



General Bernstein-Like Inequality for Additive Functionals of Markov Chains

Michał Lemańczyk¹ 

Received: 3 June 2019 / Revised: 10 March 2020 / Published online: 7 May 2020
© The Author(s) 2020

Abstract

Using the renewal approach, we prove Bernstein-like inequalities for additive functionals of geometrically ergodic Markov chains, thus obtaining counterparts of inequalities for sums of independent random variables. The coefficient in the sub-Gaussian part of our estimate is the asymptotic variance of the additive functional, i.e., the variance of the limiting Gaussian variable in the central limit theorem for Markov chains. This refines earlier results by Adamczak and Bednorz, obtained under the additional assumption of strong aperiodicity of the chain.

Keywords General Markov chain · Concentration inequality · Bernstein inequality

Mathematics Subject Classification (2010) 60E15 · 60J05

1 Introduction

Throughout this paper, we assume that $\Upsilon = (\Upsilon_n)_{n \in \mathbb{N}}$ is a Markov chain defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, taking values in a measurable (countably generated) space $(\mathcal{X}, \mathcal{B})$, with a transition function $P : \mathcal{X} \times \mathcal{B} \rightarrow [0, 1]$. Moreover, we assume that Υ is **ψ -irreducible** and **aperiodic** and **admits a unique invariant probability measure** π . As usual for any initial distribution μ on \mathcal{X} , we will write $\mathbb{P}_\mu(\Upsilon \in \cdot)$ for the distribution of the chain with Υ_0 distributed according to the measure μ . We will denote by δ_x the Dirac's mass at x , and to shorten the notation, we will use \mathbb{P}_x instead of \mathbb{P}_{δ_x} .

Research supported by the National Science Center, Poland, Grant No. 2015/18/E/ST1/00214.

✉ Michał Lemańczyk
miclem1@wp.pl

¹ Institute of Mathematics, University of Warsaw, Banacha 2, 02-097 Warsaw, Poland

We say that Υ is **geometrically ergodic** if there exists a positive number $\rho < 1$ and a real function $G : \mathcal{X} \rightarrow \mathbb{R}$ such that for every starting point $x \in \mathcal{X}$ and $n \in \mathbb{N}$,

$$\|P^n(x, \cdot) - \pi(\cdot)\|_{TV} \leq G(x)\rho^n, \tag{1.1}$$

where $\|\cdot\|_{TV}$ denotes the total variation norm of a measure and $P^n(\cdot, \cdot)$ is the n -step transition function of the chain. For equivalent conditions, we refer to Chapter 15 of [22].

We will be interested in tail inequalities for sums of random variables of the form

$$\mathbb{P}_x \left(\left| \sum_{i=0}^{n-1} f(\Upsilon_i) \right| > t \right),$$

where $f : \mathcal{X} \rightarrow \mathbb{R}$ is a measurable real function and $x \in \mathcal{X}$ is a starting point. Although our main results, stated in Sect. 4, do not require f to be bounded, we give here a version in the bounded case for the sake of simplicity. This version will be easier to compare to the Bernstein inequality for bounded random variables stated in Sect. 2 (cf. Theorem 2.1). Below for convenience, we set $\log(\cdot) = \ln(\cdot \vee e)$, where $\ln(\cdot)$ is the natural logarithm.

Theorem 1.1 (*Bernstein-like inequality for Markov chains*) *Let Υ be a geometrically ergodic Markov chain with state space \mathcal{X} , and let π be its unique stationary probability measure. Moreover, let $f : \mathcal{X} \rightarrow \mathbb{R}$ be a bounded measurable function such that $\mathbb{E}_\pi f = 0$. Furthermore, let $x \in \mathcal{X}$. Then, we can find constants $K, \tau > 0$ depending only on x and the transition probability $P(\cdot, \cdot)$ such that for all $t > 0$,*

$$\mathbb{P}_x \left(\left| \sum_{i=0}^{n-1} f(\Upsilon_i) \right| > t \right) \leq K \exp \left(-\frac{t^2}{32n\sigma_{Mrv}^2 + \tau t \|f\|_\infty \log n} \right),$$

where

$$\sigma_{Mrv}^2 = \text{Var}_\pi(f(\Upsilon_0)) + 2 \sum_{i=1}^\infty \text{Cov}_\pi(f(\Upsilon_0), f(\Upsilon_i)) \tag{1.2}$$

denotes the asymptotic variance of the process $(f(\Upsilon_i))_i$.

Remark 1.2 We refer to Theorem 4.3 for a more general counterpart of Theorem 1.1 and to Theorem 4.4 for explicit formulas for K and τ .

Let us comment briefly on the method of proof. We rely on the by now classical regeneration technique of Athreya–Ney and Nummelin (see [3,22,23]), which allows to split the sum in question into a random number of 1-dependent blocks of random lengths. In the context of tail inequalities, this approach has been successfully used, e.g., in [1,2,6,7,10,12] and provides Bernstein inequalities of optimal type under an additional assumption of strong aperiodicity of the chain (corresponding to $m = 1$

in (3.1)), which ensures that the blocks are independent and allow for a reduction to inequalities for sums of i.i.d. random variables. However, in the general case the implementation of this method available in the literature leads to loss of correlation structure and as a consequence to suboptimal sub-Gaussian coefficient in Bernstein's inequality (in place of σ_{Mrv}^2). Our main technical contribution is to propose a regeneration-based approach which allows to preserve the correlation structure and recover the correct asymptotic behavior, corresponding to the CLT for Markov chains.

The organization of the article is as follows. After a brief discussion of our results (Sect. 2), we introduce the notation and provide a short description of the regeneration method (Sect. 3). Next, we state our main theorems at their full strength (Sect. 4). At the end, we present their proofs (Sect. 7). Along the way, we develop auxiliary theorems for 1-dependent random variables (Sect. 5) and bounds on number of regenerations (Sect. 6). Some technical lemmas concerning exponential Orlicz norms are deferred to Appendix.

2 Discussion of the Main Result

Let us start by recalling the Bernstein inequality in the i.i.d. bounded case.

Theorem 2.1 (*Classical Bernstein inequality*) *If $(\xi_i)_i$ is a sequence of i.i.d. centered random variables such that $\sup_i \|\xi_i\|_\infty \leq M$, then for $\sigma^2 = \mathbb{E}\xi_i^2$ and any $t > 0$,*

$$\mathbb{P}\left(\sup_{1 \leq k \leq n} \left| \sum_{i=1}^k \xi_i \right| \geq t\right) \leq 2 \exp\left(-\frac{t^2}{2n\sigma^2 + \frac{2}{3}Mt}\right).$$

Let us recall that the CLT for Markov chains (see, e.g., [9,22,23]) guarantees that under assumptions and notation of Theorem 1.1, the sums $\frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} f(\Upsilon_i)$ converge in distribution to the normal distribution $\mathcal{N}(0, \sigma_{Mrv}^2)$. Thus, the inequality obtained in Theorem 1.1 reflects (up to constants) the asymptotic normal behavior of the sums $\frac{1}{\sqrt{n}} \sum f(\Upsilon_i)$ similarly as the classical Bernstein inequality in the i.i.d. context. Furthermore, the term $\log n$ which appears in our inequality is necessary. More precisely, one can show that if the following inequality holds for all $t > 0$:

$$\mathbb{P}_x\left(\left|\sum_{i=0}^{n-1} f(\Upsilon_i)\right| > t\right) \leq \text{const} \cdot \exp\left(-\frac{t^2}{\text{const} \cdot n\sigma^2 + \text{const}(x) \cdot a_n t \|f\|_\infty}\right) \quad (2.1)$$

for some $a_n = o(n)$ and $\sigma \in \mathbb{R}$ (const 's stand for some absolute constants, whereas $\text{const}(x)$ depends only on x and the Markov chain), then one must have $\sigma^2 \geq \text{const} \cdot \sigma_{Mrv}^2$. Moreover, it is known that for some geometrically ergodic chains a_n must grow at least logarithmically with n (see [1], Section 3.3).

Concentration inequalities for Markov chains and processes have been thoroughly studied in the literature, the (non-comprehensive) list of works concerning this topic

includes [1,2,6,7,10–13,15–17,19,20,24,25,27]. Some results are devoted to concentration for general functions of the chain (they are usually obtained under various Lipschitz or bounded difference type conditions); others specialize to additive functionals, which are the object of study in our case. Tail inequalities for additive functionals are usually counterparts of Hoeffding or Bernstein inequalities. The former ones do not take into account the variance of the additive functional and are expressed in terms of $\|f\|_\infty$ only. They can be often obtained as special cases of concentration inequalities for general function (see, e.g., [11,24,25]). Bernstein-type estimates of the form (2.1) are considered, e.g., in [1,2,6,7,10,12,13,16,17,19,20,24,27] and use various variance proxies σ^2 , which do not necessarily coincide with the limiting variance σ_{Mrv}^2 . In the continuous time case, inequalities of Bernstein type for the natural counterpart of the additive functional, involving asymptotic variance, have been obtained under certain spectral gap or Lyapunov-type conditions in [13,16]. For discrete time Markov chains, inequalities obtained in [1,2,7,10,12] by the regeneration method give (2.1) (under various types of ergodicity assumptions and with various parameters a_n) with σ^2 , which coincides with σ_{Mrv}^2 only under additional assumption of strong aperiodicity of the chain. On the other hand, the articles [19,20,25,27] provide more general results, available for non-necessarily Markovian sequences of random variables, satisfying various types of mixing conditions. The variance proxies σ^2 that are used in these references are close to the asymptotic variance and however in general do not coincide with it. For instance, the inequality obtained in [19], which is valid in particular for geometrically ergodic chains, uses (in our notation) $\sigma^2 = \text{Var}_\pi(f(\Upsilon_0)) + 2 \sum_{i=1}^{\infty} |\text{Cov}_\pi(f(\Upsilon_0), f(\Upsilon_i))|$. Comparing with (1.2), one can see that $\sigma_{Mrv}^2 \leq \sigma^2$. In fact, one can construct examples when the ratio between the two quantities is arbitrarily large or even $\sigma_{Mrv}^2 = 0$ and $\sigma^2 > 0$. Reference [27] provides an inequality for *uniformly* geometrically ergodic processes, involving a certain implicitly defined variance proxy σ_n^2 , which may be bounded from above by σ^2 from [19] or by $\text{Var}_\pi(f(\Upsilon_0)) + C\|f\|_\infty \mathbb{E}_\pi |f(\Upsilon_0)|$, where C is a constant depending on the mixing properties of the process. For a fixed process, in the non-degenerate situation, when the asymptotic variance is nonzero, it can be substituted for σ_n^2 at the cost of introducing additional multiplicative constants, depending on the chain and the function f .

To the best of our knowledge, Theorem 1.1 is therefore the first tail inequality available for general geometrically ergodic Markov chains (not necessarily strongly aperiodic), which (up to universal constants) reflects the correct limiting Gaussian behavior of additive functionals. The problem of obtaining an inequality of this type was posed in [2]. Let us remark that quantitative investigation of problems related to the central limit theorems for general aperiodic Markov chains seems to be substantially more difficult than for chains which are strongly aperiodic. For instance, optimal strong approximation results are still known only in the latter case [21].

3 Notation and Basic Properties

For any $k, l \in \mathbb{Z}$, $k \leq l$, we define **integer intervals** of consecutive integers

$$[k, l] = \{k, k+1, \dots, l\}, \quad [k, l) = \{k, k+1, \dots, l-1\}, \quad [k, \infty) = \{k, k+1, \dots\}.$$

For any process $\mathbf{X} = (X_i)_{i \in \mathbb{N}}$ and $S \subset \mathbb{N}$, we put

$$X_S = (X_i)_{i \in S}, \quad \mathcal{F}^{\mathbf{X}} = \left(\mathcal{F}_i^{\mathbf{X}} \right)_{i \in \mathbb{N}}, \quad \mathcal{F}_i^{\mathbf{X}} = \sigma(X_{[0,i]}).$$

Moreover, for $k \in \mathbb{N}$ we define the corresponding **vectorized process**

$$\mathbf{X}^{(k)} = \left(X_i^{(k)} \right)_{i \in \mathbb{N}}, \quad X_i^{(k)} = X_{[ik, (i+1)k]}.$$

Definition 3.1 (Stationarity) We say that a process $(X_n)_{n \in \mathbb{N}}$ is **stationary** if for any $k \in \mathbb{N}$ the shifted process $(X_{n+k})_{n \in \mathbb{N}}$ has the same distribution as $(X_n)_{n \in \mathbb{N}}$.

