



## New Bernstein and Hoeffding type inequalities for regenerative Markov chains

Patrice Bertail and Gabriela Ciołek

Modal'X, UPL, Univ Paris Nanterre  
200 Avenue de la République,  
Nanterre, France.

*E-mail address:* [patrice.bertail@gmail.com](mailto:patrice.bertail@gmail.com)

LTCl, Télécom ParisTech, Université Paris-Saclay  
46 Rue Barrault,  
75013 Paris, France.

Faculty of Physics and Applied Computer Science  
al. Mickiewicza 30,  
30-059 Krakow, Poland.

*E-mail address:* [gabrielaciolek@gmail.com](mailto:gabrielaciolek@gmail.com)

**Abstract.** The purpose of this paper is to present Bernstein and Hoeffding type inequalities for regenerative Markov chains. Furthermore, we generalize these results and establish exponential bounds for suprema of empirical processes over a class of functions  $\mathcal{F}$  which size is controlled by its uniform entropy number. All constants involved in the bounds of the considered inequalities are given in an explicit form which can be advantageous for practical considerations. We present the theory for regenerative Markov chains, however the inequalities are also valid in the Harris recurrent case.

### 1. Introduction

Exponential inequalities are a powerful tool to control the tail probability that a random variable  $X$  exceeds some prescribed value  $t$ . They have been extensively investigated by many researchers due to the fact that they are a crucial step in deriving many results in numerous fields such as statistics, learning theory, discrete

---

*Received by the editors April 23th, 2018; accepted November 20th, 2018.*

*2010 Mathematics Subject Classification.* 62G09, 62G20, 60J10.

*Key words and phrases.* uniform entropy, exponential inequalities, empirical processes indexed by classes of functions, regenerative Markov chain.

This research was supported by a public grant as part of the Investissement d'avenir, project reference ANR-11-LABX-0056-LMH. The work was also supported by the Polish National Science Centre NCN ( grant No. UMO2016/23/N/ST1/01355 ) and (partly) by the Ministry of Science and Higher Education. This research has also been conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01).

mathematics, statistical mechanics, information theory or convex geometry. There is a vast literature that provides a comprehensive overview of the theory of exponential inequalities in the i.i.d. setting. An interested reader is referred to [Lin and Bai \(2010\)](#), [Boucheron et al. \(2013\)](#) or [van der Vaart and Wellner \(1996\)](#).

The wealth of possible applications of exponential inequalities has naturally led to development of this theory in the dependent setting. In this paper we are particularly interested in results that establish exponential bounds for the tail probabilities of the additive functional of the regenerative Markov chain of the form

$$f(X_1) + \cdots + f(X_n),$$

where  $(X_n)_{n \in \mathbb{N}}$  is a regenerative Markov chain. It is noteworthy that when deriving exponential inequalities for Markov chains (or any other process with some dependence structure) one can not expect to recover fully the classical results from the i.i.d. case. The goal is then to get some counterparts of the inequalities for i.i.d. random variables with some extra terms that appear in the bound as a consequence of a Markovian structure of the considered process.

In the recent years such (non-)asymptotic results have been obtained for Markov chains via many approaches: martingale arguments (see [Glynn and Ormoneit, 2002](#), where Hoeffding's inequality for uniformly ergodic Markov chains has been presented), coupling techniques (see [Chazottes and Redig, 2009](#) and [Dedecker and Gouëzel, 2015](#)). In fact, [Dedecker and Gouëzel \(2015\)](#) have proved that Hoeffding's inequality holds when the Markov chain is geometrically ergodic and thus weakened the assumptions imposed on the Markov chain in [Glynn and Ormoneit \(2002\)](#). [Wintenberger \(2017\)](#) has generalized the result of [Dedecker and Gouëzel \(2015\)](#) by showing that Hoeffding's inequality is valid also for unbounded functions of geometrically ergodic Markov chains provided that the sum is correctly self-normalized. [Paulin \(2015\)](#) has presented McDiarmid inequality for Markov chains using Merton coupling and spectral methods. [Cléménçon \(2001\)](#), [Adamczak \(2008\)](#), [Bertail and Cléménçon \(2009\)](#), and [Adamczak and Bednorz \(2015\)](#) have obtained exponential inequalities for ergodic Markov chains via regeneration techniques (see [Smith, 1955](#)).

Regeneration techniques for Markov chains are particularly appealing to us mainly due to the fact that it requires much fewer restrictions on the ergodicity properties of the chain in comparison to alternative methods. In this paper we establish Hoeffding and Bernstein type of inequalities for statistics of the form  $\frac{1}{n} \sum_{i=1}^n f(X_i)$ , where  $(X_n)_{n \in \mathbb{N}}$  is a regenerative Markov chain. We show that under proper control of the size of class of functions  $\mathcal{F}$  (measured by its uniform entropy number), one can get non-asymptotic bounds on the suprema over the class of  $\mathcal{F}$  of such empirical process for regenerative Markov chains. It is noteworthy that it is easy to generalize such results from regenerative case to the Harris recurrent one, using Nummelin extension of the initial chain (see [Nummelin, 1978](#)).

