}_{n-1}]] \\ &= \mathbb{E}[e^{tS_{n-1}} \mathbb{E}[e^{td_n} | \mathcal{F}_{n-1}]] \\ &\leq \mathbb{E}[e^{tS_{n-1}}] \exp\left(\frac{2}{e^2} \frac{t^2 K_2 \delta^{-2}}{1 - t\delta^{-1}}\right) \\ &\leq \exp\left(\frac{t^2 K \delta^{-2}}{1 - t\delta^{-1}}\right), \end{aligned} \quad (3.23)$$

where

$$K = \frac{2}{e^2} \left(K_1 + K_2 \sum_{i=2}^n \left(\frac{K_{n-i}(\rho)}{K_{n-1}(\rho)}\right)^2 \right).$$

Then using the exponential Markov inequality, we deduce that, for any $x \geq 0$ and $t \in [0, \delta)$,

$$\begin{aligned} \mathbb{P}(S_n \geq x) &\leq \mathbb{E}[e^{t(S_n - x)}] \\ &\leq \exp\left(-tx + \frac{t^2 K \delta^{-2}}{1 - t\delta^{-1}}\right). \end{aligned} \quad (3.24)$$

The minimum is reached at

$$t = t(x) := \frac{x\delta^2/K}{x\delta/K + 1 + \sqrt{1 + x\delta/K}}.$$

Substituting $t = t(x)$ in (3.24), we obtain the desired inequalities (3.14) and (3.15). \square

3.3 Qualitative results when $\mathbb{E}[e^{a(G_\varepsilon(\varepsilon))^p}] < \infty$ for $p > 1$.

The next proposition follows easily from Theorem 3.2 of Liu and Watbled [20].

Proposition 3.3. *Let $p > 1$. Assume that there exist some constants $a > 0$, $K_1 \geq 1$ and $K_2 \geq 1$ such that*

$$\mathbb{E}\left[\exp\left(a(G_{X_1}(X_1))^p\right)\right] \leq K_1 \quad \text{and} \quad \mathbb{E}\left[\exp\left(a(G_\varepsilon(\varepsilon))^p\right)\right] \leq K_2. \quad (3.25)$$

Let q be the conjugate exponent of p and let $\tau > 0$ be such that

$$(q\tau)^{\frac{1}{q}} (pa)^{\frac{1}{p}} (1 - \rho) = 1.$$

Then, for any $\tau_1 > \tau$, there exist some positive numbers t_1, x_1, A, B , depending only on a, ρ, K_1, K_2, p and τ_1 , such that

$$\mathbb{E}[e^{\pm t S_n}] \leq \begin{cases} \exp(n\tau_1 t^q) & \text{if } t \geq t_1 \\ \exp(nAt^2) & \text{if } t \in [0, t_1] \end{cases} \quad (3.26)$$

and

$$\mathbb{P}(\pm S_n \geq x) \leq \begin{cases} \exp(-a_1 x^p / n^{p-1}) & \text{if } x \geq nx_1 \\ \exp(-Bx^2/n) & \text{if } x \in [0, nx_1], \end{cases} \quad (3.27)$$

where a_1 is such that $(q\tau)^{1/q} (pa_1)^{1/p} = 1$.

Remark 3.3. Assume that (3.25) is satisfied for some $p \geq 1$. From Proposition 3.2 (case $p = 1$) and Proposition 3.3 (case $p > 1$), we infer that for any $x > 0$, one can find a positive constant c_x not depending on n such that

$$\mathbb{P}(\pm S_n \geq nx) \leq \exp(-c_x n). \quad (3.28)$$

Moreover, for x large enough, one can take $c_x = a_1 x^p$.

Proof. By condition (3.25) and Proposition 2.1, it follows that

$$\mathbb{E}\left[\exp\left(a(1 - \rho)^p |d_1|^p\right)\right] \leq \mathbb{E}\left[\exp\left(a(G_{X_1}(X_1))^p\right)\right] \leq K_2$$

and, for all $i \in [2, n]$,

$$\mathbb{E}\left[\exp\left(a(1 - \rho)^p |d_i|^p\right) \mid \mathcal{F}_{i-1}\right] \leq \mathbb{E}\left[\exp\left(a(G_\varepsilon(\varepsilon))^p\right)\right] \leq K_1.$$

Let $q > 1$ and $\tau > 0$ be such that

$$\frac{1}{p} + \frac{1}{q} = 1 \quad \text{and} \quad (q\tau)^{\frac{1}{q}} (pa)^{\frac{1}{p}} (1 - \rho) = 1.$$

Then, by Theorem 3.2 of Liu and Watbled [20], for any $\tau_1 > \tau$, there exist $t_1, x_1, A, B > 0$, depending only on a, ρ, K_1, K_2, p and τ_1 , such that the claim of Proposition 3.3 holds. \square

In particular, if $p = 2$, we have the following sub-Gaussian bound.

Proposition 3.4. *Assume that there exist some constants $a > 0$, $K_1 \geq 1$ and $K_2 \geq 1$ such that*

$$\mathbb{E}\left[\exp\left(a(G_{X_1}(X_1))^2\right)\right] \leq K_1 \quad \text{and} \quad \mathbb{E}\left[\exp\left(a(G_\varepsilon(\varepsilon))^2\right)\right] \leq K_2. \quad (3.29)$$

Then, there exists a constant $c > 0$ depending only on a, ρ, K_1 and K_2 such that

$$\mathbb{E}[e^{\pm t S_n}] \leq \exp(n c t^2) \quad \text{for all } t > 0. \quad (3.30)$$

Consequently, for any $x > 0$,

$$\mathbb{P}(\pm S_n \geq x) \leq \exp\left(-\frac{x^2}{4nc}\right). \quad (3.31)$$

Remark 3.4. As quoted at the beginning of Section 3, if F satisfies only (1.2), Proposition 3.4 holds provided (3.1) is satisfied with $f(x) = \exp(ax^2)$. This condition is implied by condition (1.5), which is due to Djellout *et al* [11].

Proof. Inequality (3.30) follows directly from (3.26). Using the exponential Markov inequality, we deduce that for any $x, t \geq 0$,

$$\begin{aligned} \mathbb{P}(S_n \geq x) &\leq \mathbb{E}[e^{t(S_n - x)}] \\ &\leq \exp(-tx + n c t^2). \end{aligned} \quad (3.32)$$

The minimum is reached at $t = t(x) := x/(2nc)$. Substituting $t = t(x)$ in (3.32), we obtain the desired inequality (3.31). \square

3.4 Semi-exponential bound

In the case where $G_{X_1}(X_1)$ and $G_\varepsilon(\varepsilon)$ have semi-exponential moments, the following proposition holds. This proposition can be compared to the corresponding results in Borovkov [4] for partial sums of independent random variables, Merlevède *et al.* [22] for partial sums of weakly dependent sequences, and Fan *et al.* [13] for martingales.

Proposition 3.5. *Let $p \in (0, 1)$. Assume that there exist some positive constants K_1 and K_2 such that*

$$\mathbb{E}\left[(G_{X_1}(X_1))^2 \exp\left((G_{X_1}(X_1))^p\right)\right] \leq K_1 \quad \text{and} \quad \mathbb{E}\left[(G_\varepsilon(\varepsilon))^2 \exp\left((G_\varepsilon(\varepsilon))^p\right)\right] \leq K_2. \quad (3.33)$$

Set

$$K = K_1 + K_2 \sum_{i=2}^n \left(\frac{K_{n-i}(\rho)}{K_{n-1}(\rho)}\right)^2.$$

Then, for any $0 \leq x < K^{1/(2-p)}$,

$$\begin{aligned} \mathbb{P}(\pm S_n \geq x) &\leq \exp\left(-\frac{x^2}{2K(K_{n-1}(\rho))^2}\right) \\ &\quad + (K_{n-1}(\rho))^2 \left(\frac{x^2}{K^{1+p}}\right)^{1/(1-p)} \exp\left(-\left(\frac{K}{x(K_{n-1}(\rho))^{1-p}}\right)^{p/(1-p)}\right) \end{aligned} \quad (3.34)$$

and, for any $x \geq K^{1/(2-p)}$,

$$\begin{aligned} \mathbb{P}(\pm S_n \geq x) &\leq \exp\left(-\left(\frac{x}{K_{n-1}(\rho)}\right)^p \left(1 - \frac{K}{2} \left(\frac{K_{n-1}(\rho)}{x}\right)^{2-p}\right)\right) \\ &\quad + K \left(\frac{K_{n-1}(\rho)}{x}\right)^2 \exp\left(-\left(\frac{x}{K_{n-1}(\rho)}\right)^p\right). \end{aligned} \quad (3.35)$$

Remark 3.5. In particular, there exists a positive constant c such that, for any $x > 0$,

$$\mathbb{P}(\pm S_n \geq nx) \leq C_x \exp(-c x^p n^p), \quad (3.36)$$

where the constants C_x and c do not depend on n .