Definition 3.2 (m -dependence) Fix $m \in \mathbb{N}$. We say that $(X_n)_{n \in \mathbb{N}}$ is **m -dependent** if for any $k \in \mathbb{N}$ the process $(X_n)_{n \leq k}$ is independent of the process $(X_n)_{n \geq m+1+k}$.

Remark 3.3 Let us note that a process $(X_n)_{n \in \mathbb{N}}$ is 0-dependent iff the variables $(X_n)_{n \in \mathbb{N}}$ are independent. Finally, let us give a natural example of a 1-dependent process $(X_n)_{n \in \mathbb{N}}$. Fix an independent process $(\xi_n)_{n \in \mathbb{N}}$ and a Borel, real function $h : \mathbb{R}^2 \rightarrow \mathbb{R}$. Then, $(h(\xi_n, \xi_{n+1}))_{n \in \mathbb{N}}$ is 1-dependent. Such processes are called two-block factors. It is worth noting that there are 1-dependent processes which are not two-block factors (see [8]).

Remark 3.4 Assume that a process $(X_n)_{n \in \mathbb{N}}$ is m -dependent. Then for any $n_0 \in \mathbb{N}$, the process $(X_{n_0+k(m+1)})_{k \in \mathbb{N}}$ is independent. Moreover, if the process $(X_n)_{n \in \mathbb{N}}$ is stationary, then for any $n_0 \in \mathbb{N}$, $(X_{n_0+k(m+1)})_{k \in \mathbb{N}}$ is a collection of i.i.d. random variables.

3.1 Split Chain

As already mentioned in the Introduction, our proofs will be based on the regeneration technique which was invented independently by Nummelin and Athreya–Ney (see [3] and [23]) and was popularized by Meyn and Tweedie [22]. We will introduce the split chain and then regeneration times of the split chain. The construction of the split chain is well known, and as references, we recommend [22] (Chaps. 5,17) and [23]. We briefly recall this technique below. Let us stress that although this construction is based on the one presented in [22], our notation is slightly different. Firstly, let us recall the minorization condition for Markov chains which plays a main role in the splitting technique.

Definition 3.5 We say that a Markov chain Υ satisfies **the minorization condition** if there exists a set $C \in \mathcal{B}(\mathcal{X})$ (called a small set), a probability measure ν on \mathcal{X} (a small measure), a constant $\delta > 0$ and a positive integer $m \in \mathbb{N}$ such that $\pi(C) > 0$ and

$$P^m(x, B) \geq \delta \nu(B) \tag{3.1}$$

holds for all $x \in C$ and $B \in \mathcal{B}(\mathcal{X})$.

Remark 3.6 One can assume that $\nu(C) = 1$ (possibly at the cost of increasing m).

Remark 3.7 One can check that under assumptions of our theorem, the minorization condition (3.1) holds for some C, ν, δ and m . We refer to [22], Section 5.2 for the proof of this fact.

Fix C, m, ν and $\delta > 0$ as in (3.1). The minorization condition allows us to redefine the chain Υ together with an auxiliary regeneration structure. More precisely, we start with a splitting of the space \mathcal{X} into two identical copies on level 0 and 1, namely we consider $\overline{\mathcal{X}} = \mathcal{X} \times \{0, 1\}$. Now, we split Υ in the following way. We consider a process $\Phi = (\Upsilon, \Lambda) = (\Upsilon_i, \Lambda_i)_{i \geq 0}$ (usually called the split chain) defined on $\overline{\mathcal{X}}$. (We slightly abuse the notation by denoting the first coordinate of the split chain with the same letter as for the initial Markov chain, but it will turn out that the first coordinate of the split chain has the same distribution as the starting Markov chain, so this notation is justified.) The random variables Λ_k take values in $\{0, 1\}$. (They indicate the level on which Φ_k is.) For a fixed $x \in C$, let

$$r(x, y) = \frac{\delta \nu(dy)}{P^m(x, dy)} \tag{3.2}$$

and note that the above Radon–Nikodym derivative is well defined thanks to (3.1). Moreover, $r(x, y) \leq 1$. Now, for any $A_1, \dots, A_m \in \mathcal{B}(\mathcal{X}), k \in \mathbb{N}$ and $i \in \{0, 1\}$ set

$$\begin{aligned} &\mathbb{P}\left(\Lambda_{km} = i, \Upsilon_{[km+1, (k+1)m]} \in A_1 \times \dots \times A_m \mid \mathcal{F}_{km}^\Upsilon, \mathcal{F}_{km-m}^\Lambda, \Upsilon_{km} = x\right) \\ &= \mathbb{P}\left(\Lambda_0 = i, \Upsilon_{[1, m]} \in A_1 \times \dots \times A_m \mid \Upsilon_0 = x\right) \\ &= \int_{A_1} \dots \int_{A_m} r(x, x_m, i) P(x_{m-1}, dx_m) P(x_{m-2}, dx_{m-1}) \dots P(x, dx_1), \end{aligned} \tag{3.3}$$

where

$$r(x, y, i) = \begin{cases} \mathbb{1}_{x \in C} r(x, x_m), & \text{if } i = 1, \\ 1 - \mathbb{1}_{x \in C} r(x, x_m), & \text{if } i = 0. \end{cases} \tag{3.4}$$

Moreover, for any $k, i \in \mathbb{N}$ such that $km < i < (k + 1)m$ we set

$$\Lambda_i = \Lambda_{km}. \tag{3.5}$$

Remark 3.8 (Initial distribution for the split chain) In order to be able to set initial distribution for the split chain for arbitrary probability measure μ on \mathcal{X} , we define the **split measure** μ^* on $\overline{\mathcal{X}}$ by:

$$\mu^*(A \times \{i\}) = \begin{cases} (1 - \delta)\mu(C \cap A) + \mu(A \cap C^c), & \text{if } i = 0, \\ \delta\mu(C \cap A), & \text{if } i = 1. \end{cases} \tag{3.6}$$

Such definition ensures that $(\Upsilon_0, \Lambda_0) \sim \mu^*$ as soon as $\Upsilon_0 \sim \mu$. For convenience sake, for any $x \in \mathcal{X}$, we will write

$$\mathbb{P}_{x^*}(\cdot) = \mathbb{P}_{\delta_x^*}(\cdot). \tag{3.7}$$

Remark 3.9 (Markov-like properties of the split chain) In order to give some intuition behind the definition of the split chain, note that the distribution of the first coordinate of the split chain Φ with initial distribution μ^* coincides with that of the original Markov chain Υ which starts from μ . From now on, Υ always corresponds to this first coordinate of the split chain. One can easily generalize (3.3) to show the following **Markov-like property of the split chain**: For any $k \in \mathbb{N}$ and product measurable bounded function F , we have

$$\mathbb{E}\left(F\left(\Upsilon_{[km+1,\infty)}, \Lambda_{[km,\infty)}\right) \mid \mathcal{F}_{km}^\Upsilon, \mathcal{F}_{km-m}^\Lambda\right) = \mathbb{E}\left(F\left(\Upsilon_{[km+1,\infty)}, \Lambda_{[km,\infty)}\right) \mid \Upsilon_{km}\right). \tag{3.8}$$

This, in turn, leads to the fact that the vectorized split chain $\Phi^{(m)}$ is a Markov chain. Even more, for any product measurable bounded function F and $k \in \mathbb{N}$ we have

$$\begin{aligned} \mathbb{E}\left(F\left(\Phi_{[k,\infty)}^{(m)}\right) \mid \Phi_{[0,k)}^{(m)}\right) &= \mathbb{E}\left(F\left(\Phi_{[k,\infty)}^{(m)}\right) \mid \Phi_{k-1}^{(m)}\right) \\ &= \mathbb{E}\left(F\left(\Phi_{[k,\infty)}^{(m)}\right) \mid \Upsilon_{mk-m}, \Upsilon_{mk-1}, \Lambda_{mk-m}\right). \end{aligned}$$

Now, we can introduce the aforementioned **regeneration structure for Φ** . Firstly, we define certain stopping times. For convenience, we put $\tau_{-1} = -m$, and then, for $i \geq 0$ we define τ_i to be the i th time when the second coordinate (level coordinate) hits 1, namely

$$\tau_i = \min\{k > \tau_{i-1} \mid \Lambda_k = 1, m|k\}. \tag{3.9}$$

Now, we are ready to introduce **the random blocks and the random block process**

$$\Xi_i = \Upsilon_{[\tau_{i-1}+m, \tau_i+m)}, \quad \Xi = (\Xi_i)_{i \geq 0}, \tag{3.10}$$

where we consider Ξ_i as a random variable with values in the disjoint union $\bigsqcup_{j \geq 0} \mathcal{X}^j$. For clarity of this presentation, here and later on, we omit the measurability details.

Remark 3.10 Let us now briefly discuss the behavior of these random blocks. Firstly, by the strong Markov property of the vectorized split chain it is not hard to see that Ξ is a Markov chain. On a closer look, one can see that for any product measurable function F

$$\mathbb{E}\left(F\left(\Xi_{[i,\infty)}\right) \mid \Xi_{[0,i)}\right) = \mathbb{E}\left(F\left(\Xi_{[i,\infty)}\right) \mid \Xi_{i-1}\right) = \mathbb{E}\left(F\left(\Xi_{[i,\infty)}\right) \mid \text{pr}_m\left(\Xi_{i-1}\right)\right), \tag{3.11}$$

where $\text{pr}_m : \bigsqcup_{j \geq m} \mathcal{X}^j \rightarrow \mathcal{X}^m$ is a projection on m -last coordinates,

$$\text{pr}_m(x_0, \dots, x_j) = (x_{j-m+1}, \dots, x_j). \tag{3.12}$$

Apart from being Markovian, the sequence $(\Xi_i)_{i \geq 0}$ is 1-dependent, whereas $(\Xi_i)_{i \geq 1}$ is stationary (see [9], Corollary 2.4). The stationarity follows from the fact that for $m|k$, we have

$$\mathcal{L}(\Upsilon_{k+m} \mid \Lambda_k = 1) = \nu, \tag{3.13}$$

that is, every time k (which is a multiple of m) the split chain is on level 1 (note that this implies $\Upsilon_k \in C$) and the split chain **regenerates** and starts anew from ν . Furthermore, the lengths of Ξ_i :

$$|\Xi_i| = \tau_i - \tau_{i-1}, \tag{3.14}$$

are **independent** random variables for $i \geq 0$ and form a stationary process for $i \geq 1$. Let us add that if $m = 1$, one can show that Ξ_i 's are independent. This fact makes a crucial difference between strongly aperiodic and not strongly aperiodic Markov chains (see [5, Section 6]).

At last, let us introduce **the excursions and the excursion process**

$$\chi_i = \chi_i(f) = \sum_{j=\tau_i+m}^{\tau_{i+1}+m-1} f(\Upsilon_j), \quad \boldsymbol{\chi} = (\chi_i)_{i \geq 0}, \tag{3.15}$$

which will play a crucial role in our future considerations. By properties of the random blocks, one concludes that $\boldsymbol{\chi}$ is 1-dependent and satisfies

$$\mathbb{E}(\chi_i \mid \Xi_{[0,i]}) = \mathbb{E}(\chi_i \mid \Xi_i). \tag{3.16}$$

Moreover, $(\chi_i)_{i \geq 1}$ is stationary. Due to the Pitman occupation measure formula (see, [22], Theorem 17.3.1, page 428) which says that for any measurable real function G ,

$$\mathbb{E}_\nu \sum_{i=0}^{\tau_0/m} G(\Upsilon_{mi}, \Lambda_{mi}) = \delta^{-1} \pi(C)^{-1} \mathbb{E}_\pi G(\Upsilon_0, \Lambda_0), \tag{3.17}$$

and observation that \mathbb{P}_μ -distribution of excursion $\chi_i(f)$ ($i \geq 1$) is equal to the \mathbb{P}_ν -distribution of χ_0 , we get that for any initial distribution μ and any $i \geq 1$,

$$\mathbb{E}_\mu \chi_i = \mathbb{E}_\nu \chi_0 = \delta^{-1} \pi(C)^{-1} m \int f d\pi. \tag{3.18}$$

As a consequence, $\mathbb{E}_\pi f(\Upsilon_i) = 0$ implies that for every $i \geq 1$, $\mathbb{E}_\mu \chi_i(f) = 0$. Now, we are ready to decompose our sums into random blocks. If $m|n$, then

$$\begin{aligned} \sum_{i=0}^{n-1} f(\Upsilon_i) &= \left(\sum_{i=0}^{\tau_0/m} \Theta_i \mathbb{1}_{N>0} + \mathbb{1}_{N=0} \sum_{i=0}^{n/m-1} \Theta_i \right) + \left(\sum_{i=1}^N \chi_{i-1}(f) \right) \\ &\quad - \left(\mathbb{1}_{N>0} \sum_{k=n}^{\tau_N+m-1} f(\Upsilon_k) \right), \end{aligned} \tag{3.19}$$

where

$$\Theta_k = \Theta_k(f) = \sum_{i=0}^{m-1} f(\Upsilon_{km+i}), \quad N = \inf\{i \geq 0 \mid \tau_i + m - 1 \geq n - 1\}. \tag{3.20}$$

This decomposition will be of utmost importance in our proof.