The paper is organized as follows. In Chapter 2 we introduce the notation and preliminary assumptions for Markov chains. We also recall some classical results from the i.i.d. setting which we generalize to the Markovian case. In Chapter 3 we present main results - Bernstein and Hoeffding type inequalities for regenerative Markov chains. The main ingredient to provide a crude exponential bound (with bad constants) is based on Montgomery-Smith's inequality which allows to reduce the problem on a random number of blocks to a fixed number of independent blocks. We then proposed a refined inequality by first controlling the number of blocks in the

inequality and then applying again Montgomery-Smith’s inequality on a remainder term. Next, we generalize these results and obtain Hoeffding and Bernstein type of bounds for suprema of empirical processes over a class of functions  $\mathcal{F}$ . We also present the inequalities when the chain is Harris recurrent. Some technical parts of the proofs are postponed to the Appendix.

**2. Preliminaries**

We begin by introducing some notation and recall the key concepts of Markov chains theory (see [Meyn and Tweedie, 1993](#) for a detailed review and references). Let  $X = (X_n)_{n \in \mathbb{N}}$  be a positive recurrent,  $\psi$ -irreducible Markov chain on a countably generated state space  $(E, \mathcal{E})$  with transition probability  $\Pi$  and initial probability  $\nu$ . We assume further, that  $X$  is regenerative (see [Smith, 1955](#)), i.e. there exists a measurable set  $A$ , called an atom, such that  $\psi(A) > 0$  and for all  $(x, y) \in A^2$  we have  $\Pi(x, \cdot) = \Pi(y, \cdot)$ . In what follows,  $\mathbb{P}_\nu$  (resp.  $\mathbb{P}_A$ ) denotes the probability measure on the underlying space such that  $X_0 \sim \nu$  (resp. conditioned on  $X_0 \in A$ ), and  $\mathbb{E}_\nu(\cdot)$  is the  $\mathbb{P}_\nu$ -expectation (resp.  $\mathbb{E}_A(\cdot)$  is the  $\mathbb{P}_A$ -expectation). We define the sequence of regeneration times  $(\tau_A(j))_{j \geq 1}$  which is the sequence of successive points of time when the chain visits  $A$  and forgets its past. Throughout the paper we write  $\tau_A = \tau_A(1)$ . It is well-known that we can cut the sample path of the process into data segments of the form

$$\mathcal{B}_j = (X_{1+\tau_A(j)}, \dots, X_{\tau_A(j+1)}), \quad j \geq 1$$

according to consecutive visits of the chain to the regeneration set  $A$ . By the strong Markov property the blocks are i.i.d. random variables taking values in the torus  $\cup_{k=1}^\infty E^k$ .

In the following, we assume that the mean inter-renewal time  $\alpha = \mathbb{E}_A[\tau_A] < \infty$  and point out that in this case, the stationary distribution is a Pitman occupation measure given by

$$\forall B \in \mathcal{E}, \quad \mu(B) = \frac{1}{\mathbb{E}_A[\tau_A]} \mathbb{E}_A \left[ \sum_{i=1}^{\tau_A} \mathbb{I}_{\{X_i \in B\}} \right],$$

where  $\mathbb{I}_B$  is the indicator function of an event  $B$ . Assume that we observe  $(X_1, \dots, X_n)$ . We introduce few more pieces of notation: throughout the paper we write  $l_n = \sum_{i=1}^n \mathbb{I}\{X_i \in A\}$  for the total number of consecutive visits of the chain to the atom  $A$ , thus we observe  $l_n + 1$  data blocks. We make the convention that  $B_{l_n}^{(n)} = \emptyset$  when  $\tau_A(l_n) = n$ . Furthermore, we denote by

$$l(B_j) = \tau_A(j + 1) - \tau_A(j), \quad j \geq 1,$$

the length of regeneration blocks. Let  $f : E \rightarrow \mathbb{R}$  be  $\mu$ -integrable function. In the following, we assume without loss of generality that

$$\mu(f) = \mathbb{E}_\mu[f(X_1)] = 0.$$

We introduce the following notation for partial sums of the regeneration cycles  $f(B_i) = \sum_{i=1+\tau_A(j)}^{\tau_A(j+1)} f(X_i)$ . Then, the regenerative approach is based on the following decomposition of the sum  $\sum_{i=1}^n f(X_i)$  :

$$\sum_{i=1}^n f(X_i) = \sum_{i=1}^{l_n} f(B_i) + \Delta_n,$$

where

$$\Delta_n = \sum_{i=1}^{\tau_A} f(X_i) + \sum_{i=\tau_A(l_n-1)}^n f(X_i).$$

We denote by

$$\sigma^2(f) = \frac{1}{\mathbb{E}_A(\tau_A)} \mathbb{E}_A \left( \sum_{i=1}^{\tau_A} \{f(X_i) - \mu(f)\}^2 \right)$$

the asymptotic variance.

For the completeness of the exposition, we recall now well-known classical results concerning some exponential inequalities for independent random variables. Firstly, we present the inequality for the i.i.d. bounded random variables due to [Hoeffding \(1963\)](#).

**Theorem 2.1** (Hoeffding’s inequality). *Let  $X_1, X_2, \dots, X_n$  be independent identically distributed random variables with common expectation  $\mathbb{E}X_1$  and such that  $a_i \leq X_i \leq b_i$  ( $i = 1, \dots, n$ ), then for  $t > 0$*

$$\mathbb{P} \left( \frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}X_1 \geq t \right) \leq \exp \left( - \frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2} \right).$$

Below we recall the generalization of Hoeffding’s inequality to unbounded functions. Interested reader, can find different variations of the following inequality (depending on imposed conditions on the random variables) in [Boucheron et al. \(2013\)](#).