Remark 3.6. By a simple comparison, we find that for moderate $x \in (0, K^{1/(2-p)})$, the second item in the right hand side of (3.34) is less than the first one. Thus for moderate $x \in (0, K^{1/(2-p)})$, the bound (3.34) is a sub-Gaussian bound and is of the order

$$\exp\left(-\frac{x^2}{2K(K_{n-1}(\rho))^2}\right). \quad (3.37)$$

For all $x \geq K^{1/(2-p)}$, bound (3.35) is a semi-exponential bound and is of the order

$$\exp\left(-\frac{1}{2}\left(\frac{x}{K_{n-1}(\rho)}\right)^p\right). \quad (3.38)$$

Moreover, when $x/K^{1/(2-p)} \rightarrow \infty$, the constant $\frac{1}{2}$ in (3.38) can be improved to $1+\varepsilon$ for any given $\varepsilon > 0$.

Proof. The proof is based on a truncation argument. For given $y > 0$, set $\eta_i = d_i \mathbf{1}_{\{d_i \leq y\}}$. Then $(\eta_i, \mathcal{F}_i)_{i=1, \dots, n}$ is a sequence of supermartingale differences. Using a two term Taylor's expansion, we have, for all $t > 0$,

$$e^{t\eta_i} \leq 1 + t\eta_i + \frac{t^2 \eta_i^2}{2} e^{t\eta_i}.$$

Since $p \in (0, 1)$, it follows that

$$\eta_i^+ = d_i \mathbf{1}_{\{0 \leq d_i \leq y\}} \leq y \frac{d_i^p}{y^p} \mathbf{1}_{\{0 \leq d_i \leq y\}} \leq y^{1-p} (\eta_i^+)^p.$$

Hence,

$$e^{t\eta_i} \leq 1 + t\eta_i + \frac{t^2 \eta_i^2}{2} \exp\left(ty^{1-p} (\eta_i^+)^p\right).$$

Since $\mathbb{E}[\eta_i | \mathcal{F}_{i-1}] \leq \mathbb{E}[d_i | \mathcal{F}_{i-1}] = 0$, it follows that, for all $t > 0$,

$$\mathbb{E}[e^{t\eta_i} | \mathcal{F}_{i-1}] \leq 1 + \frac{t^2}{2} \mathbb{E}[\eta_i^2 \exp\left(ty^{1-p} (\eta_i^+)^p\right) | \mathcal{F}_{i-1}].$$

By Proposition 2.1, it follows that, for all $t > 0$,

$$\mathbb{E}[e^{t\eta_1}] \leq 1 + \frac{t^2}{2} \mathbb{E}\left[\left(K_{n-1}(\rho) G_{X_1}(X_1)\right)^2 \exp\left(ty^{1-p} \left(K_{n-1}(\rho) G_{X_1}(X_1)\right)^p\right)\right]$$

and similarly, for $i \in [2, n]$,

$$\mathbb{E}[e^{t\eta_i} | \mathcal{F}_{i-1}] \leq 1 + \frac{t^2}{2} \mathbb{E}\left[\left(K_{n-i}(\rho) G_\varepsilon(\varepsilon)\right)^2 \exp\left(ty^{1-p} \left(K_{n-i}(\rho) G_\varepsilon(\varepsilon)\right)^p\right)\right].$$

Taking $t = y^{p-1}/(K_{n-1}(\rho))^p$, by condition (3.33) and $K_{n-i}(\rho)/K_{n-1}(\rho) \leq 1$, we find that

$$\begin{aligned} \mathbb{E}[e^{t\eta_1}] &\leq 1 + \frac{1}{2} \left(\frac{y}{K_{n-1}(\rho)}\right)^{2p-2} \mathbb{E}\left[\left(G_{X_1}(X_1)\right)^2 \exp\left(\left(G_{X_1}(X_1)\right)^p\right)\right] \\ &\leq 1 + \frac{1}{2} \left(\frac{y}{K_{n-1}(\rho)}\right)^{2p-2} K_1 \\ &\leq \exp\left(\frac{1}{2} \left(\frac{y}{K_{n-1}(\rho)}\right)^{2p-2} K_1\right) \end{aligned}$$

and, for $i \in [2, n]$,

$$\begin{aligned} \mathbb{E}[e^{t\eta_i} | \mathcal{F}_{i-1}] &\leq 1 + \frac{1}{2} \left(\frac{y}{K_{n-1}(\rho)}\right)^{2p-2} \left(\frac{K_{n-i}(\rho)}{K_{n-1}(\rho)}\right)^2 \mathbb{E}\left[\left(G_\varepsilon(\varepsilon)\right)^2 \exp\left(\left(G_\varepsilon(\varepsilon)\right)^p\right)\right] \\ &\leq 1 + \frac{1}{2} \left(\frac{y}{K_{n-1}(\rho)}\right)^{2p-2} K_2 \left(\frac{K_{n-i}(\rho)}{K_{n-1}(\rho)}\right)^2 \\ &\leq \exp\left(\frac{1}{2} \left(\frac{y}{K_{n-1}(\rho)}\right)^{2p-2} K_2 \left(\frac{K_{n-i}(\rho)}{K_{n-1}(\rho)}\right)^2\right). \end{aligned}$$

Hence, by the tower property of conditional expectation, it follows that

$$\begin{aligned}
\mathbb{E} \left[e^{t \sum_{i=1}^n \eta_i} \right] &= \mathbb{E} \left[\mathbb{E} \left[e^{t \sum_{i=1}^n \eta_i} \mid \mathcal{F}_{n-1} \right] \right] \\
&= \mathbb{E} \left[e^{t \sum_{i=1}^{n-1} \eta_i} \mathbb{E} \left[e^{t \eta_n} \mid \mathcal{F}_{n-1} \right] \right] \\
&\leq \mathbb{E} \left[e^{t \sum_{i=1}^{n-1} \eta_i} \right] \exp \left(\frac{1}{2} \left(\frac{y}{K_{n-1}(\rho)} \right)^{2p-2} K_2 \left(\frac{1}{K_{n-1}(\rho)} \right)^2 \right) \\
&\leq \exp \left(\frac{1}{2} \left(\frac{y}{K_{n-1}(\rho)} \right)^{2p-2} K \right), \tag{3.39}
\end{aligned}$$

where

$$K = K_1 + K_2 \sum_{i=2}^n \left(\frac{K_{n-i}(\rho)}{K_{n-1}(\rho)} \right)^2.$$

It is easy to see that

$$\begin{aligned}
\mathbb{P}(S_n \geq x) &\leq \mathbb{P} \left(\sum_{i=1}^n \eta_i \geq x \right) + \mathbb{P} \left(\sum_{i=1}^n d_i \mathbf{1}_{\{d_i > y\}} > 0 \right) \\
&\leq \mathbb{P} \left(\sum_{i=1}^n \eta_i \geq x \right) + \mathbb{P} \left(\max_{1 \leq i \leq n} d_i > y \right) =: P_5 + \mathbb{P} \left(\max_{1 \leq i \leq n} d_i > y \right). \tag{3.40}
\end{aligned}$$