3.2 Asymptotic Variances

During the upcoming proofs, we will meet two types of asymptotic variances: σ_{Mrv}^2 associated with the process $(f(\Upsilon_i))_{i \geq 0}$ and σ_∞^2 associated with χ . The first one defined as

$$\begin{aligned} \sigma_{Mrv}^2 &= \lim_{n \rightarrow \infty} \frac{1}{n} \text{Var}(f(\Upsilon_0) + \dots + f(\Upsilon_{n-1})) \\ &= \text{Var}_\pi(f(\Upsilon_0)) + 2 \sum_{i \geq 1} \text{Cov}_\pi(f(\Upsilon_i), f(\Upsilon_0)) \end{aligned} \tag{3.21}$$

is exactly the variance of the limiting normal distribution of the sequence $\frac{1}{\sqrt{n}} \sum_{i=1}^n f(\Upsilon_i)$. The second one:

$$\sigma_\infty^2 = \lim_{n \rightarrow \infty} \frac{1}{n} \text{Var}(\chi_1 + \dots + \chi_n) = \mathbb{E}\chi_1^2 + 2\mathbb{E}\chi_1\chi_2,$$

is the variance of the limiting normal distribution of the sequence $\frac{1}{\sqrt{n}} \sum_{i=1}^n \chi_i$. Both asymptotic variances are very closely linked via the formula

$$\sigma_\infty^2 = \sigma_{Mrv}^2 \mathbb{E}(\tau_1 - \tau_0) = \sigma_{Mrv}^2 m \delta^{-1} \pi(C)^{-1}. \tag{3.22}$$

For the proof of this formula, we refer to [22] (see (17.32), page 434).

4 Main Results

In order to state our results in the general form, we need to recall the definition of the exponential Orlicz norm. For any random variable X and $\alpha > 0$, we define

$$\|X\|_{\psi_\alpha} = \inf \left\{ c > 0 \mid \mathbb{E} \exp \left(\frac{|X|^\alpha}{c^\alpha} \right) \leq 2 \right\}. \tag{4.1}$$

If $\alpha < 1$, then $\|\cdot\|_{\psi_\alpha}$ is just a quasi-norm. (For basic properties of these quasi-norms, we refer to Appendix A.) In what follows, we will deal with various underlying measures on the state space $\overline{\mathcal{X}}$. In order to stress the dependence of the Orlicz norm on the initial distribution μ of the chain Φ , we will sometimes write $\|\cdot\|_{\psi_\alpha, \mu}$ instead of $\|\cdot\|_{\psi_\alpha}$.

Before we formulate our main result, let us introduce and explain the role of the following parameters:

$$\begin{aligned} \mathbf{a} &= \left\| \sum_{k=0}^{\tau_0/m} |\Theta_k| \right\|_{\psi_\alpha, \mathbb{P}_{x^*}}, & \mathbf{b} &= \left\| \sum_{k=0}^{\tau_0/m} |\Theta_k| \right\|_{\psi_\alpha, \mathbb{P}_{\pi^*}}, \\ \mathbf{c} &= \|\chi_i(f)\|_{\psi_\alpha}, & \mathbf{d} &= \|\tau_1 - \tau_0\|_{\psi_1}, \end{aligned} \tag{4.2}$$

where $\Theta_k = \sum_{i=0}^{m-1} f(\Upsilon_{km+i})$ (cf. (3.19)). The parameter \mathbf{a} (resp. \mathbf{b}) will allow us to estimate the first (third) term on the right-hand side of (3.19), whereas the parameters \mathbf{c} and \mathbf{d} will be used to control the middle term. We note that \mathbf{d} quantifies geometric ergodicity of Υ and is finite as soon as Υ is geometrically ergodic. Let us mention that all these parameters can be bounded, for example, by means of drift conditions widely used in the theory of Markov chains (see Remark 4.2). Finally, let us remind that $\sigma_{Mrv}^2 = \text{Var}_\pi(f(\Upsilon_0)) + 2 \sum_{i=1}^\infty \text{Cov}_\pi(f(\Upsilon_0), f(\Upsilon_i))$ denotes the asymptotic variance of normalized partial sums of the process $(f(\Upsilon_i))_i$.

We are now ready to formulate the first of our main results. (Recall the definitions of the small set C and the minorization condition (3.1).)

Theorem 4.1 *Let Υ be a geometrically ergodic Markov chain and π be its unique stationary probability measure. Let $f : \mathcal{X} \rightarrow \mathbb{R}$ be a measurable function such that $\mathbb{E}_\pi f = 0$ and let $\alpha \in (0, 1]$. Moreover, assume for simplicity that $m|n$. Then for all $x \in \mathcal{X}$ and $t > 0$,*

$$\begin{aligned} \mathbb{P}_x \left(\left| \sum_{i=0}^{n-1} f(\Upsilon_i) \right| > t \right) &\leq 2 \exp \left(-\frac{t^\alpha}{(23\mathbf{a})^\alpha} \right) + 2 [\delta\pi(C)]^{-1} \exp \left(-\frac{t^\alpha}{(23\mathbf{b})^\alpha} \right) \\ &+ 6 \exp(8) \exp \left(-\frac{t^\alpha}{\frac{16}{\alpha} (27\mathbf{c})^\alpha} \right) + 6 \exp \left(-\frac{t^2}{30n\sigma_{Mrv}^2 + 8tM} \right) \\ &+ \exp(1) \exp \left(-\frac{nm}{67\delta\pi(C)\mathbf{d}^2} \right), \end{aligned} \tag{4.3}$$

where σ_{Mrv}^2 denotes the asymptotic variance for the process $(f(\Upsilon_i))_i$ given by (3.21), the parameters $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ are defined by (4.2) and $M = \mathbf{c}(24\alpha^{-3} \log n)^{\frac{1}{\alpha}}$.

Remark 4.2 For the conditions under which $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are finite, we refer to [2], where the authors give bounds on $\mathbf{a}, \mathbf{b}, \mathbf{c}$ under classical drift conditions. If f is bounded, then one easily shows that

$$\max(\mathbf{a}, \mathbf{b}) \leq 2D\|f\|_\infty, \quad \mathbf{c} \leq D\|f\|_\infty, \tag{4.4}$$

where $D = \max(\mathbf{d}, \|\tau_0\|_{\psi_1, \mathbb{P}_{x^*}}, \|\tau_0\|_{\psi_1, \mathbb{P}_{\pi^*}})$. For computable bounds on D , we refer to [4].

Let us note that in Theorem 4.1, the right-hand side of the inequality does not converge to 0 when t tends to infinity. (One of the terms depends on n but not on t .) Usually, in applications t is of order at most n and the other terms dominate on the right-hand side of the inequality, so this does not pose a problem. Nevertheless, one can obtain another version of Theorem 4.1, namely

Theorem 4.3 Under the assumptions and notation of Theorem 4.1, we have

$$\begin{aligned} \mathbb{P}_x \left(\left| \sum_{i=0}^{n-1} f(\Upsilon_i) \right| > t \right) &\leq 2 \exp\left(-\frac{t^\alpha}{(54\mathbf{a})^\alpha}\right) + 2[\delta\pi(C)]^{-1} \exp\left(-\frac{t^\alpha}{(54\mathbf{b})^\alpha}\right) \\ &+ 4 \exp(8) \exp\left(-\frac{t^\alpha}{\frac{16}{\alpha}(27\mathbf{c})^\alpha}\right) + 6 \exp\left(-\frac{t^2}{37(1+p)n\sigma_{Mrv}^2 + 18M\mathbf{d}\sqrt{K_p}t}\right), \end{aligned} \tag{4.5}$$

where $K_p = L_p + 16/L_p$ and $L_p = \frac{16}{p} + 20$.

It is well known that for geometrically ergodic chains $\|\tau_0\|_{\psi_1, \mathbb{P}_{x^*}}, \|\tau_0\|_{\psi_1, \mathbb{P}_{\pi^*}}, \|\tau_1 - \tau_0\|_{\psi_1} < \infty$ (see [4] for constructive estimates). Therefore, (4.4) and Theorem 4.1 lead to

Theorem 4.4 Let Υ be a geometrically ergodic Markov chain and π be its unique stationary, probability measure. Let $f: \mathcal{X} \rightarrow \mathbb{R}$ be a bounded, measurable function such that $\mathbb{E}_\pi f = 0$. Fix $x \in \mathcal{X}$. Moreover, assume that $\|\tau_0\|_{\psi_1, \delta_x^*}, \|\tau_0\|_{\psi_1, \pi^*}, \|\tau_1 - \tau_0\|_{\psi_1} \leq D$. Then for all $t > 0$,

$$\mathbb{P}_x \left(\left| \sum_{i=0}^{n-1} f(\Upsilon_i) \right| > t \right) \leq K \exp\left(-\frac{t^2}{32n\sigma_{Mrv}^2 + 433t\delta\pi(C)\|f\|_\infty D^2 \log n}\right), \tag{4.6}$$

where σ_{Mrv}^2 is the asymptotic variance of $(f(\Upsilon_i))_i$ and $K = \exp(10) + 2\delta^{-1}\pi(C)^{-1}$.

Remark 4.5 Theorem 4.4 implies our main Theorem 1.1 from Introduction with constants $K = (\exp(10) + 2\delta^{-1}\pi(C)^{-1})$ and $\tau = 433\delta\pi(C)D^2$.

5 Bernstein Inequalities for 1-Dependent Sequences

In this section, we will show two versions (for suprema and randomly stopped sums) of Bernstein inequality for 1-dependent random variables. They will be later used in the proofs of our main theorems. In what follows for a 1-dependent sequence of random variables $(X_i)_{i \geq 0}$, σ_∞^2 denotes the asymptotic variance of normalized partial sums, i.e.,

$$\sigma_\infty^2 = \mathbb{E}X_1^2 + 2\mathbb{E}X_1X_2.$$

Lemma 5.1 (*Bernstein inequality for suprema of partial sums*) *Let $(X_i)_{i \geq 0}$ be a 1-dependent sequence of centered random variables such that $\mathbb{E} \exp(c^{-\alpha}|X_i|^\alpha) \leq 2$ for some $\alpha \in (0, 1]$ and $c > 0$. Assume that there exists a filtration $(\mathcal{F}_i)_{i \geq 0}$ such that for $Z_i = X_i + \mathbb{E}(X_{i+1}|\mathcal{F}_i) - \mathbb{E}(X_i|\mathcal{F}_{i-1})$ we have the following:*

- (0) X_i is \mathcal{F}_i measurable,
- (1) $(Z_i)_{i \geq 1}$ is stationary,
- (2) $(Z_i)_{i \geq 1}$ is m -dependent with $m = 1$ or $m = 2$,
- (3) $(\mathbb{E}(X_i|\mathcal{F}_{i-1}))_{i \geq 1}$ is stationary,
- (4) $\mathbb{E}(X_i|\mathcal{F}_{i-1})$ is independent of X_{i+1} for any $i \geq 1$.

Then,

$$\mathbb{E}Z_i^2 = \sigma_\infty^2, \quad \|Z_i\|_{\psi_\alpha} \leq c(8/\alpha)^{\frac{1}{\alpha}}. \tag{5.1}$$

Moreover, for any $t > 0$ and $n \in \mathbb{N}$,

$$\mathbb{P}\left(\sup_{1 \leq k \leq n} \left| \sum_{i=1}^k X_i \right| > t\right) \leq K_m \exp\left(-\frac{t^\alpha}{u_m c^\alpha}\right) + L_m \exp\left(-\frac{t^2}{v_{n,m} \sigma_\infty^2 + w_{n,m} t}\right) \tag{5.2}$$

where $u_m = \frac{16 \cdot 8^\alpha (m+1)^\alpha}{\alpha}$, $v_{n,m} = 5(m+1)(n+m+1)$, $w_{n,m} = 2(m+1)(24\alpha^{-3} \log n)^{\frac{1}{\alpha}} c$, $K_m = 2(m+1) \exp(8)$ and $L_m = 2(m+1)$.