**Theorem 2.2** (Bernstein’s inequality). *Let  $X_1, \dots, X_n$  be independent random variables with expectation  $\mathbb{E}X_l$  for  $X_l$ ,  $l \geq 1$  respectively, such that, for all integers  $p \geq 2$ ,*

$$\mathbb{E}|X_l|^p \leq p! R^{p-2} \sigma_l^2 / 2 \text{ for all } l \in \{1, \dots, n\}.$$

Then, for all  $t > 0$ ,

$$\mathbb{P} \left( \left| \sum_{i=1}^n (X_i - \mathbb{E}X_i) \right| \geq t \right) \leq 2 \exp \left( - \frac{t^2}{2(\sigma^2 + Rt)} \right),$$

where  $\sigma^2 = \sum_{i=1}^n \sigma_i^2$ .

The purpose of this paper is to derive similar bounds for Markov chains using the nice regenerative structure of Markov chains.

### 3. Exponential inequalities for the tail probability for suprema of empirical processes for Markov chains

In the following, we denote  $\bar{f}(x) = f(x) - \mu(f)$ . Moreover, we write respectively  $\bar{f}(B_1) = \sum_{i=1}^{\tau_A} \bar{f}(X_i)$  and  $|\bar{f}|(B_1) = \sum_{i=1}^{\tau_A} |\bar{f}|(X_i)$ . We will work under following conditions.

- A1. (Bernstein’s block moment condition) There exists a positive constant  $M_1$  such that for any  $p \geq 2$  and for every  $f \in \mathcal{F}$

$$\mathbb{E}_A |\bar{f}(B_1)|^p \leq \frac{1}{2} p! \sigma^2(f) M_1^{p-2}. \tag{3.1}$$

- A2. (Non-regenerative block exponential moment assumption) There exists  $\lambda_0 > 0$  such that for every  $f \in \mathcal{F}$  we have

$$\mathbb{E}_\nu \left[ \exp \left[ \lambda_0 \left| \sum_{i=1}^{\tau_A} \bar{f}(X_i) \right| \right] \right] < \infty.$$

- A3. (Exponential block moment assumption) There exists  $\lambda_1 > 0$  such that for every  $f \in \mathcal{F}$  we have

$$\mathbb{E}_A [\exp [\lambda_1 |\bar{f}| (B_1)]] < \infty.$$

*Remark 3.1.* It is noteworthy to mention that assumption A1 implies the existence of some exponential moments of  $\bar{f}(B_1)$  :

$$\mathbb{E}_A \exp(\lambda \bar{f}(B_1)) \leq \exp \left( \frac{\lambda^2/2}{1 - M_1 |\lambda|} \right) \text{ for all } \lambda < \frac{1}{M_1}.$$

In this section, we formulate two Bernstein type inequalities for Markov chains, one is established via simple use of Montgomery-Smith's inequality (see [Montgomery-Smith, 1993](#) and [de la Peña and Giné, 1999](#)) which results in much larger constants (comparing to the i.i.d. setting) in the dominating parts of the bound. The second Bernstein's bound contains small constants in the main counterparts of the bound, however at a cost of having an extra term in the bound. Since Montgomery-Smith's inequality is a crucial tool in our considerations, we provide that result below.

**Theorem 3.2** (Montgomery-Smith's inequality). *If  $X_i$ ,  $i \in \mathbb{N}$  are independent and identically distributed random variables, then for  $1 \leq k \leq n < \infty$  and all  $t > 0$  we have*

$$\mathbb{P} \left( \max_{1 \leq k \leq n} \left\| \sum_{i=1}^k X_i \right\| > t \right) \leq 9 \mathbb{P} \left( \left\| \sum_{i=1}^n X_i \right\| > t/30 \right).$$

Montgomery-Smith's inequality is particularly useful when dealing with random number of blocks  $l_n$  when establishing tail bound for  $\sum_{i=1}^{l_n} f(B_i)$  after proper centering. Using [Theorem 3.2](#) we can smoothly switch to consideration of a fixed number of independent blocks.

Before we state the theorems, we will give a short discussion on already existing results for exponential inequalities for Markov chains.

**Remarks 3.3.** *Since there is plenty of results concerning exponential inequalities for Markov chains under many assumptions, it may be difficult to compare their strength (measured by assumptions imposed on the chain) and applicability. Thus, before we present the proofs of [Theorem 3.4](#) and [Theorem 3.8](#), we make a short comparison of our result to already existing inequalities for Markov chains. We also strongly recommend seeing an exhaustive overview on the recent results of that type in [Adameczak and Bednorz \(2015\)](#).*

- (1) *The bounds obtained in this paper are related to the Fuk and Nagaev sharp bound inequality obtained in [Bertail and Cléménçon \(2009\)](#). It is also based on the regeneration properties and decomposition of the chain. However, our techniques of proof differ and allow us to obtain a better rate in the main subgaussian part of the inequality under the hypotheses. The proofs of the inequalities are simplified and do not require the partitioning arguments which was used in [Bertail and Cléménçon \(2009\)](#).*