For the first item of (3.40), by the exponential Markov's inequality and (3.39), we have

$$P_5 \leq \mathbb{E} \left[e^{t(\sum_{i=1}^n \eta_i - x)} \right] \leq \exp \left(-tx + \frac{1}{2} \left(\frac{y}{K_{n-1}(\rho)} \right)^{2p-2} K \right). \tag{3.41}$$

For the second item of (3.40), we have the following estimation:

$$\begin{aligned}
\mathbb{P} \left(\max_{1 \leq i \leq n} d_i > y \right) &\leq \sum_{i=1}^n \mathbb{P}(d_i > y) \\
&\leq \sum_{i=1}^n \mathbb{P} \left(\frac{d_i}{K_{n-1}(\rho)} > \frac{y}{K_{n-1}(\rho)} \right) \\
&\leq \frac{\exp \left(- \left(\frac{y}{K_{n-1}(\rho)} \right)^p \right)}{\left(\frac{y}{K_{n-1}(\rho)} \right)^2} \sum_{i=1}^n \mathbb{E} \left[\left(\frac{d_i}{K_{n-1}(\rho)} \right)^2 e^{|d_i/K_{n-1}(\rho)|^p} \right].
\end{aligned}$$

By Proposition 2.1 and $K_{n-i}(\rho)/K_{n-1}(\rho) \leq 1$ again, it is easy to see that

$$\begin{aligned}
&\sum_{i=1}^n \mathbb{E} \left[\left(\frac{d_i}{K_{n-1}(\rho)} \right)^2 e^{|d_i/K_{n-1}(\rho)|^p} \right] \\
&\leq \mathbb{E} \left[(G_{X_1}(X_1))^2 \exp \left((G_{X_1}(X_1))^p \right) \right] + \left(\frac{K_{n-i}(\rho)}{K_{n-1}(\rho)} \right)^2 \sum_{i=2}^n \mathbb{E} \left[(G_\varepsilon(\varepsilon))^2 \exp \left((G_\varepsilon(\varepsilon))^p \right) \right] \leq K.
\end{aligned}$$

Thus

$$\mathbb{P} \left(\max_{1 \leq i \leq n} d_i > y \right) \leq \frac{K}{\left(\frac{y}{K_{n-1}(\rho)} \right)^2} \exp \left(- \left(\frac{y}{K_{n-1}(\rho)} \right)^p \right). \tag{3.42}$$

Combining (3.40), (3.41) and (3.42) together, it is to see that

$$\begin{aligned}
\mathbb{P}(S_n \geq x) &\leq \exp \left(-tx + \frac{1}{2} \left(\frac{y}{K_{n-1}(\rho)} \right)^{2p-2} K \right) \\
&\quad + \frac{K}{\left(\frac{y}{K_{n-1}(\rho)} \right)^2} \exp \left(- \left(\frac{y}{K_{n-1}(\rho)} \right)^p \right). \tag{3.43}
\end{aligned}$$

Recall that $t = y^{p-1}/(K_{n-1}(\rho))^p$. Taking

$$y = \begin{cases} (K/x)^{1/(1-p)} & \text{if } 0 \leq x < K^{1/(2-p)}, \\ x & \text{if } x \geq K^{1/(2-p)}, \end{cases} \quad (3.44)$$

we obtain the desired inequalities. \square

3.5 McDiarmid inequality

In this section, we consider the case where the increments d_k are bounded. We shall use an improved version of the well known inequality by McDiarmid, which has been recently stated by Rio [28]. For this inequality, we do not assume that (1.3) holds. Hence, Proposition 3.6 applies to any Markov chain $X_n = F(X_{n-1}, \varepsilon_n)$, for F satisfying (1.2).

As in Rio [28], let

$$\ell(t) = (t - \ln t - 1) + t(e^t - 1)^{-1} + \ln(1 - e^{-t}) \quad \text{for all } t > 0,$$

and let

$$\ell^*(x) = \sup_{t>0} (xt - \ell(t)) \quad \text{for all } x > 0,$$

be the Young transform of $\ell(t)$. As quoted by Rio [28], the following inequalities hold

$$\ell^*(x) \geq (x^2 - 2x) \ln(1 - x) \geq 2x^2 + x^4/6. \quad (3.45)$$

Let also $(X'_1, (\varepsilon'_i)_{i \geq 2})$ be an independent copy of $(X_1, (\varepsilon_i)_{i \geq 2})$.

Proposition 3.6. *Assume that there exist some positive constants M_k such that*

$$\|d(X_1, X'_1)\|_\infty \leq M_1 \quad \text{and} \quad \|d(F(X_{k-1}, \varepsilon_k), F(X_{k-1}, \varepsilon'_k))\|_\infty \leq M_k \quad \text{for } k \in [2, n]. \quad (3.46)$$

Let

$$M^2(n, \rho) = \sum_{k=1}^n (K_{n-k}(\rho) M_k)^2 \quad \text{and} \quad D(n, \rho) = \sum_{k=1}^n K_{n-k}(\rho) M_k.$$

Then, for any $t \geq 0$,

$$\mathbb{E}[e^{\pm t S_n}] \leq \exp\left(\frac{D^2(n, \rho)}{M^2(n, \rho)} \ell\left(\frac{M^2(n, \rho) x}{D(n, \rho)}\right)\right) \quad (3.47)$$

and, for any $x \in [0, D(n, \rho)]$,

$$\mathbb{P}(\pm S_n > x) \leq \exp\left(-\frac{D^2(n, \rho)}{M^2(n, \rho)} \ell^*\left(\frac{x}{D(n, \rho)}\right)\right). \quad (3.48)$$

Consequently, for any $x \in [0, D(n, \rho)]$,

$$\mathbb{P}(\pm S_n > x) \leq \left(\frac{D(n, \rho) - x}{D(n, \rho)}\right)^{\frac{2D(n, \rho)x - x^2}{M^2(n, \rho)}}. \quad (3.49)$$

Remark 3.7. Since $(x^2 - 2x) \ln(1 - x) \geq 2x^2$, inequality (3.49) implies the following McDiarmid inequality

$$\mathbb{P}(\pm S_n > x) \leq \exp\left(-\frac{2x^2}{M^2(n, \rho)}\right).$$

Remark 3.8. Taking $\Delta(n, \rho) = K_{n-1}(\rho) \max_{1 \leq k \leq n} M_k$, we obtain the upper bound: for any $x \in [0, n\Delta(n, \rho)]$,

$$\mathbb{P}(\pm S_n > x) \leq \exp\left(-n\ell^*\left(\frac{x}{n\Delta(n, \rho)}\right)\right) \leq \exp\left(-\frac{2x^2}{n\Delta^2(n, \rho)}\right).$$

Remark 3.9. If F satisfies (1.3), then one can take $M_1 = \|d(X_1, X'_1)\|_\infty$ and $M_k = C\|\delta(\varepsilon_1, \varepsilon'_1)\|_\infty$ for $k \in [2, n]$.

Proof. Let

$$u_{k-1}(x_1, \dots, x_{k-1}) = \operatorname{ess\,inf}_{\varepsilon_k} g_k(x_1, \dots, F(x_{k-1}, \varepsilon_k))$$

and

$$v_{k-1}(x_1, \dots, x_{k-1}) = \operatorname{ess\,sup}_{\varepsilon_k} g_k(x_1, \dots, F(x_{k-1}, \varepsilon_k))$$

From the proof of Proposition 2.1, it follows that

$$u_{k-1}(X_1, \dots, X_{k-1}) \leq d_k \leq v_{k-1}(X_1, \dots, X_{k-1}).$$

By Proposition 2.1 and condition (3.46), we have

$$v_{k-1}(X_1, \dots, X_{k-1}) - u_{k-1}(X_1, \dots, X_{k-1}) \leq K_{n-k}(\rho)M_k.$$

Now, following exactly the proof of Theorem 3.1 of Rio [28] with $\Delta_k = K_{n-k}(\rho)M_k$ we obtain the inequalities (3.47) and (3.48). Since $\ell^*(x) \geq (x^2 - 2x)\ln(1-x)$, inequality (3.49) follows from (3.48). \square

3.6 Fuk-Nagaev type bound

The next proposition follows easily from Corollary 2.3 of Fan *et al.* [12].