Proof Firstly, we will show that if X_i 's are centered, independent random variables with common variance σ_∞^2 and $\mathbb{E} \exp(c^{-\alpha}|X_i|^\alpha) \leq 2$, then (5.2) holds with $u_0 = 2 \cdot 6^\alpha$, $v_{n,0} = \frac{72}{25} n$, $w_{n,0} = \frac{8}{5} c (3\alpha^{-2} \log n)^{\frac{1}{\alpha}}$, $K_0 = \exp(8)$ and $L_0 = 2$ (allowing for a slight abuse of precision we consider this the $m = 0$ case of the lemma). Indeed, by Lemma 4.1 in [2] for $\lambda = (2^{1/\alpha} c)^{-1}$,

$$\mathbb{E} \exp\left(\lambda^\alpha \sum_{i=0}^{n-1} (|U_i|^\alpha + (\mathbb{E}|U_i|)^\alpha)\right) \leq \exp(8), \tag{5.3}$$

where $U_i = X_i \mathbb{1}_{|X_i| > M_0}$ stands for the ‘‘unbounded’’ part of X_i and $M_0 = c(3\alpha^{-2} \log n)^{\frac{1}{\alpha}}$. Define the ‘‘bounded’’ part of X_i , $B_i = X_i \mathbb{1}_{|X_i| \leq M_0}$ and notice that

$X_i = \overline{B}_i + \overline{U}_i$, where $\overline{B}_i = B_i - \mathbb{E}B_i$ and $\overline{U}_i = U_i - \mathbb{E}U_i$. Using the union bound, we get for $p = 1/6$

$$\begin{aligned} \mathbb{P}\left(\sup_{1 \leq k \leq n} \left| \sum_{i=1}^k X_i \right| > t\right) &\leq \mathbb{P}\left(\sup_{1 \leq k \leq n} \left| \sum_{i=1}^k \overline{U}_i \right| > tp\right) \\ &\quad + \mathbb{P}\left(\sup_{1 \leq k \leq n} \left| \sum_{i=1}^k \overline{B}_i \right| > t(1-p)\right). \end{aligned}$$

Consider first the unbounded part. Using the subadditivity of $x \rightarrow x^\alpha$, Markov’s inequality and then (5.3), we get

$$\begin{aligned} \mathbb{P}\left(\sup_{1 \leq k \leq n} \left| \sum_{i=1}^k \overline{U}_i \right| > tp\right) &\leq \mathbb{P}\left(\exp\left(\lambda^\alpha \sum_{i=1}^n |\overline{U}_i|^\alpha\right) > \exp(\lambda pt)^\alpha\right) \\ &\leq \exp(8) \exp\left(-\frac{t^\alpha p^\alpha}{2c^\alpha}\right) = \exp(8) \exp\left(-\frac{t^\alpha}{2(6c)^\alpha}\right). \end{aligned}$$

As for the “bounded” part, notice that $\mathbb{E}\overline{B}_i^2 \leq \mathbb{E}B_i^2 \leq \mathbb{E}X_i^2 = \sigma_\infty^2$. Therefore, using the classical Bernstein inequality we get

$$\mathbb{P}\left(\sup_{1 \leq k \leq n} \left| \sum_{i=1}^k \overline{B}_i \right| > t(1-p)\right) \leq 2 \exp\left(-\frac{t^2(1-p)^2}{2n\sigma_\infty^2 + \frac{4}{3}t(1-p)M_0}\right).$$

Combining the three last estimates and substituting $p = 1/6$ allow to finish the proof for independent random variables.

We will now use the independent case to prove the tail estimate (5.2), assuming (5.1), the proof of which we postpone. Note that (5.2) is trivial unless $t \geq w_m \log(2(m+1))$ (as the right-hand side exceeds 1). Therefore, from now on we will consider only t satisfying this lower bound. In particular, setting $p = 1/5$, we have $t \geq \frac{2}{p}(2/\alpha)^{\frac{1}{\alpha}}c$ and $t \geq 4^{\frac{1}{\alpha}} \frac{2c}{p} \log(n)^{\frac{1}{\alpha}}$. Using the union bound and the assumption 3), we get (denoting for brevity $\mathbb{E}_i(\cdot) = \mathbb{E}(\cdot | \mathcal{F}_i)$)

$$\begin{aligned} \mathbb{P}\left(\sup_{1 \leq k \leq n} \left| \sum_{i=1}^k X_i \right| > t\right) &\leq \mathbb{P}\left(\sup_{1 \leq k \leq n} \left| \sum_{i=1}^k Z_i \right| > t(1-p)\right) \\ &\quad + \mathbb{P}\left(\sup_{1 \leq i \leq n} |\mathbb{E}_i X_{i+1} - \mathbb{E}_0 X_1| > tp\right) \\ &\leq \mathbb{P}\left(\sup_{1 \leq k \leq n} \left| \sum_{i=1}^k Z_i \right| > t(1-p)\right) \\ &\quad + 2\mathbb{P}\left(\sup_{1 \leq i \leq n} |\mathbb{E}_{i-1} X_i| > \frac{tp}{2}\right). \end{aligned} \tag{5.4}$$

By another application of the union bound together with Lemma A.5 and stationarity of $(\mathbb{E}_{i-1} X_i)_i$, we obtain

$$2\mathbb{P}\left(\sup_{1 \leq i \leq n} |\mathbb{E}_{i-1} X_i| > \frac{tp}{2}\right) \leq 2n\mathbb{P}\left(|\mathbb{E}_0 X_1| > \frac{tp}{2}\right) \leq 12n \exp\left(-\frac{p^\alpha t^\alpha}{2(2c)^\alpha}\right).$$

Notice that

$$\begin{aligned} 12n \exp\left(-\frac{p^\alpha t^\alpha}{2(2c)^\alpha}\right) &= 12\left[n \exp\left(-\frac{p^\alpha t^\alpha}{4(2c)^\alpha}\right)\right] \exp\left(-\frac{p^\alpha t^\alpha}{4(2c)^\alpha}\right) \\ &\leq 12 \exp\left(-\frac{p^\alpha t^\alpha}{4(2c)^\alpha}\right), \end{aligned}$$

where the inequality is a consequence of the estimate $t \geq 4^{\frac{1}{\alpha}} \frac{2c}{p} \log(n)^{\frac{1}{\alpha}}$. It follows that

$$2\mathbb{P}\left(\sup_{1 \leq i \leq n} |\mathbb{E}_{i-1} X_i| > \frac{pt}{2}\right) \leq 12 \exp\left(-\frac{p^\alpha t^\alpha}{4(2c)^\alpha}\right) = 12 \exp\left(-\frac{t^\alpha}{4(10c)^\alpha}\right). \tag{5.5}$$

In order to deal with $\mathbb{P}\left(|\sum_{i=1}^n Z_i| > t(1-p)\right)$, we start with splitting this sum into $m + 1$ parts and using the union bound, namely

$$\mathbb{P}\left(\sup_{1 \leq k \leq n} \left|\sum_{i=1}^k Z_i\right| > t(1-p)\right) \leq \sum_{j=0}^m \mathbb{P}\left(\sup_{1 \leq k \leq n} \left|\sum_{1 \leq i \leq k, m+1|i-j} Z_i\right| > \frac{t(1-p)}{m+1}\right).$$

Now, to each summand on the right-hand side of the above inequality we will apply the estimate for the independent case obtained at the beginning of this proof. Setting $M = (24\alpha^{-3} \log n)^{\frac{1}{\alpha}} c$ and taking into account (5.1), we obtain

$$\begin{aligned} &\frac{1}{m+1} \mathbb{P}\left(\sup_{1 \leq k \leq n} \left|\sum_{i=1}^k Z_i\right| > t(1-p)\right) \\ &\leq \frac{1}{m+1} \sum_{j=0}^m \mathbb{P}\left(\sup_{1 \leq k \leq n} \left|\sum_{1 \leq i \leq k, m+1|i-j} Z_i\right| > \frac{t(1-p)}{m+1}\right) \\ &\leq \exp(8) \exp\left(-\frac{t^\alpha}{\frac{16}{\alpha}(8(m+1)c)^\alpha}\right) \\ &\quad + 2 \exp\left(-\frac{(1-p)^2 t^2}{\frac{72}{25}(m+1)[(n+m+1)\sigma_\infty^2 + \frac{8}{5}(1-p)tM]}\right) \end{aligned}$$

$$\begin{aligned} &\leq \exp(8) \exp\left(-\frac{t^\alpha}{\frac{16}{\alpha}(8(m+1)c)^\alpha}\right) \\ &\quad + 2 \exp\left(-\frac{t^2}{(m+1)[5(n+m+1)\sigma_\infty^2 + 2tM]}\right). \end{aligned} \tag{5.6}$$

Finally, using (5.4), (5.5) and (5.6) we get

$$\begin{aligned} &\mathbb{P}\left(\sup_{1 \leq k \leq n} \left| \sum_{i=1}^k X_i \right| > t\right) \leq 12 \exp\left(-\frac{t^\alpha}{4(10c)^\alpha}\right) \\ &\quad + (m+1) \exp(8) \exp\left(-\frac{t^\alpha}{\frac{16}{\alpha}(8(m+1)c)^\alpha}\right) \\ &\quad + 2(m+1) \exp\left(-\frac{t^2}{5(m+1)(n+m+1)\sigma_\infty^2 + 2(m+1)tM}\right). \end{aligned}$$

To conclude (5.2), it is now enough to note that the second summand on the right-hand side above dominates the first one.

To finish the proof of the lemma, it remains to show (5.1). Firstly, we address the variance of Z_i , which can be easily calculated by using the properties of conditional expectation. We have (recall the notation $\mathbb{E}_i(\cdot) = \mathbb{E}(\cdot | \mathcal{F}_i)$)

$$\mathbb{E}Z_i^2 = \mathbb{E}\left[X_i^2 + \mathbb{E}_i^2 X_{i+1} + \mathbb{E}_{i-1}^2 X_i - 2\mathbb{E}_i X_{i+1} \mathbb{E}_{i-1} X_i - 2X_i \mathbb{E}_{i-1} X_i + 2X_i \mathbb{E}_i X_{i+1}\right].$$

Since $\mathbb{E}X_i \mathbb{E}_{i-1} X_i = \mathbb{E}\mathbb{E}_{i-1}^2 X_i$, $\mathbb{E}\mathbb{E}_i X_{i+1} \mathbb{E}_{i-1} X_i = \mathbb{E}X_{i+1} \mathbb{E}_{i-1} X_i$ and $X_i \mathbb{E}_i X_{i+1} = \mathbb{E}_i(X_i X_{i+1})$, we obtain

$$\begin{aligned} \mathbb{E}Z_i^2 &= \mathbb{E}\left(X_i^2 + \mathbb{E}_i^2 X_{i+1} - \mathbb{E}_{i-1}^2 X_i - 2X_{i+1} \mathbb{E}_{i-1} X_i + 2X_i X_{i+1}\right) \\ &= \mathbb{E}\left(X_i^2 + 2X_i X_{i+1}\right) - 2\mathbb{E}\left(X_{i+1} \mathbb{E}_{i-1} X_i\right) + \mathbb{E}\left(\mathbb{E}_i^2 X_{i+1} - \mathbb{E}_{i-1}^2 X_i\right). \end{aligned}$$

The variance formula in (5.1) follows by observing that due to 3), $\mathbb{E}\left(\mathbb{E}_i^2 X_{i+1} - \mathbb{E}_{i-1}^2 X_i\right) = 0$, whereas by 4), $\mathbb{E}\left(X_{i+1} \mathbb{E}_{i-1} X_i\right) = 0$.

Now, we will demonstrate the upper bound on $\|Z_i\|_{\psi_\alpha}$ in (5.1). Using the triangle inequality (cf. Lemma A.1) twice and then Lemma A.3, we obtain

$$\begin{aligned} \|Z_i\|_{\psi_\alpha} &\leq 2^{\frac{1}{\alpha}-1} \|X_i\|_{\psi_\alpha} + 2^{\frac{1}{\alpha}-1} \|\mathbb{E}_i X_{i+1} - \mathbb{E}_0 X_1\|_{\psi_\alpha} \leq 2^{\frac{1}{\alpha}} \|X_i\|_{\psi_\alpha} + 2^{\frac{2}{\alpha}-1} \|\mathbb{E}_0 X_1\|_{\psi_\alpha} \\ &\leq 2^{\frac{1}{\alpha}} \|X_i\|_{\psi_\alpha} + 2^{\frac{2}{\alpha}-1} (2/\alpha)^{\frac{1}{\alpha}} \|X_1\|_{\psi_\alpha} \leq \|X_1\|_{\psi_\alpha} \left(2^{\frac{1}{\alpha}} + \frac{1}{2} (8/\alpha)^{\frac{1}{\alpha}}\right) \leq c(8/\alpha)^{\frac{1}{\alpha}}. \end{aligned} \tag{5.7}$$

This concludes the proof of the lemma. □

Remark 5.2 If $(X)_{i \geq 0}$ is a 1-dependent, centered and stationary Markov chain such that $\|X_i\|_\infty \leq M < \infty$, then the assumptions of the above lemma are satisfied with $m = 2$

and $\mathcal{F}_i = \sigma \{X_j \mid j \leq i\}$. If $(\xi_i)_{i \geq 0}$ are i.i.d. random variables and $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ is a bounded, Borel function such that $X_i = f(\xi_i, \xi_{i+1})$ are centered, then we can take $\mathcal{F}_i = \sigma \{\xi_j \mid j \leq i + 1\}$ and notice that the assumptions of the above lemma are satisfied with $m = 1$.