- (2) It is noteworthy that we do not impose condition of stationarity of the considered Markov chain as in [Dedecker and Gouëzel \(2015\)](#) and [Chazottes and Redig \(2009\)](#) or any restrictions on the starting point of the chain as in [Dedecker and Gouëzel \(2015\)](#). Moreover, [Adamczak and Bednorz \(2015\)](#) use the assumption of strong aperiodicity for Harris Markov chain. We state a remark that this condition can be relaxed and we can only assume that Harris Markov chain is aperiodic (see [Remark 3.19](#)).
- (3) Many results concerning exponential inequalities for Markov chains are established for bounded functions  $f$  (see for instance [Adamczak, 2008](#), [Cléménçon, 2001](#), [Dedecker and Gouëzel, 2015](#), [Paulin, 2015](#) and the references therein). Our inequalities work for unbounded functions satisfying Bernstein's block moment condition without uniform ergodicity conditions. Moreover, all terms involved in our inequalities are given by explicit formulas. Thus, the results can be directly used in practical considerations. Note also that all the constants are given in simple, easy to interpret form and they do not depend on other underlying parameters.
- (4) [Wintenberger \(2017\)](#) has established exponential inequalities in unbounded case extending the result of [Dedecker and Gouëzel \(2015\)](#) to the case when the chain can start from any  $x \in E$ . However, the constant involved in the bound of the [Theorem 2.1](#) (obtained for bounded and unbounded functions) is very large.
- (5) As mentioned in the paper of [Adamczak \(2008\)](#), there is many exponential inequalities that satisfy spectral gaps (see for instance [Gao et al., 2014](#), [Lezaud, 2001](#)). Spectral gap inequalities allow to recover the Bernstein type inequality at its full strength. We need to mention that the geometric ergodicity assumption does not ensure in the non-reversible case that considered Markov chains admit a spectral gap (see [Theorem 1.4](#) in [Kontoyiannis and Meyn, 2012](#)).

We formulate a Bernstein type inequality for Markov chains below.

**Theorem 3.4.** *Assume that  $X = (X_n)_{n \in \mathbb{N}}$  is a regenerative positive recurrent Markov chain. Then, under assumptions A1 – A3, we have*

$$\mathbb{P}_\nu \left[ \left| \sum_{i=1}^n f(X_i) - \mu(f) \right| \geq x \right] \leq 18 \exp \left[ -\frac{x^2}{2 \times 90^2 (n\sigma^2(f) + M_1 x/90)} \right] \\ + C_1 \exp \left[ -\frac{\lambda_0 x}{3} \right] + C_2 \exp \left[ -\frac{\lambda_1 x}{3} \right],$$

where

$$C_1 = \mathbb{E}_\nu \left[ \exp \left| \lambda_0 \sum_{i=1}^{\tau_A} \bar{f}(X_i) \right| \right], \\ C_2 = \mathbb{E}_A \left[ \exp[\lambda_1 |\bar{f}|] (B_1) \right].$$

*Remark 3.5.* Observe that we do not impose a moment condition on  $\mathbb{E}_A[\tau_A]^p < \infty$  for  $p \geq 2$ . At the first glance, this might be surprising since one usually assumes the existence of  $\mathbb{E}_A[\tau_A]^2 < \infty$  when proving central limit theorem for regenerative Markov chains. A simple analysis of the proof of the central limit theorem in

a Markovian case (see for instance [Meyn and Tweedie, 1993](#)) reveals that it is sufficient to require only  $\mathbb{E}_A[\tau_A] < \infty$  when we consider centered function  $\bar{f}$  instead of  $f$ .

*Proof:* Firstly, we consider the sum of random variables of the following form

$$Z_n(\bar{f}) = \sum_{i=1}^{l_n} \bar{f}(B_j). \tag{3.2}$$

Furthermore, we have that  $S_n(\bar{f}) = Z_n(\bar{f}) + \Delta_n(\bar{f})$ .

We recall, that  $l_n$  is random and correlated with blocks itself. In order to apply Bernstein’s inequality for i.i.d. random variables we apply the Montgomery-Smith’s inequality (see [Montgomery-Smith, 1993](#)). It follows easily that

$$\begin{aligned} \mathbb{P}_A \left[ \left| \sum_{i=1}^{l_n} \bar{f}(B_i) \right| \geq x/3 \right] &\leq \mathbb{P}_A \left[ \max_{1 \leq k \leq l_n} \left| \sum_{i=1}^k \bar{f}(B_i) \right| \geq x/3 \right] \\ &\leq 9\mathbb{P}_A \left[ \left| \sum_{i=1}^n \bar{f}(B_i) \right| \geq x/90 \right] \end{aligned} \tag{3.3}$$

and under Bernstein’s condition A1 we obtain

$$\mathbb{P}_A \left[ \left| \sum_{i=1}^n \bar{f}(B_i) \right| \geq x/90 \right] \leq 2 \exp \left[ -\frac{x^2}{2 \times 90^2 (M_1 x/90 + n\sigma^2(f))} \right]. \tag{3.4}$$

Next, we want to control the remainder term  $\Delta_n$ .

$$\Delta_n = \sum_{i=1}^{\tau_A} \bar{f}(X_i) + \sum_{i=\tau_A(l_n-1)}^n \bar{f}(X_i). \tag{3.5}$$

The control of  $\Delta_n$  is guaranteed by Markov’s inequality, i.e.

$$\mathbb{P}_\nu \left[ \left| \sum_{i=1}^{\tau_A} \bar{f}(X_i) \right| \geq \frac{x}{3} \right] \leq \mathbb{E}_\nu \left[ \exp \left| \lambda_0 \sum_{i=1}^{\tau_A} \bar{f}(X_i) \right| \right] \exp \left[ -\frac{\lambda_0 x}{3} \right].$$