Proposition 3.7. *Assume that there exist two positive constants V_1 and V_2 such that*

$$\mathbb{E}[(G_{X_1}(X_1))^2] \leq V_1 \quad \text{and} \quad \mathbb{E}[(G_\varepsilon(\varepsilon))^2] \leq V_2.$$

Let

$$V = V_1(K_{n-1}(\rho))^2 + V_2 \sum_{i=2}^n (K_{n-i}(\rho))^2. \quad (3.50)$$

Then, for any $x, y > 0$,

$$\mathbb{P}(\pm S_n > x) \leq H_n\left(\frac{x}{yK_{n-1}(\rho)}, \frac{\sqrt{V}}{yK_{n-1}(\rho)}\right) + \mathbb{P}\left(\max\left\{G_{X_1}(X_1), \max_{2 \leq i \leq n} G_\varepsilon(\varepsilon_i)\right\} > y\right), \quad (3.51)$$

where

$$H_n(x, v) = \left\{ \left(\frac{v^2}{x+v^2}\right)^{x+v^2} \left(\frac{n}{n-x}\right)^{n-x} \right\}^{\frac{n}{n+v^2}} \mathbf{1}_{\{x \leq n\}} \quad (3.52)$$

with the convention that $(+\infty)^0 = 1$ (which applies when $x = n$).

Proof. We apply Corollary 2.3 of Fan *et al.* [12] with the truncature level $yK_{n-1}(\rho)$. By Proposition 2.1, $|d_1| \leq K_{n-1}(\rho)G_{X_1}(X_1)$ and $|d_i| \leq K_{n-i}(\rho)G_\varepsilon(\varepsilon_i)$ for $i \in [2, n]$. Hence

$$\mathbb{E}[d_1^2 \mathbf{1}_{\{d_1 \leq yK_{n-1}(\rho)\}}] \leq (K_{n-1}(\rho))^2 \mathbb{E}[(G_{X_1}(X_1))^2] \leq (K_{n-1}(\rho))^2 V_1$$

and, for $i \in [2, n]$,

$$\mathbb{E}[d_i^2 \mathbf{1}_{\{d_i \leq yK_{n-1}(\rho)\}} | \mathcal{F}_{i-1}] \leq (K_{n-i}(\rho))^2 \mathbb{E}[(G_\varepsilon(\varepsilon))^2] \leq (K_{n-i}(\rho))^2 V_2.$$

It follows from Corollary 2.3 of Fan *et al.* [12] that

$$\mathbb{P}(S_n > x) \leq H_n\left(\frac{x}{yK_{n-1}(\rho)}, \frac{\sqrt{V}}{yK_{n-1}(\rho)}\right) + \mathbb{P}\left(\max_{1 \leq i \leq n} d_i > yK_{n-1}(\rho)\right).$$

Inequality (3.51) follows by applying Proposition 2.1 again. \square

In particular, if $G_{X_1}(X_1)$ and $G_\varepsilon(\varepsilon)$ are bounded, then Proposition 3.7 implies the following Hoeffding bound.

Proposition 3.8. Assume that there exist some positive constants M , V_1 and V_2 such that

$$G_{X_1}(X_1) \leq M, \quad G_\varepsilon(\varepsilon) \leq M, \quad \mathbb{E}[(G_{X_1}(X_1))^2] \leq V_1 \quad \text{and} \quad \mathbb{E}[(G_\varepsilon(\varepsilon))^2] \leq V_2.$$

Then, for any $x > 0$,

$$\mathbb{P}(\pm S_n > x) \leq H_n \left(\frac{x}{MK_{n-1}(\rho)}, \frac{\sqrt{V}}{MK_{n-1}(\rho)} \right), \quad (3.53)$$

where $H_n(x, v)$ and V are defined by (3.52) and (3.50), respectively.

Remark 3.10. According to Remark 2.1 of Fan *et al.* [12], for any $x \geq 0$ and any $v > 0$, it holds

$$H_n(x, v) \leq B(x, v) := \left(\frac{v^2}{x + v^2} \right)^{x+v^2} e^x \quad (3.54)$$

$$\leq B_1(x, v) := \exp \left\{ -\frac{x^2}{2(v^2 + \frac{1}{3}x)} \right\}. \quad (3.55)$$

Note that (3.54) and (3.55) are respectively known as Bennett's and Bernstein's bounds. Then, inequality (3.53) also implies Bennett's and Bernstein's bounds

$$\mathbb{P}(\pm S_n > x) \leq B \left(\frac{x}{MK_{n-1}(\rho)}, \frac{\sqrt{V}}{MK_{n-1}(\rho)} \right) \leq B_1 \left(\frac{x}{MK_{n-1}(\rho)}, \frac{\sqrt{V}}{MK_{n-1}(\rho)} \right).$$

We now consider the case where the random variables $G_{X_1}(X_1)$ and $G_\varepsilon(\varepsilon)$ 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

$$\|Z\|_{w,p}^p = \sup_{x>0} x^p \mathbb{P}(|Z| > x). \quad (3.56)$$

Proposition 3.9. Let $p > 2$. Assume that there exist some positive constants V_1 , V_2 , $A_1(p)$ and $A_2(p)$ such that

$$\begin{aligned} \mathbb{E}[(G_{X_1}(X_1))^2] &\leq V_1, & \mathbb{E}[(G_\varepsilon(\varepsilon))^2] &\leq V_2, \\ \|G_{X_1}(X_1)\|_{w,p}^p &\leq A_1(p) & \text{and} & \|G_\varepsilon(\varepsilon)\|_{w,p}^p &\leq A_2(p). \end{aligned}$$

Let V be defined by (3.50), and let

$$A(p) = A_1(p) + (n-1)A_2(p).$$

Then, for any $x, y > 0$,

$$\mathbb{P}(\pm S_n > x) \leq H_n \left(\frac{x}{yK_{n-1}(\rho)}, \frac{\sqrt{V}}{yK_{n-1}(\rho)} \right) + \frac{A(p)}{y^p}, \quad (3.57)$$

where $H_n(x, v)$ is defined by (3.52).

Remark 3.11. Assume that $G_{X_1}(X_1)$ and $G_\varepsilon(\varepsilon)$ have a weak moment of order $p > 2$. Taking

$$y = \frac{3nx}{2pK_{n-1}(\rho) \ln(n)}$$

in inequality (3.57), we infer that, for any $x > 0$,

$$\mathbb{P}(\pm S_n > nx) \leq \frac{C_x (\ln(n))^p}{n^{p-1}},$$

for some positive C_x not depending on n .

If the martingale differences d_i have p th moments ($p \geq 2$), then we have the following Fuk-type inequality (cf. Corollary 3' of Fuk [15]).

Proposition 3.10. *Let $p \geq 2$. Assume that there exist some positive constants $V_1, V_2, A_1(p)$ and $A_2(p)$ such that*

$$\begin{aligned} \mathbb{E}[(G_{X_1}(X_1))^2] &\leq V_1, & \mathbb{E}[(G_\varepsilon(\varepsilon))^2] &\leq V_2, \\ \mathbb{E}[(G_{X_1}(X_1))^p] &\leq A_1(p) & \text{and} & \mathbb{E}[(G_\varepsilon(\varepsilon))^p] &\leq A_2(p). \end{aligned} \quad (3.58)$$

Let V be defined by (3.50), and let

$$A(p) = A_1(p)(K_{n-1}(\rho))^p + A_2(p) \sum_{i=2}^n (K_{n-i}(\rho))^p.$$

Then, for any $x > 0$,

$$\mathbb{P}(|S_n| > x) \leq 2\left(1 + \frac{2}{p}\right)^p \frac{A(p)}{x^p} + 2 \exp\left(-\frac{2}{(p+2)^2 e^p} \frac{x^2}{V}\right). \quad (3.59)$$

Remark 3.12. Since $A(p)$ is of order n , it easy to see that the term

$$\exp\left(-\frac{2}{(p+2)^2 e^p} \frac{(nx)^2}{V}\right)$$

is decreasing at an exponential order, and that the term

$$2\left(1 + \frac{2}{p}\right)^p \frac{A(p)}{(nx)^p}$$

is of order n^{1-p} . Thus, for any $x > 0$ and all n ,

$$\mathbb{P}(|S_n| > nx) \leq \frac{C_x}{n^{p-1}},$$

for some positive C_x not depending on n . Note that the last inequality is optimal under the stated condition, even if S_n is a sum of iid random variables.