Remark 5.3 It is worth noticing that σ_∞^2 may be equal to 0 in case of 1-dependent processes $(X_i)_{i \in \mathbb{N}}$. Take for example $X_i = \xi_{i+1} - \xi_i$ where $(\xi_i)_{i \in \mathbb{N}}$ are i.i.d. random variables. It turns out (cf. [14]) that the reverse is true, that is, if for a 1-dependent, bounded stationary process $(X_i)_{i \in \mathbb{N}}$ we have $\sigma_\infty^2 = 0$, then there exists an i.i.d. process $(\xi_i)_{i \in \mathbb{N}}$ such that $X_i = \xi_{i+1} - \xi_i$.

Lemma 5.4 (Bernstein inequality for random sums) *Let $(X_i)_{i \geq 0}$ be a 1-dependent sequence of centered random variables such that $\mathbb{E} \exp(c^{-\alpha} |X_i|^\alpha) \leq 2$ for some $\alpha \in (0, 1]$ and $c \geq 1$. Moreover, let $N \leq n \in \mathbb{N}$ be an \mathbb{N} -valued bounded random variable. Assume that we can find a filtration $\mathcal{F} = (\mathcal{F}_i)_{i \geq 0}$ such that for $Z_i = X_i + \mathbb{E}(X_{i+1} | \mathcal{F}_i) - \mathbb{E}(X_i | \mathcal{F}_{i-1})$ we have the following:*

- (0) X_i is \mathcal{F}_i measurable,
- (1) N is a stopping time with respect to \mathcal{F} ,
- (2) $(Z_i)_{i \geq 1}$ is stationary,
- (3) For each $j \in \mathbb{N}$ process, $(Z_i)_{i \geq j+3}$ is independent of \mathcal{F}_j ,
- (4) $(\mathbb{E}(X_i | \mathcal{F}_{i-1}))_{i \geq 1}$ is stationary,
- (5) $\mathbb{E}(X_i | \mathcal{F}_{i-1})$ is independent of X_{i+1} for all $i \geq 1$.

Then for any $t > 0$ and $a > 0$,

$$\mathbb{P} \left(\left| \sum_{i=1}^N X_i \right| > t \right) \leq 4 \exp(8) \exp \left(-\frac{t^\alpha}{uc^\alpha} \right) + 9 \exp \left(-\frac{t^2}{v\sigma_\infty^2 + wt} \right), \tag{5.8}$$

where $u = \frac{16 \cdot 26^\alpha}{\alpha}$, $v = 102a$, $w = 14M \max(2, \sqrt{\|(\lceil N/3 \rceil - a + 1)_+\| \psi_1})$ and $M = c(24\alpha^{-3} \log n)^{\frac{1}{\alpha}}$.

Proof Observe that 0) and 3) imply 2-dependence of the process $(Z_i)_{i \geq 1}$. Therefore, the filtration \mathcal{F} satisfies all the assumptions of Lemma 5.1 and thus (5.1) holds. Note also that without loss of generality, we may assume that $t \geq w \log 9$. (Otherwise, the right-hand side of (5.8) is at least one.) Fix $s = (8\sqrt{2} \log 9)^{-1}$. Using the union bound, we get $\mathbb{E}_i(\cdot) = \mathbb{E}(\cdot | \mathcal{F}_i)$

$$\mathbb{P} \left(\left| \sum_{i=1}^N X_i \right| > t \right) \leq \mathbb{P} \left(\left| \sum_{i=1}^N Z_i \right| > t(1-s) \right) + 2\mathbb{P} \left(\sup_{1 \leq i \leq n} |\mathbb{E}_{i-1} X_i| > \frac{ts}{2} \right). \tag{5.9}$$

Now using Lemma A.5, $ts/2 \geq c(\frac{2}{\alpha})^{\frac{1}{\alpha}}$, $t \geq w \log 9$ and $n \exp \left(-\frac{(st)^\alpha}{4(2c)^\alpha} \right) \leq 1$, we obtain

$$2\mathbb{P} \left(\sup_{1 \leq i \leq n} |\mathbb{E}_{i-1} X_i| > \frac{st}{2} \right) \leq 2n\mathbb{P} \left(|\mathbb{E}_0 X_1| > \frac{st}{2} \right) \leq 12 \exp \left(-\frac{(st)^\alpha}{4(2c)^\alpha} \right). \tag{5.10}$$

Next, we take care of the other term on the right-hand side of (5.9). Firstly, we split the sum

$$\mathbb{P}\left(\left|\sum_{i=1}^N Z_i\right| > t(1-s)\right) \leq \sum_{j=0}^2 \mathbb{P}\left(\left|\sum_{1 \leq i \leq N, 3|(i+j)} Z_i\right| > \frac{t(1-s)}{3}\right). \tag{5.11}$$

Now, we will consider the j th summand of the above sum. Let us take $r = \frac{3}{8\sqrt{2}\log(9)}$ and notice that there is function $f_j : \mathbb{N} \rightarrow \mathbb{N}$ such that for any $n \in \mathbb{N}$, $\lfloor \frac{n}{3} \rfloor \leq f_j(n) \leq \lceil \frac{n}{3} \rceil$ and

$$\begin{aligned} \mathbb{P}\left(\left|\sum_{1 \leq i \leq N, 3|i+j} Z_i\right| > t(1-s)/3\right) &= \mathbb{P}\left(\left|\sum_{1 \leq i \leq f_j(N)} Z_{3i-j}\right| > t(1-s)/3\right) \\ &\leq \mathbb{P}\left(\left|\sum_{1 \leq i \leq \lceil N/3 \rceil + 1} Z_{3i-j}\right| > t(1-r)(1-s)/3\right) + \mathbb{P}\left(2 \sup_{k \leq n+6} |Z_k| > tr(1-s)/3\right). \end{aligned} \tag{5.12}$$

Due to $\|Z_i\|_{\psi_\alpha} \leq c(8/\alpha)^{\frac{1}{\alpha}}$ (cf. (5.1)) and Lemma A.4 along with $t \geq w \log(9)$, $n \geq 2$ (for $n = 1$, the result of the lemma is trivial), we get

$$\begin{aligned} \mathbb{P}\left(2 \sup_{k \leq n+6} |Z_k| > \frac{tr(1-s)}{3}\right) &\leq (n+6)\mathbb{P}\left(|Z_k| > \frac{tr(1-s)}{3}\right) \\ &\leq 2(n+6) \exp\left(-\frac{\alpha(tr(1-s))^\alpha}{8(3c)^\alpha}\right) \leq 2 \exp\left(-\frac{\alpha(tr(1-s))^\alpha}{16(3c)^\alpha}\right). \end{aligned} \tag{5.13}$$

To handle the first summand on the right-hand side of (5.12), let us fix j and denote $\gamma_i := Z_{3i+3-j}$, $\mathcal{G}_i := \mathcal{F}_{3i-j}$, $T := \lceil N/3 + 1 \rceil \leq \lceil n/3 \rceil + 1$. Using the assumptions on the filtration \mathcal{F} and (5.1), it is straightforward to check that the following properties hold:

1. γ_i are independent,
2. $\mathbb{E}\gamma_i = 0$, $\mathbb{E}\gamma_i^2 = \sigma_\infty^2$, $\|\gamma_i\|_{\psi_\alpha} \leq c(8/\alpha)^{\frac{1}{\alpha}}$,
3. γ_{i-1} is \mathcal{G}_i measurable,
4. γ_i is independent of \mathcal{G}_i ,
5. T is a stopping time with respect to the filtration \mathcal{G}_i .

This is precisely the setting of Proposition 4.4. ii) from [2] which applied with $\epsilon := 1$, $p := \frac{\sqrt{2}}{\sqrt{2}-1}$ and $q := \sqrt{2}$ gives that for any $a > 0$,

$$\begin{aligned}
 & \mathbb{P} \left(\left| \sum_{1 \leq i \leq \lceil N/3 \rceil + 1} Z_{3i-j} \right| > t(1-r)(1-s)/3 \right) \\
 & \leq \exp(8) \exp \left(-\frac{(t(1-r)(1-s))^\alpha}{2(3(2+\sqrt{2})\hat{c})^\alpha} \right) \\
 & \quad + 3 \exp \left(-\frac{(t(1-r)(1-s))^2}{72a\sigma_\infty^2 + 6\sqrt{2}\mu(1-r)(1-s)t} \right),
 \end{aligned} \tag{5.14}$$

where

$$\mu = \max \left(\frac{8M}{3}, 2\sigma_\infty \sqrt{\|(\lceil N/3 \rceil - a + 1)_+\| \psi_1} \right), \quad \hat{c} = c \left(\frac{8}{\alpha} \right)^{\frac{1}{\alpha}}.$$

Using (5.1), Lemma A.2 with $Y = \frac{\alpha Z^\alpha}{8c^\alpha}$ and $\beta = \frac{2}{\alpha}$, together with the gamma function estimate $\Gamma(x) \leq \left(\frac{x}{2}\right)^{x-1}$ for $x \geq 2$ (see Theorem 1 in [18]), we get

$$\sigma_\infty^2 = \mathbb{E}Z_1^2 \leq 2c^2 \left(\frac{8}{\alpha}\right)^{\frac{2}{\alpha}} \Gamma\left(\frac{2}{\alpha} + 1\right) \leq \frac{4}{\alpha} c^2 \left(\frac{8}{\alpha}\right)^{\frac{2}{\alpha}} \Gamma\left(\frac{2}{\alpha}\right) \leq 4c^2 \left(\frac{8}{\alpha^2}\right)^{\frac{2}{\alpha}},$$

which implies that $\sigma_\infty \leq \frac{2}{3}M$ and as a consequence,

$$\mu \leq \frac{4}{3}Mb, \quad b = \max \left(2, \sqrt{\|(\lceil N/3 \rceil - a + 1)_+\| \psi_1} \right).$$

Therefore, (5.14) reduces to

$$\begin{aligned}
 & \mathbb{P} \left(\left| \sum_{1 \leq i \leq \lceil N/3 \rceil + 1} Z_{3i-j} \right| > t(1-r)(1-s)/3 \right) \\
 & \leq \exp(8) \exp \left(-\frac{(t(1-r)(1-s))^\alpha}{2(3(2+\sqrt{2})\hat{c})^\alpha} \right) \\
 & \quad + 3 \exp \left(-\frac{(t(1-r)(1-s))^2}{72a\sigma_\infty^2 + 8\sqrt{2}Mb(1-r)(1-s)t} \right).
 \end{aligned}$$

Combining the above inequality with (5.9)–(5.13), we obtain

$$\begin{aligned}
 & \mathbb{P} \left(\left| \sum_{i=1}^N X_i \right| > t \right) \leq 12 \exp \left(-\frac{(st)^\alpha}{4(2c)^\alpha} \right) + 6 \exp \left(-\frac{\alpha(tr(1-s))^\alpha}{16(3c)^\alpha} \right) \\
 & \quad + 3 \exp(8) \exp \left(-\frac{(t(1-r)(1-s))^\alpha}{2(3(2+\sqrt{2})\hat{c})^\alpha} \right) \\
 & \quad + 9 \exp \left(-\frac{(t(1-r)(1-s))^2}{72a\sigma_\infty^2 + 8\sqrt{2}Mb(1-r)(1-s)t} \right).
 \end{aligned}$$

To conclude, it is now enough to recall that $r = 3(8\sqrt{2} \log(9))^{-1}$, $s = (8\sqrt{2} \log 9)^{-1}$ and do some elementary calculations. □

6 Bounds on the Number of Regenerations

We will now obtain a bound on the stopping time N , introduced in (3.20). To this end, we will use the ψ_1 version of Bernstein inequality, which follows easily from the classical moment version of this inequality (see, e.g., Lemma 2.2.11 in [26]), by observing that for $k \geq 2$, $\mathbb{E}|\xi|^k \leq k! \|\xi\|_{\psi_1}^k = k! M^{k-2} v/2$, where $M = \|\xi\|_{\psi_1}$, $v = 2\|\xi\|_{\psi_1}^2$.