We deal similarly with the last term of  $\Delta_n$ . We complement the data  $1 + \tau_A(l_n) + 1$  by observations up to the next regeneration time  $1 + \tau_A(l_n + 1)$  and obtain

$$\begin{aligned} \mathbb{P}_\nu \left[ \left| \sum_{i=1+\tau_A(l_n)+1}^n \bar{f}(X_i) \right| \geq \frac{x}{3} \right] &\leq \mathbb{P}_\nu \left[ \sum_{i=1+\tau_A(l_n)+1}^n |\bar{f}|(X_i) \geq \frac{x}{3} \right] \\ &\leq \mathbb{P}_\nu \left[ \sum_{i=1+\tau_A(l_n)+1}^{1+\tau_A(l_n+1)} |\bar{f}|(X_i) \geq \frac{x}{3} \right] \\ &\leq \mathbb{E}_A \left[ \exp[\lambda_1 |\bar{f}|(B_1)] \right] \exp \left[ -\frac{\lambda_1 x}{3} \right]. \end{aligned}$$

□

We note that although the Montgomery-Smith’s inequality allows to obtain easily Bernstein’s bound for Markov chains, the constants are rather large. Interestingly, under an additional assumption on  $\mathbb{E}_A[\tau_A]^p$  we can obtain the Bernstein type inequality for regenerative Markov chains with much smaller constants for the dominating counterparts of the bound.

A4. (Block length moment assumption) There exists a positive constant  $M_2$  such that for any  $p \geq 2$

$$\mathbb{E}_A[\tau_A]^p \leq p!M_2^{p-2}\mathbb{E}_A[\tau_A^2]$$

and

$$\mathbb{E}_\nu[\tau_A]^p \leq p!M_2^{p-2}\mathbb{E}_\nu[\tau_A^2].$$

Before we formulate Bernstein’s inequality for regenerative Markov chains we introduce a lemma which provides a bound for tail probability of  $\sqrt{n} \left(\frac{l_n}{n} - \frac{1}{\alpha}\right)$  which will be crucial for the proof of Bernstein’s bound but also may be of independent interest.

**Lemma 3.6.** *Suppose that condition A4 holds. Then*

$$\mathbb{P}_\nu \left( n^{1/2} \left( \frac{l_n}{n} - \frac{1}{\alpha} \right) \geq x \right)$$

is bounded by

$$\exp \left( \frac{-\frac{1}{2}(\alpha x \sqrt{n} - 2\alpha)^2}{(E_\nu \tau_A^2 + (\frac{n}{\alpha} + x \sqrt{n}) E_A \tau_A^2) + (\alpha x \sqrt{n} + E_\nu \tau_A) M_2 (E_\nu \tau_A^2 + (\frac{n}{\alpha} + x \sqrt{n}) E_A \tau_A^2)^{1/2}} \right).$$

Proof of Lemma 3.6 is postponed to the Appendix.

*Remark 3.7.* Note that when  $n \rightarrow \infty$ , the dominating part in the exponential term is of order

$$\begin{aligned} & \frac{1}{2} \frac{\alpha^2 x^2}{E_A \tau_A^2 / \alpha + \alpha^{1/2} x M_2 (E_A \tau_A^2)^{1/2}} + O(n^{-1/2}) \\ &= \frac{1}{2} \frac{\alpha^2 x^2}{E_A \tau_A^2 / \alpha (1 + \alpha x M_2 (E_A \tau_A^2 / \alpha)^{-1/2})} + O(n^{-1/2}) \\ &= \frac{1}{2} \frac{(\alpha x)^2 / (E_A \tau_A^2 / \alpha)}{(1 + \alpha x M_2 (E_A \tau_A^2 / \alpha)^{-1/2})} + O(n^{-1/2}), \end{aligned}$$

thus we have a Gaussian tail with the right variance for moderate  $x$  and an exponential tail for large  $x$  and, in consequence, the constants are asymptotically optimal.

Now we are ready to state an alternative Bernstein type inequality for regenerative Markov chains, where under additional condition on the length of the blocks we can obtain much better inequality in terms of constants.

**Theorem 3.8.** *Assume that  $X = (X_n)_{n \in \mathbb{N}}$  is a regenerative positive recurrent Markov chain. Then, under assumptions A1-A4 we have for any  $a > 0$ , for  $x > 0$  and  $N > 0$*

$$\begin{aligned} \mathbb{P}_\nu \left[ \left| \sum_{i=1}^n f(X_i) - \mu(f) \right| \geq x \right] &\leq 2 \exp \left[ \frac{-x^2}{2 \times 3^2 (1+a)^2 \left( \lfloor \frac{n}{\alpha} \rfloor \sigma^2(f) + \frac{M_1}{3} \frac{x}{1+a} \right)} \right] \\ &+ 18 \exp \left[ \frac{-a^2 x^2}{2 \times 90^2 (1+a)^2 \left( N \sqrt{n} \sigma^2(f) + \frac{M_1}{90} \frac{ax}{1+a} \right)} \right] \\ &+ \mathbb{P}_\nu \left( n^{1/2} \left[ \frac{l_n}{n} - \frac{1}{\alpha} \right] > N \right) + C_1 \exp \left[ -\frac{\lambda_0 x}{3} \right] + C_2 \exp \left[ -\frac{\lambda_1 x}{3} \right], \end{aligned} \tag{3.6}$$



where

$$C_1 = \mathbb{E}_\nu \left[ \exp \left| \lambda_0 \sum_{i=1}^{\tau_A} \bar{f}(X_i) \right| \right],$$

$$C_2 = \mathbb{E}_A [\exp[\lambda_1 |\bar{f}|(B_1)]] .$$

*Remark 3.9.* In the proof of Theorem 3.8 we are interested in bounding for some  $t > 0$  the probability

$$\mathbb{P}_A \left[ \left| \sum_{i=\min(\lfloor \frac{n}{\alpha} \rfloor, l_n)}^{\max(\lfloor \frac{n}{\alpha} \rfloor, l_n)} \bar{f}(B_i) \right| \geq t \right] .$$