Proof. By Proposition 2.1 and condition (3.58), it follows that

$$\begin{aligned} \sum_{i=1}^n \mathbb{E}[|d_i|^p | \mathcal{F}_{i-1}] &\leq \mathbb{E}[|K_{n-1}(\rho)G_{X_1}(X_1)|^p] + \sum_{i=2}^n \mathbb{E}[|K_{n-i}(\rho)G_\varepsilon(\varepsilon_i)|^p] \\ &\leq (K_{n-1}(\rho))^p \mathbb{E}[|G_{X_1}(X_1)|^p] + \sum_{i=2}^n (K_{n-i}(\rho))^p \mathbb{E}[|G_\varepsilon(\varepsilon_i)|^p] \\ &\leq A_1(p)(K_{n-1}(\rho))^p + A_2(p) \sum_{i=2}^n (K_{n-i}(\rho))^p = A(p). \end{aligned}$$

Notice that $A(2) = V$. Using Corollary 3' of Fuk [15], we obtain the desired inequality. \square

3.7 von Bahr-Esseen bound

In the first proposition of this section, we assume that the dominating random variables $G_{X_1}(X_1)$ and $G_\varepsilon(\varepsilon_k)$ have only a moment of order $p \in [1, 2]$. For similar inequalities in the case where the X_i 's are independent, we refer to Pinelis [24].

Proposition 3.11. *Let $p \in [1, 2]$. Assume that*

$$\mathbb{E}\left[\left(G_{X_1}(X_1)\right)^p\right] \leq A_1(p) \quad \text{and} \quad \mathbb{E}\left[\left(G_\varepsilon(\varepsilon)\right)^p\right] \leq A_2(p). \quad (3.60)$$

Then

$$\|S_n\|_p \leq (A(n, \rho, p))^{1/p}, \quad (3.61)$$

where

$$A(n, \rho, p) = A_1(p)(K_{n-1}(\rho))^p + 2^{2-p} A_2(p) \sum_{k=2}^n (K_{n-k}(\rho))^p. \quad (3.62)$$

Remark 3.13. The constant 2^{2-p} in (3.62) can be replaced by the more precise constant \tilde{C}_p described in Proposition 1.8 of Pinelis [24].

Remark 3.14. Assume that F satisfies only (1.2). Then, it follows from the proof of Proposition 3.11 that the inequality (3.61) remains true if the second condition of (3.60) is replaced by

$$\sup_{k \in [2, n]} \mathbb{E} \left[\left(H_\varepsilon(X_{k-1}, \varepsilon_k) \right)^p \right] \leq A_2(p).$$

Proof. Using an improvement of the von Bahr-Esseen inequality (see inequality (1.11) in Pinelis [24]), we have

$$\|S_n\|_p^p \leq \|d_1\|_p^p + \tilde{C}_p \sum_{k=2}^n \|d_k\|_p^p,$$

where the constant \tilde{C}_p is described in Proposition 1.8 of Pinelis [24], and is such that $\tilde{C}_p \leq 2^{2-p}$ for any $p \in [1, 2]$. By Proposition 2.1, it follows that

$$\begin{aligned} \|S_n\|_p^p &\leq \left((K_{n-1}(\rho))^p \mathbb{E} \left[\left(G_{X_1}(X_1) \right)^p \right] + \tilde{C}_p \sum_{k=2}^n (K_{n-k}(\rho))^p \mathbb{E} \left[\left(G_\varepsilon(\varepsilon) \right)^p \right] \right) \\ &\leq \left(A_1(p)(K_{n-1}(\rho))^p + \tilde{C}_p A_2(p) \sum_{k=2}^n (K_{n-k}(\rho))^p \right), \end{aligned}$$

which gives the desired inequality. \square

We now consider the case where the variables $G_{X_1}(X_1)$ and $G_\varepsilon(\varepsilon_k)$ have only a weak moment of order $p \in (1, 2)$. Recall that the weak moment $\|Z\|_{w,p}^p$ has been defined by (3.56).

Proposition 3.12. *Let $p \in (1, 2)$. Assume that*

$$\left\| G_{X_1}(X_1) \right\|_{w,p}^p \leq A_1(p) \quad \text{and} \quad \left\| G_\varepsilon(\varepsilon) \right\|_{w,p}^p \leq A_2(p). \quad (3.63)$$

Then, for any $x > 0$,

$$\mathbb{P}(|S_n| > x) \leq \frac{C_p B(n, \rho, p)}{x^p}, \quad (3.64)$$

where

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

and

$$B(n, \rho, p) = A_1(p)(K_{n-1}(\rho))^p + A_2(p) \sum_{k=2}^n (K_{n-k}(\rho))^p.$$

Remark 3.15. Assume that F satisfies only (1.2). Then, it follows from the proof of Proposition 3.12 that the inequality (3.64) remains true if the second condition of (3.63) is replaced by

$$\sup_{k \in [2, n]} \left\| H_\varepsilon(X_{k-1}, \varepsilon_k) \right\|_{w,p}^p \leq A_2(p).$$

Proof. This proof is based on a truncation argument. For given $x > 0$, let

$$\begin{aligned} \xi_1 &= d_1 \mathbf{1}_{\{d_1 \leq x\}}, & \xi'_1 &= d_1 \mathbf{1}_{\{d_1 > x\}}, \\ \xi_k &= d_k \mathbf{1}_{\{d_k \leq x\}} \quad \text{and} \quad \xi'_k &= d_k \mathbf{1}_{\{d_k > x\}}. \end{aligned}$$

Define

$$\begin{aligned}\eta_1 &= \xi_1 - \mathbb{E}[\xi_1], & \eta'_1 &= \xi'_1 - \mathbb{E}[\xi'_1], \\ \eta_k &= \xi_k - \mathbb{E}[\xi_k | \mathcal{F}_{k-1}] \quad \text{and} \quad \eta'_k &= \xi'_k - \mathbb{E}[\xi'_k | \mathcal{F}_{k-1}].\end{aligned}$$

It is obvious that

$$\mathbb{P}(|S_n| > x) \leq \mathbb{P}\left(\left|\sum_{k=1}^n \eta_k\right| > \frac{x}{2}\right) + \mathbb{P}\left(\left|\sum_{k=1}^n \eta'_k\right| > \frac{x}{2}\right). \quad (3.65)$$

Applying Markov's inequality, we get

$$\mathbb{P}\left(\left|\sum_{k=1}^n \eta'_k\right| > \frac{x}{2}\right) \leq \frac{2}{x} \sum_{k=1}^n \|\eta'_k\|_1 \leq \frac{4}{x} \sum_{k=1}^n \|\xi'_k\|_1. \quad (3.66)$$

Recall that, if Z is any real-valued random variable such that

$$\mathbb{P}(|Z| > x) \leq H(x) \quad (3.67)$$

for a tail function H , then

$$\mathbb{E}(|Z| \mathbf{1}_{\{|Z| > a\}}) \leq \int_0^{H(a)} Q(u) du, \quad (3.68)$$

where Q is the cadlag inverse of H . Using Proposition 2.1, we have

$$\mathbb{P}(|d_k| > x) \leq H_k(x), \quad (3.69)$$

where $H_1(x) = \min\{1, x^{-p} A_1(p)(K_{n-1}(\rho))^p\}$ and $H_k(x) = \min\{1, x^{-p} A_2(p)(K_{n-k}(\rho))^p\}$ if $k \in [2, n]$. Hence, applying (3.68), we obtain

$$\|\xi'_1\|_1 \leq (A_1(p))^{1/p} K_{n-1}(\rho) \int_0^{H_1(x)} u^{-1/p} du \leq \frac{p}{p-1} A_1(p) (K_{n-1}(\rho))^p x^{1-p}. \quad (3.70)$$