Lemma 6.1 (*ψ_1 Bernstein’s inequality.*) *If $(\xi_i)_i$ is a sequence of independent centered random variables such that $\sup_i \|\xi_i\|_{\psi_1} \leq \tau$, then*

$$\mathbb{P}\left(\sum_{i=1}^n \xi_i \geq t\right) \leq \exp\left(-\frac{t^2}{4n\tau^2 + 2\tau t}\right).$$

Lemma 6.2 *If $\|\tau_1 - \tau_0\|_{\psi_1} \leq d$, then for any $p > 0$,*

$$\mathbb{P}\left(N > \left[(1+p)n [\mathbb{E}(\tau_1 - \tau_0)]^{-1}\right]\right) \leq \exp(1) \exp\left(-\frac{pn\mathbb{E}(\tau_1 - \tau_0)}{K_p d^2}\right), \tag{6.1}$$

where $K_p = L_p + 16/L_p$ and $L_p = \frac{16}{p} + 20$. Moreover, the function $p \rightarrow K_p$ is decreasing on \mathbb{R}_+ (in particular; $K_p \geq K_\infty = \frac{104}{5}$) and if $p = 2/3$, then $\frac{1}{p}K_p \leq 67$.

Proof For convenience, let $T_i = \tau_i - \tau_{i-1}$ for $i \geq 1$. Firstly, notice that without loss of generality, we may assume that $np \geq L_p \mathbb{E}T_1$. Indeed, otherwise, using $\mathbb{E}T_1 \leq d$ we obtain

$$\exp(1) \exp\left(-\frac{pn\mathbb{E}T_1}{K_p d^2}\right) \geq \exp(1) \exp\left(-\frac{L_p \mathbb{E}^2 T_1}{K_p d^2}\right) \geq \exp\left(1 - \frac{L_p}{K_p}\right) \geq 1.$$

Thus, from now on we consider n such that $np \geq L_p \mathbb{E}T_1$. For $A = (1+p)n [\mathbb{E}T_1]^{-1} \geq 1$, we get

$$\begin{aligned} \mathbb{P}(N > \lceil A \rceil) &\leq \mathbb{P}(\tau_{\lceil A \rceil} - \tau_0 \leq n) \leq \mathbb{P}\left(\sum_{i=0}^{\lceil A \rceil - 1} T_{i+1} - \mathbb{E}T_{i+1} \leq n - A\mathbb{E}T_1\right) \\ &= \mathbb{P}\left(\sum_{i=0}^{\lceil A \rceil - 1} T_{i+1} - \mathbb{E}T_{i+1} \leq n - (1+p)n\right) \\ &= \mathbb{P}\left(\sum_{i=0}^{\lceil A \rceil - 1} T_{i+1} - \mathbb{E}T_{i+1} \leq -np\right). \end{aligned} \tag{6.2}$$

Now, we have $\|T_{i+1} - \mathbb{E}T_{i+1}\|_{\psi_1} \leq 2d$, so using Lemma 6.1, $\mathbb{E}T_1 \leq d$ and $np \geq L_p \mathbb{E}T_1$, we get

$$\begin{aligned} \mathbb{P}\left(N > \left\lceil (1+p)n [\mathbb{E}T_1]^{-1} \right\rceil\right) &\leq \exp\left(-\frac{p^2n^2}{4(A+1)4d^2 + 4dnp}\right) \\ &= \exp\left(-\frac{p^2n^2}{16d^2 \left[(1+p)n [\mathbb{E}T_1]^{-1} + 1\right] + 4dnp}\right) \\ &= \exp\left(-\frac{pn\mathbb{E}T_1}{16d^2 \left(\frac{1+p}{p} + \frac{\mathbb{E}T_1}{pn}\right) + 4d\mathbb{E}T_1}\right) \\ &\leq \exp\left(-\frac{pn\mathbb{E}T_1}{16d^2 \left(\frac{1+p}{p} + \frac{1}{L_p}\right) + 4d^2}\right) = \exp\left(-\frac{pn\mathbb{E}T_1}{K_p d^2}\right) \\ &\leq \exp\left(1 - \frac{pn\mathbb{E}T_1}{K_p d^2}\right), \end{aligned}$$

which finishes the proof of (6.1). The properties of K_p follow from easy computations. □

The following lemma is a standard consequence of the tail estimates given in Lemma 6.2. Its proof, based on integration by parts, is analogous to that of Lemma 5.4 in [2] and is therefore omitted.

Lemma 6.3 *Suppose that $\|\tau_1 - \tau_0\|_{\psi_1} \leq d$ for some $d > 0$. Then for any $p > 0$,*

$$\|(N - a)_+\|_{\psi_1} \leq \frac{4K_p d^2}{[\mathbb{E}(\tau_1 - \tau_0)]^2} \leq \frac{4K_p d^2}{m^2},$$

where $a = (1+p)n [\mathbb{E}(\tau_1 - \tau_0)]^{-1}$, $K_p = L_p + \frac{16}{L_p}$ and $L_p = \frac{16}{p} + 20$. Moreover,

$$\frac{d^2 K_p}{[\mathbb{E}(\tau_1 - \tau_0)]^2} \geq K_p \geq K_\infty.$$

7 Proofs of Theorems 4.1, 4.3 and 4.4

In this section, we will prove our main results. The structure of proofs of Theorems 4.1 and 4.3 is similar, and they contain a common part, which we will present first in Sects. 7.1 and 7.2. The proof of Theorem 4.1 will be concluded in Sect. 7.3 and the proof of Theorem 4.3 in Sect. 7.4. Theorem 4.4 will be obtained as a corollary to Theorem 4.1 in Sect. 7.5.

Let us thus pass to the proofs of Theorems 4.1 and 4.3. Assume that $m|n$. The argument will be based on the approach of [1] and [2] (see also [10] and [12]) and will

rely on the decomposition

$$\left| \sum_{i=0}^{n-1} f(\Upsilon_i) \right| \leq H_n + M_n + T_n, \tag{7.1}$$

where

$$H_n = \left| \sum_{i=0}^{\tau_0/m} \Theta_i \mathbb{1}_{N>0} + \mathbb{1}_{N=0} \sum_{i=0}^{n/m-1} \Theta_i \right|, \quad M_n = \left| \sum_{i=1}^N \chi_{i-1}(f) \right|, \tag{7.2}$$

$$T_n = \left| \mathbb{1}_{N>0} \sum_{k=n}^{\tau_N+m-1} f(\Upsilon_k) \right|, \quad N = \inf\{i \geq 0 \mid \tau_i + m - 1 \geq n - 1\}.$$

The proof will be divided into three main steps. In the first two (common for both theorems), we will get easy bounds on tails of H_n and T_n . The main, third step will be devoted to obtaining two different estimates on the tail of M_n . To this end, we will use Lemmas 5.1, 6.2 (for the proof of Theorem 4.1) and Lemmas 5.4, 6.3 (for Theorem 4.3).

7.1 Estimate on H_n

Using $\{N = 0\} \subset \{\tau_0 \geq n - m\}$, the definition of \mathbf{a} (see (4.2)) and Lemma A.4, we get

$$\begin{aligned} \mathbb{P}_{x^*}(H_n > t) &\leq \mathbb{P}_{x^*} \left(\mathbb{1}_{N>0} \sum_{i=0}^{\tau_0/m} |\Theta_i| + \mathbb{1}_{N=0} \sum_{i=0}^{n/m-1} |\Theta_i| > t \right) \\ &\leq \mathbb{P}_{x^*} \left(\sum_{i=0}^{\tau_0/m} |\Theta_i| > t \right) \\ &\leq 2 \exp \left(-\frac{t^\alpha}{\mathbf{a}^\alpha} \right). \end{aligned} \tag{7.3}$$

7.2 Estimate on T_n

By repeating verbatim the easy argument presented in the proof of Theorem 5.1 in [2], we obtain

$$\mathbb{P}(|T_n| > t) \leq 2 [\delta\pi(C)]^{-1} \exp \left(-\frac{t^\alpha}{\mathbf{b}^\alpha} \right). \tag{7.4}$$

We skip the details.

7.3 Proof of Theorem 4.1

Recall that $M = c(24\alpha^{-3} \log n)^{\frac{1}{\alpha}}$ and note that without loss of generality, we can assume that $t \geq 8M \log 6$. Otherwise, (4.3) is trivial as the right-hand side is greater than or equal to 1. Fix $p = 2/3$. We have $(A := \lceil (p + 1)n(\mathbb{E}(\tau_1 - \tau_0))^{-1} \rceil)$

$$\begin{aligned} \mathbb{P}(M_n \geq t) &= \mathbb{P}(M_n \geq t, N \leq A) + \mathbb{P}(M_n \geq t, N > A) \\ &\leq \mathbb{P}\left(\sup_{1 \leq k \leq A} \left| \sum_{i=1}^k \chi_{i-1} \right| \geq t\right) + \mathbb{P}(N > A). \end{aligned} \tag{7.5}$$

To control the first summand on the right-hand side of the above inequality, we will apply Lemma 5.1 with $m = 2$, $X_i := \chi_i = F(\Xi_{i+1})$ (cf. (3.16)), $c := c$ and $n := A$. Assuming that the assumptions of the lemma are satisfied (we will verify them later on), we obtain (in the first line, we use stationarity of $(\Xi_i)_{i \geq 1}$):

$$\begin{aligned} P &:= \mathbb{P}\left(\sup_{1 \leq k \leq A} \left| \sum_{i=1}^k \chi_{i-1} \right| \geq t\right) = \mathbb{P}\left(\sup_{1 \leq k \leq A} \left| \sum_{i=1}^k F(\Xi_{i+1}) \right| \geq t\right) \\ &\leq 6 \exp(8) \exp\left(-\frac{t^\alpha}{\frac{16}{\alpha}(24c)^\alpha}\right) \\ &\quad + 6 \exp\left(-\frac{t^2}{15(\lceil (p + 1)n(\mathbb{E}(\tau_1 - \tau_0))^{-1} \rceil + 3)\sigma_\infty^2 + 6tM}\right) \\ &\leq 6 \exp(8) \exp\left(-\frac{t^\alpha}{\frac{16}{\alpha}(24c)^\alpha}\right) \\ &\quad + 6 \exp\left(-\frac{t^2}{15((p + 1)n(\mathbb{E}(\tau_1 - \tau_0))^{-1} + 4)\sigma_\infty^2 + 6tM}\right) \end{aligned} \tag{7.6}$$

Recall that by (3.22), $\sigma_\infty^2 = \sigma_{Mrv}^2 \mathbb{E}(\tau_1 - \tau_0)$. We will now obtain a comparison between σ_∞^2 and tM , which will allow us to reduce the above estimate to one in which the sub-Gaussian coefficient is expressed only in terms of σ_{Mrv}^2 . Thanks to Lemma A.2 applied with $\Lambda := (\chi_1/c)^\alpha$ and $\beta := 2/\alpha$, we have

$$\sigma_\infty^2 \leq 3\mathbb{E}\chi_1^2 \leq 3c^2\Gamma(2/\alpha + 1) \leq 3c^2(2/\alpha)^{\frac{2}{\alpha}+1},$$

where the last inequality is a consequence of equation 4 in [18]. Moreover, recalling the definition of M and using the assumption $t \geq 8 \log(6)M$, we obtain

$$tM \geq 8 \log(6)M^2 = 8 \log(6)c^2(24\alpha^{-3} \log(n))^{\frac{2}{\alpha}} \geq 16 \cdot 8 \log(6)c^2(2/\alpha)^{\frac{2}{\alpha}+1} \geq 76\sigma_\infty^2.$$

The last inequality in combination with (7.6) yields

$$\begin{aligned}
 P &\leq 6 \exp(8) \exp\left(-\frac{t^\alpha}{\frac{16}{\alpha}(24\mathbf{c})^\alpha}\right) + 6 \exp\left(-\frac{t^2}{15(p+1)n\sigma_{Mrv}^2 + 7tM}\right) \\
 &\leq 6 \exp(8) \exp\left(-\frac{t^\alpha}{\frac{16}{\alpha}(24\mathbf{c})^\alpha}\right) + 6 \exp\left(-\frac{t^2}{25n\sigma_{Mrv}^2 + 7tM}\right).
 \end{aligned}
 \tag{7.7}$$