We control this quantity by using truncation argument for the total number of regeneration times  $l_n$ , i.e. for some  $N > 0$  we cut  $\sqrt{n} \left[ \frac{l_n}{n} - \frac{1}{\alpha} \right] \leq N$ . The magnitude of  $N$  (we want it to be relatively small) is significant since it appears in the final bound for  $\mathbb{P}_\nu [|\sum_{i=1}^n f(X_i) - \mu(f)| \geq x]$ . Observe that if we choose  $N = \log(n)$ , then by Lemma 3.6 we can see that

$$\mathbb{P}_\nu \left( n^{1/2} \left[ \frac{l_n}{n} - \frac{1}{\alpha} \right] \geq \log(n) \right) = o \left( \frac{1}{n} \right)$$

and in that case the second term in (3.6) remains small uniformly in  $x$ .

*Proof:* We start by the obvious observation that

$$\begin{aligned} \mathbb{P}_\nu \left[ \left| \sum_{i=1}^n f(X_i) - \mu(f) \right| \geq x \right] &\leq \mathbb{P}_A \left[ \left| \sum_{i=1}^{l_n} \bar{f}(B_i) \right| \geq x/3 \right] \\ &+ \mathbb{P}_\nu \left[ \left| \sum_{i=1}^{\tau_A} \bar{f}(X_i) \right| \geq x/3 \right] + \mathbb{P}_A \left[ \left| \sum_{i=\tau_A(l_n-1)}^n \bar{f}(X_i) \right| \geq x/3 \right]. \end{aligned} \tag{3.7}$$

*Remark 3.10.* Instead of dividing  $x$  by 3 in (3.7), one can use a different splitting to improve a little bit the final constants.

The bounds for the first and last non-regenerative blocks can be handled the same way as in Theorem 3.4. Next, we observe that, for any  $a > 0$ , we have

$$\begin{aligned} &\mathbb{P}_A \left[ \left| \sum_{i=1}^{l_n} \bar{f}(B_i) \right| \geq x/3 \right] \\ &\leq \mathbb{P}_A \left[ \left| \sum_{i=1}^{\lfloor \frac{n}{\alpha} \rfloor} \bar{f}(B_i) \right| \geq \frac{x}{3(1+a)} \right] + \mathbb{P}_A \left[ \left| \sum_{l_{n_1}}^{l_{n_2}} \bar{f}(B_i) \right| \geq \frac{ax}{3(1+a)} \right], \end{aligned} \tag{3.8}$$

where  $l_{n_1} = \min(\lfloor \frac{n}{\alpha} \rfloor, l_n)$  and  $l_{n_2} = \max(\lfloor \frac{n}{\alpha} \rfloor, l_n)$ . We observe that  $\sum_{i=1}^{\lfloor \frac{n}{\alpha} \rfloor} \bar{f}(B_i)$  is a sum of independent, identically distributed and sub-exponential random variables. Thus, we can directly apply Bernstein's bound and obtain

$$\mathbb{P}_A \left[ \left| \sum_{i=1}^{\lfloor \frac{n}{\alpha} \rfloor} \bar{f}(B_i) \right| \geq \frac{x}{3(1+a)} \right] \leq 2 \exp \left[ \frac{-x^2}{2 \times 3^2(1+a)^2 \left( \lfloor \frac{n}{\alpha} \rfloor \sigma^2(f) + M_1 x/3(1+a) \right)} \right]. \tag{3.9}$$

The control of  $\sum_{l_{n_1}}^{l_{n_2}} \bar{f}(B_i)$  is slightly more challenging due to the fact that  $l_n$  is random and correlated with the blocks itself. In the following, we will make use of the Montgomery-Smith’s inequality. Notice however, that since we expect the number of terms in this sum to be at most of the order  $\sqrt{n}$ , this term will be much smaller than the leading term (3.9) and will be asymptotically negligible. We have

$$\begin{aligned} \mathbb{P}_A \left( \left| \sum_{l_{n_1}}^{l_{n_2}} \bar{f}(B_i) \right| \geq \frac{ax}{3(1+a)} \right) &\leq \mathbb{P}_A \left[ \left| \sum_{l_{n_1}}^{l_{n_2}} \bar{f}(B_i) \right| \geq \frac{ax}{3(1+a)}, \sqrt{n} \left[ \frac{l_n}{n} - \frac{1}{\alpha} \right] \leq N \right] \\ &\quad + \mathbb{P}_\nu \left( \sqrt{n} \left[ \frac{l_n}{n} - \frac{1}{\alpha} \right] > N \right) = I + II. \end{aligned} \tag{3.10}$$

Firstly, we will bound term  $I$  in (3.10) using Montgomery-Smith’s inequality and the fact that if  $\sqrt{n} \left[ \frac{l_n}{n} - \frac{1}{\alpha} \right] \leq N$ , then  $l_{n_1} - l_{n_2} \leq \sqrt{n}N$ .