Similarly, for $k \in [2, n]$,

$$\|\xi'_k\|_1 \leq (A_2(p))^{1/p} K_{n-k}(\rho) \int_0^{H_k(x)} u^{-1/p} du \leq \frac{p}{p-1} A_2(p) (K_{n-k}(\rho))^p x^{1-p}. \quad (3.71)$$

Consequently, from (3.66), (3.70) and (3.71),

$$\mathbb{P}\left(\left|\sum_{k=1}^n \eta'_k\right| > \frac{x}{2}\right) \leq \frac{4pB(n, \rho, p)}{(p-1)x^p}. \quad (3.72)$$

On the other hand, the η_k 's being martingales differences,

$$\mathbb{P}\left(\left|\sum_{k=1}^n \eta_k\right| > \frac{x}{2}\right) \leq \frac{4}{x^2} \sum_{k=1}^n \|\eta_k\|_2^2 \leq \frac{4}{x^2} \sum_{k=1}^n \|\xi_k\|_2^2 \quad (3.73)$$

Recall that, if Z is any real-valued random variable satisfying (3.67),

$$\mathbb{E}(Z^2 \mathbf{1}_{|Z| \leq a}) \leq \mathbb{E}((Z \wedge a)^2) \leq \int_0^1 \min\{Q^2(u), a^2\} du \leq 2 \int_{H(a)}^1 Q^2(u) du. \quad (3.74)$$

Using (3.69) and (3.74), we obtain

$$\|\xi_1\|_2^2 \leq 2(A_1(p))^{2/p} (K_{n-1}(\rho))^2 \int_{H_1(x)}^1 u^{-2/p} du \leq \frac{2p}{2-p} A_1(p) (K_{n-1}(\rho))^p x^{2-p}. \quad (3.75)$$

Similarly, for $k \in [2, n]$,

$$\|\xi_k\|_2^2 \leq 2(A_2(p))^{2/p} (K_{n-k}(\rho))^2 \int_{H_k(x)}^1 u^{-2/p} du \leq \frac{2p}{2-p} A_2(p) (K_{n-k}(\rho))^p x^{2-p}. \quad (3.76)$$

Consequently, from (3.73), (3.75) and (3.76),

$$\mathbb{P}\left(\left|\sum_{k=1}^n \eta_k\right| > \frac{x}{2}\right) \leq \frac{8pB(n, \rho, p)}{(p-2)x^p}. \quad (3.77)$$

Inequality (3.64) follows from (3.65), (3.72) and (3.77). \square

3.8 Marcinkiewicz-Zygmund bound

We now assume that the dominating random variables $G_{X_1}(X_1)$ and $G_\varepsilon(\varepsilon_k)$ have a moment of order $p \geq 2$.

Proposition 3.13. *Let $p \geq 2$. Assume that*

$$\mathbb{E}\left[\left(G_{X_1}(X_1)\right)^p\right] \leq A_1(p) \quad \text{and} \quad \mathbb{E}\left[\left(G_\varepsilon(\varepsilon)\right)^p\right] \leq A_2(p). \quad (3.78)$$

Then

$$\|S_n\|_p \leq \sqrt{A(n, \rho, p)}, \quad (3.79)$$

where

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

Remark 3.16. Assume that F satisfies only (1.2). Then, it follows from the proof of Proposition 3.13 that the inequality (3.79) remains true if the second condition of (3.78) is replaced by

$$\sup_{k \in [2, n]} \mathbb{E}\left[\left(H_\varepsilon(X_{k-1}, \varepsilon_k)\right)^p\right] \leq A_2(p).$$

Proof. Using Theorem 2.1 of Rio [27], we have

$$\|S_n\|_p^2 \leq \|d_1\|_p^2 + (p-1) \sum_{k=2}^n \|d_k\|_p^2.$$

By Proposition 2.1 and condition (3.78), it follows that

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

which gives the desired inequality. \square

3.9 Burkholder-Rosenthal bounds

When the dominating random variables $G_{X_1}(X_1)$ and $G_\varepsilon(\varepsilon_k)$ have a moment of order $p \geq 2$, one can prove the following proposition. For similar inequalities in the case where the X_i 's are independent, we refer to Pinelis [25].

Proposition 3.14. *Assume that there exist two constants $V_1 \geq 0$ and $V_2 \geq 0$ such that*

$$\mathbb{E}\left[\left(G_{X_1}(X_1)\right)^2\right] \leq V_1 \quad \text{and} \quad \mathbb{E}\left[\left(G_\varepsilon(\varepsilon)\right)^2\right] \leq V_2. \quad (3.80)$$

Let

$$V = V_1 (K_{n-1}(\rho))^2 + V_2 \sum_{k=2}^n (K_{n-k}(\rho))^2. \quad (3.81)$$

For any $p \geq 2$, there exist two positive constants $C_1(p)$ and $C_2(p)$ such that

$$\|S_n\|_p \leq C_1(p) \sqrt{V} + C_2(p) \left\| \max \left\{ K_{n-1}(\rho) G_{X_1}(X_1), \max_{2 \leq i \leq n} K_{n-i}(\rho) G_\varepsilon(\varepsilon_i) \right\} \right\|_p. \quad (3.82)$$

Remark 3.17. According to the proof of Theorem 4.1 of Pinelis [23], one can take $C_1(p) = 60c$ and $C_2(p) = 120\sqrt{ce^{p/c}}$ for any $c \in [1, p]$.

Proof. Applying Proposition 2.1, we have $|d_1| \leq K_{n-1}(\rho)G_{X_1}(X_1)$ and $|d_k| \leq K_{n-k}(\rho)G_\varepsilon(\varepsilon_k)$ for $k \in [2, n]$, and consequently

$$\mathbb{E}[d_1^2] \leq (K_{n-1}(\rho))^2 V_1 \quad \text{and} \quad \mathbb{E}[d_k^2 | \mathcal{F}_{k-1}] \leq (K_{n-k}(\rho))^2 V_2 \quad \text{for } k \in [2, n].$$

Then the proposition follows directly from Theorem 4.1 of Pinelis [23]. \square

We now consider the case where the random variables $G_{X_1}(X_1)$ and $G_\varepsilon(\varepsilon)$ have a weak moment of order $p > 2$. Recall that the weak moment $\|Z\|_{w,p}^p$ has been defined by (3.56).

Proposition 3.15. *Assume that (3.80) holds, and let V be defined by (3.81). Then, for any $p \geq 2$, there exist two positive constants $C_1(p)$ and $C_2(p)$ such that*

$$\mathbb{P}(|S_n| > t) \leq \frac{1}{t^p} \left\{ C_1(p)V^{p/2} + C_2(p) \left\| \max \left\{ K_{n-1}(\rho)G_{X_1}(X_1), \max_{2 \leq i \leq n} K_{n-i}(\rho)G_\varepsilon(\varepsilon_i) \right\} \right\|_{w,p}^p \right\}. \quad (3.83)$$

Remark 3.18. Assume that F satisfies only (1.2). Then, it follows from the proofs of Propositions 3.14 and 3.15 that the inequalities (3.82) and (3.83) remain true if the second condition of (3.80) is replaced by

$$\sup_{k \in [2, n]} \left\| \mathbb{E} \left[(H_\varepsilon(X_{k-1}, \varepsilon_k))^2 \middle| X_{k-1} \right] \right\|_{p/2} \leq V_2,$$

and by taking $H_\varepsilon(X_{k-1}, \varepsilon_k)$ instead of $G_\varepsilon(\varepsilon_k)$ in the second terms on right hand of (3.82) and (3.83).

Proof. It is the same as that of Proposition 3.14, by applying Theorem 6.3 in Chazottes and Gouézel [5]. \square

4 Application to the Wasserstein distance between the empirical distribution and the invariant distribution

4.1 Definition and upper bounds

Recall that the Wasserstein distance $W_1(\nu_1, \nu_2)$ between two probability measures ν_1, ν_2 on (\mathcal{X}, d) is defined by

$$W_1(\nu_1, \nu_2) = \inf_{\lambda \in M(\nu_1, \nu_2)} \int d(x, y) \lambda(dx, dy),$$

where $M(\nu_1, \nu_2)$ is the set of probability measures on $\mathcal{X} \times \mathcal{X}$ with margins ν_1 and ν_2 .