In order to justify the above inequality, it remains to verify the assumptions of Lemma 5.1. To this end, take

$$\mathcal{F}_i = \sigma\{\Xi_j \mid j \leq i + 1\}, \quad Z_i = \chi_i + \mathbb{E}(\chi_{i+1}|\mathcal{F}_i) - \mathbb{E}(\chi_i|\mathcal{F}_{i-1}).$$

We will now strongly rely on the properties of the stationary sequence of 1-dependent blocks $(\Xi_i)_{i \geq 1}$ stated in Remark 3.10 together with (3.15) and (3.16). Since $\chi_i = F(\Xi_{i+1})$, the assumption 0) of Lemma 5.1 is trivially satisfied. To prove 2), observe that $\mathbb{E}(\chi_{i+1}|\mathcal{F}_i) = \mathbb{E}(\chi_{i+1}|\Xi_{i+1}) = G(\Xi_{i+1})$ for some measurable function G , and so the sequence $(Z_i)_{i \geq 1} = (F(\Xi_{i+1}) + G(\Xi_{i+1}) - G(\Xi_i))_{i \geq 1}$ is stationary (as a function of the stationary sequence $(\Xi_i)_{i \geq 1}$). The sequence $(\Xi_i)_{i \geq 1}$ is 1-dependent, which clearly implies that $(Z_i)_{i \geq 1}$ is 2-dependent, i.e., the assumption 2) of the lemma. The assumption 3), i.e., the stationarity of the sequence $(\mathbb{E}(\chi_i|\mathcal{F}_{i-1}))_{i \geq 1} = (G(\Xi_i))_{i \geq 1}$, follows again by stationarity of $(\Xi_i)_{i \geq 1}$. Finally, using once more the fact that $(\Xi_i)_{i \geq 0}$ is 1-dependent, we obtain that for any $i \geq 1$, the random variable $\mathbb{E}(\chi_i|\mathcal{F}_{i-1}) = G(\Xi_i)$ is independent of $\chi_{i+1} = F(\Xi_{i+2})$, which ends the verification of the assumptions of Lemma 5.1 and proves (7.7).

Thus, in order to get a bound on $\mathbb{P}(M_n > t)$ it suffices to estimate the second term on the right-hand side of (7.5). To this aim, we use Lemma 6.2 with $p = 2/3$ and $d = \mathbf{d}$ obtaining

$$\mathbb{P}\left(N > \left\lceil (1+p)n(\mathbb{E}(\tau_1 - \tau_0))^{-1} \right\rceil\right) \leq \exp(1) \exp\left(-\frac{n\mathbb{E}(\tau_1 - \tau_0)}{67\mathbf{d}^2}\right).$$

In combination with (7.5) and (7.7), this gives

$$\begin{aligned}
 \mathbb{P}(M_n \geq t) &\leq 6 \exp(8) \exp\left(-\frac{t^\alpha}{\frac{16}{\alpha}(24\mathbf{c})^\alpha}\right) + 6 \exp\left(-\frac{t^2}{25n\sigma_{Mrv}^2 + 7tM}\right) \\
 &\quad + \exp(1) \exp\left(-\frac{n\mathbb{E}(\tau_1 - \tau_0)}{67\mathbf{d}^2}\right).
 \end{aligned}$$

Combining the above inequality with (7.3) and (7.4), we get

$$\mathbb{P}_x\left(\left|\sum_{i=0}^{n-1} f(\Upsilon_i)\right| > t\right) \leq \mathbb{P}\left(H_n \geq \frac{1 - \sqrt{5/6}}{2}t\right) + \mathbb{P}\left(M_n \geq \sqrt{5/6}t\right)$$

$$\begin{aligned}
 &+ \mathbb{P}\left(T_n \geq \frac{1 - \sqrt{5/6}}{2}t\right) \\
 &\leq 2 \exp\left(-\frac{t^\alpha}{(23\mathbf{a})^\alpha}\right) + 2[\delta\pi(C)]^{-1} \exp\left(-\frac{t^\alpha}{(23\mathbf{b})^\alpha}\right) \\
 &+ \exp(1) \exp\left(-\frac{n\mathbb{E}(\tau_1 - \tau)}{67\mathbf{d}^2}\right) \\
 &+ 6 \exp(8) \exp\left(-\frac{t^\alpha}{\frac{16}{\alpha}(27\mathbf{c})^\alpha}\right) + 6 \exp\left(-\frac{t^2}{30n\sigma_{Mrv}^2 + 8tM}\right).
 \end{aligned}$$

In order to finish the proof of Theorem 4.1, it is enough to substitute $\mathbb{E}(\tau_1 - \tau_0) = \delta^{-1}\pi(C)^{-1}m$.

7.4 Proof of Theorem 4.3

Recall that $M = \mathbf{c}(24\alpha^{-3} \log n)^{\frac{1}{\alpha}}$, and let $p > 0$ be a parameter which will be fixed later on. We are going to apply Lemma 5.4 with $X_i := \chi_i = F(\Xi_{i+1})$, $c := \mathbf{c}$, $\mathcal{F}_i := \sigma\{\Xi_j \mid 0 \leq j \leq i + 1\}$. Clearly, N is a stopping time with respect to \mathcal{F} . The remaining assumptions of Lemma 5.1 can be verified in the same manner as in the proof of Theorem 4.1. Let $a = (1 + p)^{\frac{n}{3}} [\mathbb{E}(\tau_1 - \tau_0)]^{-1}$. By Lemma 6.3, we get

$$\begin{aligned}
 \|(\lceil N/3 \rceil - a + 1)_+\|_{\psi_1} &\leq \frac{1}{3} \left\| \left(N - (1 + p)n(\mathbb{E}(\tau_1 - \tau_0))^{-1} \right)_+ \right\|_{\psi_1} + \frac{2}{\log 2} \\
 &\leq \frac{4}{3} \mathbf{d}^2 K_p + \frac{2}{\log 2} \leq \left(\frac{4}{3} + \frac{7}{50} \right) \mathbf{d}^2 K_p,
 \end{aligned}$$

where the last inequality follows from (recall the definition of K_∞ from Lemma 6.2)

$$\frac{7}{50} K_p \geq \frac{7}{50} K_\infty = \frac{7}{50} \cdot \frac{104}{5} \geq \frac{2}{\log 2}.$$

Therefore, $\max(2, \sqrt{\|(\lceil N/3 \rceil - a + 1)_+\|_{\psi_1}}) \leq \sqrt{4/3 + 7/50} \sqrt{K_p} \cdot \mathbf{d}$ and we get that for arbitrary $p > 0$,

$$\begin{aligned}
 \mathbb{P}\left(\left|\sum_{i=1}^N \chi_{i-1}\right| > t\right) &\leq 4 \exp(8) \exp\left(-\frac{t^\alpha}{\frac{16}{\alpha}(26\mathbf{c})^\alpha}\right) \\
 &+ 9 \exp\left(-\frac{t^2}{34(1 + p)\sigma_{Mrv}^2 + 17M\mathbf{d}t\sqrt{K_p}}\right).
 \end{aligned}$$

Using the above inequality together with (7.3) and (7.4), we obtain

$$\mathbb{P}_x\left(\left|\sum_{i=0}^{n-1} f(\Upsilon_i)\right| > t\right) \leq \mathbb{P}\left(H_n \geq \frac{t}{54}\right) + \mathbb{P}\left(M_n \geq \frac{26t}{27}\right) + \mathbb{P}\left(T_n \geq \frac{t}{54}\right)$$

$$\begin{aligned} &\leq 2 \exp\left(-\frac{t^\alpha}{(54\mathbf{a})^\alpha}\right) + 2 [\delta\pi(C)]^{-1} \exp\left(-\frac{t^\alpha}{(54\mathbf{b})^\alpha}\right) \\ &\quad + 4 \exp(8) \exp\left(-\frac{t^\alpha}{\frac{16}{\alpha}(27\mathbf{c})^\alpha}\right) \\ &\quad + 9 \exp\left(-\frac{t^2}{37(1+p)\sigma_{Mrv}^2 + 18M\mathbf{d}t\sqrt{K_p}}\right), \end{aligned}$$

which concludes the proof of Theorem 4.3.

7.5 Proof of Theorem 4.4.

Denote $M = \|f\|_\infty$ and notice that for $t > nM$, the left-hand side of (4.6) vanishes, so we may assume that $t \leq nM$. Using (4.4), one can easily see that if $m|n$, then Theorem 4.1 applied with $\alpha = 1$ implies that

$$\begin{aligned} \mathbb{P}_x\left(\left|\sum_{i=0}^{n-1} f(\Upsilon_i)\right| > t\right) &\leq \left(2 + 2 [\delta\pi(C)]^{-1}\right) \exp\left(-\frac{t}{46DM}\right) \\ &\quad + 6 \exp(8) \exp\left(-\frac{t}{432DM}\right) \tag{7.8} \\ &\quad + 6 \exp\left(-\frac{t^2}{30n\sigma_{Mrv}^2 + 192tDM}\right) + \exp(1) \exp\left(-\frac{nm}{67\delta\pi(C)D^2}\right). \end{aligned}$$

The assumption $t \leq nM$ yields

$$\exp\left(-\frac{nm}{67\delta\pi(C)D^2}\right) \leq \exp\left(-\frac{tm}{67\delta\pi(C)MD^2}\right),$$

which plugged into (7.8) gives after some elementary calculations that (recall $K = \exp(10) + 2 [\delta\pi(C)]^{-1}$)

$$\mathbb{P}_x\left(\left|\sum_{i=0}^{n-1} f(\Upsilon_i)\right| > t\right) \leq K \exp\left(-\frac{t^2}{30n\sigma_{Mrv}^2 + 432tD^2M\delta\pi(C)\log n}\right), \tag{7.9}$$

proving the theorem in the special case $m|n$.

Now, we consider the case $m \nmid n$. Define $\lceil n \rceil_m$ to be the smallest integer greater or equal to n , which is divisible by m . Notice that without loss of generality, we can assume that $t > 4330D^2M\delta\pi(C)$. (Otherwise, the assertion of the theorem is trivial as the right-hand side of (4.6) exceeds one.) Since $D^2\delta\pi(C) > m$ (recall $\mathbb{E}(\tau_1 - \tau_0) = \delta^{-1}\pi(C)^{-1}m$), this implies that $t \geq 4330Mm$. Moreover, as $t \leq nM$, we also obtain that $n \geq 4330m$.

Thus, for $p = 1/4330$ we have $\left| \sum_{i=n}^{\lceil n \rceil_m} f(\Upsilon_i) \right| \leq Mm \leq pt$, and as a consequence,

$$\mathbb{P}_x \left(\left| \sum_{i=0}^{n-1} f(\Upsilon_i) \right| > t \right) \leq \mathbb{P}_x \left(\left| \sum_{i=0}^{\lceil n \rceil_m - 1} f(\Upsilon_i) \right| > (1 - p)t \right). \tag{7.10}$$

Now using (7.9) and the inequality $n > 4330m$, we get

$$\begin{aligned} \mathbb{P}_x \left(\left| \sum_{i=0}^{n-1} f(\Upsilon_i) \right| > t \right) &\leq K \exp \left(- \frac{t^2}{31 \lceil n \rceil_m \sigma_{Mrv}^2 + 433t D^2 M \delta \pi(C) \log n} \right) \\ &\leq K \exp \left(- \frac{t^2}{31(n + m) \sigma_{Mrv}^2 + 433t D^2 M \delta \pi(C) \log n} \right) \\ &\leq K \exp \left(- \frac{t^2}{32n \sigma_{Mrv}^2 + 433t D^2 M \delta \pi(C) \log n} \right). \end{aligned}$$

This concludes the proof of Theorem 4.4.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

A Orlicz Exponential Norm

At the beginning, recall the definition of the exponential Orlicz quasi-norm (4.1) and note that if $\alpha \geq 1$, then $\| \cdot \|_{\psi_\alpha}$ is a norm, whereas for $0 < \alpha < 1$, $\| \cdot \|_{\psi_\alpha}$ is only a quasi-norm. More precisely, we have the following version of the triangle inequality (see Lemma 3.7 in [2]).

Lemma A.1 (Triangle inequality for $\alpha \leq 1$) *Fix $0 < \alpha \leq 1$. Then, for any random variables X, Y we have*

$$\|X + Y\|_{\psi_\alpha} \leq \left(\|X\|_{\psi_\alpha}^\alpha + \|Y\|_{\psi_\alpha}^\alpha \right)^{1/\alpha} \leq 2^{1/\alpha-1} (\|X\|_{\psi_\alpha} + \|Y\|_{\psi_\alpha}).$$

Now, we present a moment estimation for random variables with bounded exponential moment.