$$\begin{aligned} &\mathbb{P}_A \left[ \left| \sum_{l_{n_1}}^{l_{n_2}} \bar{f}(B_i) \right| \geq \frac{ax}{3(1+a)}, \sqrt{n} \left[ \frac{l_n}{n} - \frac{1}{\alpha} \right] \leq N \right] \\ &\leq \mathbb{P}_A \left( \max_{1 \leq k \leq N\sqrt{n}} \left| \sum_{i=1}^k \bar{f}(B_i) \right| \geq \frac{ax}{3(1+a)} \right) \\ &\leq 9\mathbb{P}_A \left( \left| \sum_{i=1}^{N\sqrt{n}} \bar{f}(B_i) \right| \geq \frac{ax}{90(1+a)} \right) \\ &\leq 18 \exp \left[ \frac{-a^2 x^2}{2 \times 90^2 (1+a)^2 \left( N\sqrt{n}\sigma^2(f) + \frac{M_1}{90} \frac{ax}{1+a} \right)} \right]. \end{aligned}$$

Lemma 3.6 allows to control term  $II$ . □

3.1. *Maximal inequalities under uniform entropy.* In empirical processes theory for processes indexed by class of functions, it is important to assess the complexity of considered classes. The information about entropy of  $\mathcal{F}$  helps us to inspect how large our class is. Generally, control of entropy of certain classes may be crucial step when investigating asymptotic behaviour of empirical processes indexed by a class of functions. In our setting, we will measure the size of class of functions  $\mathcal{F}$  via covering numbers and uniform entropy number. The following definition is due to [van der Vaart and Wellner \(1996\)](#).

**Definition 3.11** (Covering and uniform entropy number). The *covering number*  $N_p(\epsilon, Q, \mathcal{F})$  is the minimal number of balls  $\{g : \|g - f\|_{L^p(Q)} < \epsilon\}$  of radius  $\epsilon$  needed to cover the set  $\mathcal{F}$ . The *entropy* (without bracketing) is the logarithm of the covering number. We define uniform entropy number as  $N_p(\epsilon, \mathcal{F}) = \sup_Q N_p(\epsilon, Q, \mathcal{F})$ , where the supremum is taken over all discrete probability measures  $Q$ .

In the following we state assumptions on the size of considered class of functions  $\mathcal{F}$ . Rather than considering the assumptions  $A2$  and  $A3$ , we impose the assumptions on the first and the last non-regenerative blocks for the envelope  $F$  of  $\mathcal{F}$ .

A2'. (Non-regenerative block exponential moment assumption) There exists  $\lambda_0 > 0$  such that

$$\mathbb{E}_\nu \left[ \exp \left[ 2\lambda_0 \left| \sum_{i=1}^{\tau_A} \bar{F}(X_i) \right| \right] \right] < \infty.$$

A3'. (Exponential block moment assumption) There exists  $\lambda_1 > 0$  such that

$$\mathbb{E}_A [\exp [2\lambda_1 |\bar{F}| (B_1)]] < \infty.$$

A5. (Uniform entropy number condition)  $N_1(\epsilon, \mathcal{F}) < \infty$ .

Before we formulate Bernstein concentration type inequality for unbounded classes of functions, we introduce one more piece of notation, let

$$\sigma_m^2 = \max_{f \in \mathcal{F}} \sigma^2(f) > \eta > 0.$$

**Theorem 3.12.** *Assume that  $X = (X_n)_{n \in \mathbb{N}}$  is a regenerative positive recurrent Markov chain. Then, under assumptions A1, A2', A3' and A5 and for any  $x > 0$ ,  $0 < \epsilon < x/2$  and for all  $n \geq 1$  we have*

$$\begin{aligned} & \mathbb{P}_\nu \left[ \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(X_i) - \mu(f) \right| \geq x \right] \\ & \leq N_1(\epsilon, \mathcal{F}) \left\{ 18 \exp \left[ -\frac{(x - 2\epsilon)^2 n}{2 \times 90^2 (\sigma_m^2 + M_1(x - 2\epsilon)/90)} \right] \right. \\ & \quad \left. + C_1 \exp \left[ -\frac{\lambda_0(x - 2\epsilon)n}{3} \right] + C_2 \exp \left[ -\frac{\lambda_1(x - 2\epsilon)n}{3} \right] \right\}, \end{aligned} \tag{3.11}$$

where

$$C_1 = \mathbb{E}_\nu \left[ \exp \left| 2\lambda_0 \sum_{i=1}^{\tau_A} F(X_i) \right| \right] \quad \text{and} \quad C_2 = \mathbb{E}_A [\exp[2\lambda_1 |F| (B_1)]]$$

and  $F$  is an envelope function for  $\mathcal{F}$ .

*Remark 3.13.* Notice that our bound depends on a notion of uniform entropy number over a certain class of probability measures. However, for some classes of functions, this uniformity holds naturally, see for instance [Zou et al. \(2009\)](#).