Let $\Lambda_1(\mathcal{X})$ be the set of functions from (\mathcal{X}, d) to \mathbb{R} such that

$$|g(x) - g(y)| \leq d(x, y).$$

Recall that $W_1(\nu_1, \nu_2)$ can be expressed *via* its dual form (see for instance the equality (5.11) in Villani [30])

$$W_1(\nu_1, \nu_2) = \sup_{g \in \Lambda_1(\mathcal{X})} |\nu_1(g) - \nu_2(g)|.$$

Let μ_n be the empirical distribution of the random variables X_1, X_2, \dots, X_n , that is

$$\mu_n = \frac{1}{n} \sum_{k=1}^n \delta_{X_k},$$

and let μ be the unique invariant distribution of the chain. It is easy to see that the function f defined by

$$nW_1(\mu_n, \mu) = f(X_1, X_2, \dots, X_n) := \sup_{g \in \Lambda_1(\mathcal{X})} \left| \sum_{i=1}^n (g(X_i) - \mu(g)) \right|,$$

is separately Lipschitz, and satisfies (2.1). Hence, all the inequalities of Section 3 apply to

$$S_n = nW_1(\mu_n, \mu) - n\mathbb{E}[W_1(\mu_n, \mu)].$$

Let us only give some qualitative consequences of these inequalities:

- If (3.25) holds for some $p \geq 1$, then there exist some positive constants A, B and C such that

$$\mathbb{P}\left(|W_1(\mu_n, \mu) - \mathbb{E}[W_1(\mu_n, \mu)]| > x\right) \leq \begin{cases} 2 \exp(-nAx^p) & \text{if } x \geq C \\ 2 \exp(-nBx^2) & \text{if } x \in [0, C]. \end{cases} \quad (4.1)$$

This follows from Proposition 3.2 (case $p = 1$) and Proposition 3.3 (case $p > 1$).

- If (3.33) holds for some $p \in (0, 1)$, then there exist some positive constants A, B, C, D and L such that

$$\mathbb{P}\left(|W_1(\mu_n, \mu) - \mathbb{E}[W_1(\mu_n, \mu)]| > x\right) \leq \begin{cases} C \exp(-n^p Ax^p) & \text{if } x \geq Ln^{-(1-p)/(2-p)} \\ D \exp(-nBx^2) & \text{if } x \in [0, Ln^{-(1-p)/(2-p)}]. \end{cases}$$

This follows from Proposition 3.5.

- If (3.63) holds for some $p \in (1, 2)$, then there exists a positive constant C such that

$$\mathbb{P}\left(|W_1(\mu_n, \mu) - \mathbb{E}[W_1(\mu_n, \mu)]| > x\right) \leq \frac{C}{n^{p-1}x^p}.$$

This follows from Proposition 3.12.

- If (3.63) holds for some $p \geq 2$, then there exists a positive constant C such that

$$\mathbb{P}\left(|W_1(\mu_n, \mu) - \mathbb{E}[W_1(\mu_n, \mu)]| > x\right) \leq \frac{C}{n^{p/2}x^p}.$$

This follows from Proposition 3.15.

And for the moment bounds of S_n :

- If (3.60) for some $p \in [1, 2]$, then

$$\left\|W_1(\mu_n, \mu) - \mathbb{E}[W_1(\mu_n, \mu)]\right\|_p^p \leq \frac{C}{n^{p-1}}. \quad (4.2)$$

This follows from Proposition 3.11.

- If (3.78) holds for some $p \geq 2$, then

$$\left\|W_1(\mu_n, \mu) - \mathbb{E}[W_1(\mu_n, \mu)]\right\|_p^p \leq \frac{C}{n^{p/2}}. \quad (4.3)$$

This follows from Proposition 3.14.

Let us now give some references on the subject.

As already mentioned, the subgaussian bound (4.1) for $p = 2$ is proved in the paper by Djellout *et al.* [11]. Notice that these authors also consider the Wasserstein metrics W_r for $r \geq 1$, with cost function $c(x, y) = (d(x, y))^r$.

In the iid case, when $X_i = \varepsilon_i$, some very precise results are given in the paper by Gozlan and Leonard [17], for a more general class of Wasserstein metrics (meaning that the cost function is not necessary a distance). In the case of W_1 , they have obtained deviation inequalities under some conditions of the Laplace transform of some convex and increasing function of $d(x_0, X_1)$ (see their Theorem 10 combined with their Theorem 7). In particular, *via* their Lemma 1, they have obtained a Cramér-type inequality for W_1 similar to what we get in Proposition 3.2.

In the dependent case, another important reference is the recent paper by Chazottes and Gouëzel [5]. These authors consider separately Lipschitz functionals of iterates of maps that

can be modeled by Young towers. They obtain exponential or polynomial bounds according as the covariances between Lipschitz functions of the iterates decrease with an exponential or polynomial rate. See their Section 7.3 for the applications to the Wasserstein distance W_1 . Note that the Markov chains associated to the maps considered by Chazottes and Gouëzel do not in general satisfy the one step contraction, and are much more difficult to handle than the class of Markov chains of the present paper.

4.2 Discussion

Of course, the next question is that of the behavior of $\mathbb{E}[W_1(\mu_n, \mu)]$, because it can give us information on $W_1(\mu_n, \mu)$ through the preceding inequalities. For instance, from (4.2), we infer that if (3.60) holds for some $p \in [1, 2]$, then

$$\mathbb{E}[W_1(\mu_n, \mu)] \leq \|W_1(\mu_n, \mu)\|_p \leq \mathbb{E}[W_1(\mu_n, \mu)] + \frac{C}{n^{(p-1)/p}}. \quad (4.4)$$

In the same way, from (4.3), we infer that if (3.78) holds for some $p \geq 2$, then

$$\mathbb{E}[W_1(\mu_n, \mu)] \leq \|W_1(\mu_n, \mu)\|_p \leq \mathbb{E}[W_1(\mu_n, \mu)] + \frac{C}{\sqrt{n}}. \quad (4.5)$$

Let us first quote that, if $\mathbb{E}[G_{X_1}(X_1)] < \infty$ and $\mathbb{E}[G_\varepsilon(\varepsilon)] < \infty$, then $\mathbb{E}[W_1(\mu_n, \mu)]$ converges to 0. Indeed, the Markov chain $(X_i)_{i \geq 1}$ satisfies the strong law of large numbers:

$$\lim_{n \rightarrow \infty} \mu_n(f) = \mu(f) \quad \text{almost surely,}$$

for any f such that $f(x) \leq C(1 + d(x_0, x))$. Hence, it follows from Theorem 6.9 in Villani [30] that $W_1(\mu_n, \mu)$ converges to 0 almost surely, and that $\mathbb{E}[W_1(\mu_n, \mu)]$ converges to 0.

The question of the rate of convergence to 0 of $\mathbb{E}[W_1(\mu_n, \mu)]$ is delicate, and has a long history. Let us recall some known results in the iid case, when $X_i = \varepsilon_i$.

- If $\mathcal{X} = \mathbb{R}$ and $d(x, y) = |x - y|$, and if $\int |x| \sqrt{\mathbb{P}(|X_1| > x)} dx < \infty$, then

$$\lim_{n \rightarrow \infty} \sqrt{n} \mathbb{E}[W_1(\mu_n, \mu)] = c$$

with $c \neq 0$ as soon as X_i is not almost surely constant. This follows from del Barrio *et al.* [1] and can be easily extended to our Markov setting.