Lemma A.2 *If Y is nonnegative random variable such that $\mathbb{E} \exp(Y) \leq 2$, then for any $\beta > 0$, we have*

$$\mathbb{E}Y^\beta \leq 2\Gamma(\beta + 1).$$

Furthermore, if $\beta \in \mathbb{N}$, then one can replace the constant 2 with 1.

Proof If β is a natural number, then the claim follows from Taylor’s expansion of $\exp(x)$. The general case is obtained by Markov’s inequality, namely

$$\begin{aligned} \mathbb{E}Y^\beta &= \int_0^\infty \mathbb{P}(Y^\beta \geq t) dt = \int_0^\infty \mathbb{P}\left(\exp(Y) \geq \exp\left(t^{\frac{1}{\beta}}\right)\right) dt \\ &\leq \int_0^\infty \frac{2}{\exp\left(t^{\frac{1}{\beta}}\right)} dt = \int_0^\infty \exp(-s) 2\beta s^{\beta-1} ds = 2\beta\Gamma(\beta). \end{aligned} \tag{A.1}$$

□

The next lemma allows to pass from the ψ_α -norm of a random variable to the norm of its conditional expectation.

Lemma A.3 *(Orlicz’s norm of conditional mean value) Let $0 < \alpha \leq 1$. Assume that a random variable X satisfies $\|X\|_{\psi_\alpha} < \infty$. Moreover, let \mathcal{F} be some sigma field. Then,*

$$\|\mathbb{E}(X|\mathcal{F})\|_{\psi_\alpha} \leq \left(1 + \frac{\log\left(\alpha \exp\left(\frac{1-\alpha}{\alpha}\right)\right)}{\log(2)}\right)^{\frac{1}{\alpha}} \|X\|_{\psi_\alpha} \leq \left(\frac{2}{\alpha}\right)^{\frac{1}{\alpha}} \|X\|_{\psi_\alpha}.$$

Proof Set $\phi_\alpha(x) = \exp(x^\alpha)$ for $x \geq 0$ and notice that ϕ_α is concave on $(0, x_\alpha)$ and convex on (x_α, ∞) , where $x_\alpha = \left(\frac{1-\alpha}{\alpha}\right)^{1/\alpha}$. Define Ψ_α to be a smallest convex function bigger or equal to ϕ_α which is equal to ϕ_α on (x_α, ∞) , that is,

$$\Psi_\alpha(x) = \begin{cases} \alpha \exp\left(\frac{1-\alpha}{\alpha}\right) (x x_\alpha^{\alpha-1} + 1), & \text{if } 0 \leq x \leq x_\alpha, \\ \phi_\alpha(x), & \text{if } x \geq x_\alpha. \end{cases}$$

Clearly, Ψ_α is a convex function on \mathbb{R}_+ and it is easy to see that $\phi_\alpha \leq \Psi_\alpha \leq \alpha \exp\left(\frac{1-\alpha}{\alpha}\right) \phi_\alpha$. Using these properties, Jensen’s inequality and the definition of the Orlicz norm, we get

$$\mathbb{E}\phi_\alpha\left(\frac{|\mathbb{E}(X|\mathcal{F})|}{\|X\|_{\psi_\alpha}}\right) \leq \mathbb{E}\Psi_\alpha\left(\frac{|\mathbb{E}(X|\mathcal{F})|}{\|X\|_{\psi_\alpha}}\right) \leq \mathbb{E}\Psi_\alpha\left(\frac{|X|}{\|X\|_{\psi_\alpha}}\right) \leq 2\alpha \exp\left(\frac{1-\alpha}{\alpha}\right).$$

Put $c_\alpha = \left(1 + \frac{\log\left(\alpha \exp\left(\frac{1-\alpha}{\alpha}\right)\right)}{\log(2)}\right)^{\frac{1}{\alpha}} \geq 1$ and note that due to Jensen’s inequality,

$$\mathbb{E}\phi_\alpha\left(\frac{|\mathbb{E}(X|\mathcal{F})|}{c_\alpha \|X\|_{\psi_\alpha}}\right) \leq \left(\mathbb{E}\phi_\alpha\left(\frac{|\mathbb{E}(X|\mathcal{F})|}{\|X\|_{\psi_\alpha}}\right)\right)^{\frac{1}{c_\alpha}} \leq 2,$$

which completes the proof. □

Now, we give two concentration inequalities which are valid for random variables with finite Orlicz norm. The first one is an easy consequence of the Markov inequality; therefore, we omit the proof.

Lemma A.4 *For any random variable X with $\|X\|_{\psi_\alpha} < \infty$ and $t > 0$, we have*

$$\mathbb{P}(|X| \geq t) \leq 2 \exp\left(-\frac{t^\alpha}{\|X\|_{\psi_\alpha}^\alpha}\right).$$

Lemma A.5 *(Tail inequality for conditional mean value) Let $0 < \alpha \leq 1$. Assume that a random variable X satisfies $\|X\|_{\psi_\alpha} < \infty$. Moreover, let \mathcal{F} be some sigma field. Then for any $t \geq \left(\frac{2}{\alpha}\right)^{1/\alpha} \|X\|_{\psi_\alpha}$, we have*

$$\mathbb{P}(|\mathbb{E}(X|\mathcal{F})| > t) \leq 6 \exp\left(-\frac{t^\alpha}{2\|X\|_{\psi_\alpha}^\alpha}\right).$$

Proof Fix $c > \|X\|_{\psi_\alpha}$ and $t \geq \left(\frac{2}{\alpha}\right)^{1/\alpha} c$. Then in particular, we have $\alpha \left(\frac{t}{c}\right)^\alpha \geq 2$. Using the Markov and Jensen inequalities along with $\Gamma(x) \leq x^x/e^{x-1}$ ([18], Thm. 1) and Lemma A.2 with $Y = (|X|/c)^\alpha$, $\beta = t^\alpha/c^\alpha$, we get

$$\begin{aligned} \mathbb{P}(|\mathbb{E}(X|\mathcal{F})| > t) &\leq \mathbb{P}\left(|\mathbb{E}(X|\mathcal{F})|^{\alpha \frac{t}{c}} > t^\alpha \frac{t}{c^\alpha}\right) \leq t^{-\alpha \left(\frac{t}{c}\right)^\alpha} \mathbb{E}|\mathbb{E}(X|\mathcal{F})|^{\alpha \left(\frac{t}{c}\right)^\alpha} \\ &\leq t^{-\alpha \left(\frac{t}{c}\right)^\alpha} \mathbb{E}|X|^{\alpha \left(\frac{t}{c}\right)^\alpha} \\ &= (t/c)^{-\alpha \left(\frac{t}{c}\right)^\alpha} \mathbb{E}|X/c|^{\alpha \left(\frac{t}{c}\right)^\alpha} \leq 2e (t/c)^\alpha \exp\left(- (t/c)^\alpha\right) \\ &\leq 2e \exp\left(- (1/2) (t/c)^\alpha\right), \end{aligned}$$

where in the last inequality we used the estimate $xe^{-x} \leq e^{-\frac{x}{2}}$ which is valid for all $x \in \mathbb{R}$. Now, it is enough to take limit $c \rightarrow \|X\|_{\psi_\alpha}$ and notice that $2e \leq 6$. □

References

1. Adamczak, R.: A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.* **13**(34), 1000–1034 (2008)
2. Adamczak, R., Bednorz, W.: Exponential concentration inequalities for additive functionals of Markov chains. *ESAIM Probab. Stat.* **19**, 440–481 (2015). ISSN: 1292-8100
3. Athreya, K.B., Ney, P.: A new approach to the limit theory of recurrent Markov chains. *Trans. Am. Math. Soc.* **245**, 493–501 (1978). ISSN: 0002-9947
4. Baxendale, P.H.: Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.* **15**(1B), 700–738 (2005). ISSN: 1050-5164
5. Bednorz, W., Łatuszyński, K., Latała, R.: A regeneration proof of the central limit theorem for uniformly ergodic Markov chains. *Electron. Commun. Probab.* **13**, 85–98 (2008)
6. Bertail, P., Ciolek, G.: New Bernstein and Hoeffding type inequalities for regenerative Markov chains. *ALEA Lat. Am. J. Probab. Math. Stat.* **16.1**, 259–277 (2019)

7. Bertail, P., Cléménçon, S.: Regenerative block bootstrap for Markov chains. *Bernoulli* **12**(4), 689–712 (2006). ISSN: 1350-7265
8. Burton, R.M., Goulet, M., Meester, R.: On 1-dependent processes and k-block factors. *Ann. Probab.* **21**(4), 2157–2168 (1993). ISSN: 0091-1798
9. Chen, X.: Limit theorems for functionals of ergodic Markov chains with general state space. *Mem. Am. Math. Soc.* **139.664**, xiv+203 (1999). ISSN: 0065-9266
10. Cléménçon, S.J.M.: Moment and probability inequalities for sums of bounded additive functionals of regular Markov chains via the Nummelin splitting technique. *Stat. Probab. Lett.* **55**(3), 227–238 (2001). ISSN: 0167-7152
11. Dedecker, J., Gouëzel, S.: Subgaussian concentration inequalities for geometrically ergodic Markov chains. *Electron. Commun. Probab.* **20**(64), 12 (2015)
12. Douc, R., Guillin, A., Moulines, E.: Bounds on regeneration times and limit theorems for subgeometric Markov chains. *Ann. Inst. Henri Poincaré Probab. Stat.* **44**(2), 239–257 (2008). ISSN: 0246-0203
13. Gao, F., Guillin, A., Wu, L.: Bernstein-type concentration inequalities for symmetric Markov processes. *Theory Probab. Appl.* **58**(3), 358–382 (2014). ISSN: 0040-585X
14. Janson, S.: On degenerate sums of m-dependent variables. *J. Appl. Probab.* **52**(4), 1146–1155 (2015). ISSN: 0021-9002
15. Jiang, B., Sun, Q., Fan, J.: Bernstein’s inequality for general Markov chains. In: arXiv preprint [arXiv:1805.10721](https://arxiv.org/abs/1805.10721) (2018)
16. Lezaud, P.: Chernoff and Berry–Esséen inequalities for Markov processes. *ESAIM Probab. Stat.* **5**, 183–201 (2001). ISSN: 1292-8100
17. Lezaud, P.: Chernoff-type bound for finite Markov chains. *Ann. Appl. Probab.* **8**(3), 849–867 (1998). ISSN: 1050-5164
18. Li, X., Chen, C.: Inequalities for the gamma function. In: *JIPAM. J. Inequal. Pure Appl. Math.* **8.1**, Article 28, 3 (2007) ISSN: 1443-5756
19. Merlevède, F., Peligrad, M., Rio, E.: A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probab. Theory Relat. Fields* **151**(3–4), 435–474 (2011). ISSN: 0178-8051
20. Merlevède, F., Peligrad, M., Rio, E.: High dimensional probability V: the Luminy volume. In: Christian Houdré et al. (Ed.) Vol. 5. Institute of Mathematical Statistics (IMS) Collections. Including papers from the 5th International Conference (HDP V) held in Luminy, May 26, 30, 2008. Institute of Mathematical Statistics, Beachwood, OH, pp. viii+356 (2009) ISBN: 978-0-940600-78-2
21. Merlevède, F., Rio, E.: Strong approximation for additive functionals of geometrically ergodic Markov chains. *Electron. J. Probab.* **20**, 27 (2015)
22. Meyn, S.P., Tweedie, R.L.: Markov chains and stochastic stability. *Communications and Control Engineering Series*. Springer-Verlag London, Ltd., London, pp. xvi+ 548 (1993). ISBN: 3-540-19832-6
23. Nummelin, E.: General Irreducible Markov Chains and Nonnegative Operators. *Cambridge Tracts in Mathematics*, vol. 83. Cambridge University Press, Cambridge (1984)
24. Paulin, D.: Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electron. J. Probab.* **20**(79), 32 (2015)
25. Samson, P.-M.: Concentration of measure inequalities for Markov chains and Φ -mixing processes. *Ann. Probab.* **28**(1), 416–461 (2000)
26. van der Vaart, A.W., Wellner, J.A.: Weak convergence and empirical processes. *Springer Series in Statistics. With Applications to Statistics*. Springer-Verlag, New York, pp. xvi+508 (1996). ISBN: 0-387-94640-3
27. Wintenberger, O.: Deviation inequalities for sums of weakly dependent time series. *Electron. Commun. Probab.* **15**, 489–503 (2010)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.