- If $\mathcal{X} = \mathbb{R}^\ell$ and $d(x, y) = \|x - y\|$ for some norm $\|\cdot\|$, let us recall some recent results by Fournier and Guillin [14] (see also Dereich *et al.* [8]). In Theorem 1 of Fournier and Guillin [14], the following upper bounds are proved: Assume that $p > 1$ and that $\int \|x\|^p \mu(dx) < \infty$, then

$$\mathbb{E}[W_1(\mu_n, \mu)] \leq \begin{cases} C(n^{-1/2} + n^{-(p-1)/p}) & \text{if } \ell = 1 \text{ and } p \neq 2 \\ C(n^{-1/2} \ln(1+n) + n^{-(p-1)/p}) & \text{if } \ell = 2 \text{ and } p \neq 2 \\ C(n^{-1/\ell} + n^{-(p-1)/p}) & \text{if } \ell > 2 \text{ and } p \neq \ell/(\ell-1). \end{cases} \quad (4.6)$$

Combining this upper bound with (4.4) and (4.5), we obtain the following proposition

Proposition 4.1. *Let X_1, \dots, X_n be an iid sequence of \mathbb{R}^ℓ -valued random variables, with common distribution μ . Let $p > 1$ and assume that $\int \|x\|^p \mu(dx) < \infty$. Then the quantity $\|W_1(\mu_n, \mu)\|_p$ satisfies the upper bound (4.6).*

Note that Fournier and Guillin [14] consider the case of W_r metrics, and the upper bound (4.6) is just a particular case of their Theorem 1. Note also that an extension of inequality (4.6) to ρ -mixing Markov chains is given in Theorem 15 of the same paper.

In their Theorem 2, Fournier and Guillin [14] give some deviation inequalities for

$$\mathbb{P}(W_r(\mu_n, \mu) > x).$$

For $r = 1$, these results are different from ours, since they do not deal with concentration around the mean. In particular their upper bounds depend on the dimension ℓ , and for $r = 1$ and $\ell \geq 3$ they are useless for $x = yn^{-\alpha}$ as soon as $\alpha \in (1/\ell, 1/2]$. This is coherent with our upper bounds of Section 4.1 since in that case $\mathbb{E}[W_1(\mu_n, \mu)]$ can be of order $n^{-1/\ell}$. Let us note, however, that the results of Section 4.1 give always an efficient upper bound for the concentration of $W_1(\mu_n, \mu)$ around $\mathbb{E}[W_1(\mu_n, \mu)]$ for any $x = yn^{-\alpha}$ with $\alpha \in [0, 1/2]$, that is in the whole range from small to large deviations, whatever the dimension of \mathcal{X} .

- Concerning the behavior of $\mathbb{E}[W_1(\mu_n, \mu)]$ in the infinite dimensional case, let us mention the upper bound (15) in Boissard [3]. This upper bound involves the covering numbers of an increasing sequence of compact sets K_t for which $\mu(K_t^c)$ tends to zero as t tends to infinity. Some extensions to a class of Markov chains are given in Section 2.4 of the same paper. In particular, his results apply to one step contracting Markov chains satisfying (1.2) (again, this follows from Proposition 3.1 of Djellout *et al.* [11]).

Acknowledgements. Jérôme Dedecker is partially supported by the French ANR project Top-Data.

References

- [1] del Barrio, E., Giné, E. and Matrán, C., 1999. Central limit theorems for the Wasserstein distance between the empirical and the true distributions. *Ann. Probab.* **27**, 1009–1071.
- [2] Bobkov, S. and Götze, F., 1999. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.* **163**, 1–28.
- [3] Boissard, E., 2011. Simple bounds for the convergence of empirical and occupation measures in 1-Wasserstein distance. *Elect. Journal Probab.* **16**, 2296–2333.
- [4] Borovkov, A. A., 2000. Estimates for the distribution of sums and maxima of sums of random variables when the Cramér condition is not satisfied. *Sib. Math. J.* **41**, 811–848.
- [5] Chazottes, J.R. and Gouëzel, S., 2012. Optimal concentration inequalities for dynamical systems. *Commun. Math. Physics* **316**, 843–889.
- [6] Collet, P., Martinez, S., and Schmitt, B., 2002. Exponential inequalities for dynamical measures of expanding maps of the interval. *Probab. Theory Relat. Fields* **123**, 301–322.
- [7] De La Peña, V. H., 1999. A general class of exponential inequalities for martingales and ratios. *Ann. Probab.* **27**, 537–564.
- [8] Dereich, S., Scheutzow, M. and Schottstedt, R., 2013. Constructive quantization: approximation by empirical measures. *Ann. Inst. Henri Poincaré Probab. Stat.* **49**, 1183–1203.
- [9] Delyon, B., Juditsky, A. and Liptser, R., 2006. Moderate deviation principle for ergodic Markov chain. Lipschitz summands. In “From stochastic calculus to mathematical finance” (pp. 189-209). Springer Berlin Heidelberg.
- [10] Diaconis, P., Freedman, D. 1999. Iterated random functions. *SIAM review* **41**, No. 1, 45–76.
- [11] Djellout, H., Guillin, A. and Wu, L., 2004. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Probab.* **32**, 2702–2732.
- [12] Fan, X., Grama, I., Liu, Q., 2012. Hoeffding’s inequality for supermartingales. *Stochastic Process. Appl.* **122**, 3545–3559.
- [13] Fan, X., Grama, I. and Liu, Q., 2012. Large deviation exponential inequalities for supermartingales. *Electron. Commun. Probab.* **17**, no. 59, 1–8.
- [14] Fournier, N. and Guillin, A., 2013. On the rate of convergence in Wasserstein distance of the empirical measure. arxiv:1312.2128v1.
- [15] Fuk, D.Kh., 1973. Some probabilistic inequalities for martingales. *Sib. Math. J.* **14**, 185–193.
- [16] van de Geer, S., 1995. Exponential inequalities for martingales, with application to maximum likelihood estimation for counting process. *Ann. Statist.* **23**, 1779–1801.
- [17] Gozlan, N. and Léonard, C. 2007. A large deviation approach to some transportation cost inequalities. *Probab. Theory Relat. Fields* **139**, 235–283.

- [18] Kontorovich, L. and Ramanan, K., 2008. Concentration inequalities for dependent random variables via the martingale method. *Ann. Probab.* **36**, No. 6, 2126–2158.
- [19] Lesigne, E. and Volný, D., 2001. Large deviations for martingales. *Stochastic Process. Appl.* **96**, 143–159.
- [20] Liu, Q. and Watbled, F., 2009. Exponential inequalities for martingales and asymptotic properties of the free energy of directed polymers in a random environment. *Stochastic Process. Appl.* **119**, 3101–3132.
- [21] McDiarmid, C., 1989. On the method of bounded differences. *Surveys of combinatorics. London Math. Soc. Lectures Notes Ser.* **141**, 148–188.
- [22] Merlevède, F., Peligrad, M., and Rio, E., 2011. A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probab. Theory Relat. Fields* **151**, 435–474.
- [23] Pinelis, I., 1994. Optimum bounds for the distribution of martingales in Banach spaces. *Ann. Probab.* **7**, 1679–1706.
- [24] Pinelis, I., Best possible bounds of the von Bahr–Esseen type, arXiv:1008.5350v2 [math.PR] 18 Nov 2013
- [25] Pinelis, I., 2013. Optimal re-centering bounds with applications to Rosenthal-type concentration of measure inequalities. *High Dimensional Probab. VI. Progress Probab.* **66**, 81–93.
- [26] Rio, E., 2000. Inégalités de Hoeffding pour les fonctions Lipschitziennes de suites dépendantes. *C. R. Acad. Sci. Paris Sér. I Math.* **330**, 905–908.
- [27] Rio, E., 2009. Moment inequalities for sums of dependent random variables under projective conditions. *J. Theor. Probab.* **22**, 146–163.
- [28] Rio, E., 2013. On McDiarmids concentration inequality. *Electron. Commun. Probab.* **18**, no. 44, 1–11.
- [29] Rosenblatt, M., 1956. A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. USA* **42**, 43–47.
- [30] Villani, C., 2009. Optimal transport: old and new. *Grundlehren der mathematischen wissenschaften* **338**, Springer-Verlag Berlin Heidelberg.
- [31] Yurinskii, V. V., 1974. Exponential bounds for large deviations. *Theory Probab. Appl.* **19**, 154–155.