Observe that if  $\mathcal{F}$  belongs to a ball of a Hölder space  $\mathbb{C}^P(E')$  on a compact set  $E'$  of an Euclidean space endowed with the norm

$$\|f\|_{\mathbb{C}^P(E')} = \sup_{x \in E'} |f(x)| + \sup_{x_1 \in E', x_2 \in E'} \left( \frac{|f(x_1) - f(x_2)|}{d(x_1, x_2)^P} \right)$$

then we have

$$M = \sup_{x \in X} F(x) < \infty$$

as well as

$$L = \sup_{f, g \in \mathcal{F}, f \neq g} \sup_z \frac{|f(z) - g(z)|}{\|f - g\|_{\mathbb{C}^P(E')}} < \infty$$

so that we can directly control the empirical sum by the obvious inequality

$$\sup_{f, g \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(X_i) - g(X_i) \right| \leq L \|f - g\|_{\mathbb{C}^P(E')}.$$

We refer to [Cucker and Smale \(2002\)](#) and [Zou et al. \(2009\)](#) for more details. See also examples of such classes of functions used in statistical learning in the latter.

It follows that if we replace the notion of uniform covering number  $N_1(\epsilon, \mathcal{F})$  with respect to the norm  $\|\cdot\|_{L^1(Q)}$  by the covering numbers  $N_{\mathbb{C}^p}(\epsilon, \mathcal{F})$  with respect to  $\|\cdot\|_{\mathbb{C}^p(E')}$  (which does not depend on underlying probability provided that  $N_1(\epsilon, \mathcal{F})$  is replaced by  $N_{\mathbb{C}^p}(\frac{\epsilon}{L}, \mathcal{F})$  in the inequality).

*Proof of Theorem 3.12:* We choose functions  $g_1, g_2, \dots, g_M$ , where  $M = N_1(\epsilon, \mathcal{F})$  such that

$$\min_j \mathbb{Q} |f - \mu(f) - g_j + \mu(g_1)| \leq 2\epsilon \text{ for each } f \in \mathcal{F},$$

where  $Q$  is any discrete probability measure. We also assume that  $g_1, g_2, \dots, g_M$  belong to  $\mathcal{F}$  and satisfy conditions  $A1, A2', A3'$ . We write  $f^*$  for the  $g_j$ , where the minimum is achieved.

Next, by definition of uniform covering numbers we obtain

$$\begin{aligned} & \mathbb{P}_\nu \left[ \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n (f(X_i) - \mu(f)) \right| \geq x \right] \\ & \leq \mathbb{P}_\nu \left\{ \sup_{f \in \mathcal{F}} \left[ \left| \frac{1}{n} \sum_{i=1}^n |f(X_i) - \mu(f) - f^*(X_i) + \mu(f^*)| \right| \right. \right. \\ & \qquad \qquad \qquad \left. \left. + \left| \frac{1}{n} \sum_{i=1}^n |f^*(X_i) - \mu(f^*)| \right| \right] \geq x \right\} \\ & \leq \mathbb{P}_\nu \left[ \max_{j \in \{1, \dots, N_1(\epsilon, \mathcal{F})\}} \left| \frac{1}{n} \sum_{i=1}^n g_j(X_i) - \mu(g_j) \right| \geq x - 2\epsilon \right] \\ & \leq N_1(\epsilon, \mathcal{F}) \max_{j \in \{1, \dots, N_1(\epsilon, \mathcal{F})\}} \mathbb{P}_\nu \left\{ \frac{1}{n} \left| \sum_{i=1}^n g_j(X_i) - \mu(g_j) \right| \geq x - 2\epsilon \right\}. \end{aligned}$$

We set the notation that

$$\bar{g}_j = g_j - \mu(g_j).$$

In what follows, our reasoning is analogous as in the proof of Theorem 3.4. Instead of taking any  $f \in \mathcal{F}$ , we work with the functions  $\bar{g}_j \in \mathcal{F}$ . Thus, we consider now the processes

$$Z_n(\bar{g}_j) = \sum_{i=1}^{l_n} \bar{g}_j(B_i) \tag{3.12}$$

and

$$S_n(\bar{g}_j) = Z_n(\bar{g}_j) + \Delta_n(\bar{g}_j).$$

Under the assumptions  $A1, A2'$  and  $A3'$  for  $\bar{g}_j$ , we get the analogous bound to that from Theorem 3.4 for  $Z_n(\bar{g}_j)$ . Note that in order to control the first and the last non-regenerative blocks we use conditions  $A2'$  and  $A3'$ . Thus, consequently an envelope  $\bar{F}$  appears in the bounds. □

Below we will formulate a maximal version of Theorem 3.8.

**Theorem 3.14.** *Assume that  $X = (X_n)_{n \in \mathbb{N}}$  is a regenerative positive recurrent Markov chain. Then, under assumptions  $A1, A2', A3', A4 - A5$  and for any  $x > 0$ ,*

any  $0 < \epsilon < x/2$ , any  $N > 0$  and for all  $n \geq 1$  we have

$$\begin{aligned} & \mathbb{P}_\nu \left[ \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(X_i) - \mu(f) \right| \geq x \right] \\ & \leq N_1(\epsilon, \mathcal{F}) \left\{ 2 \exp \left[ \frac{-(x-2\epsilon)^2 n}{2 \times 3^2 (1+a)^2 \left( \frac{\sigma^2(f)}{\mathbb{E}_A[\tau_A]} + \frac{M_1}{3} \frac{x-2\epsilon}{1+a} \right)} \right] \right. \\ & \quad + 18 \exp \left[ \frac{-a^2(x-2\epsilon)^2 n^2}{2 \times 90^2 (1+a)^2 \left( N \sqrt{n} \sigma^2(f) + \frac{M_1}{90} \frac{a(x-2\epsilon)}{1+a} \right)} \right] \\ & \quad + \mathbb{P}_\nu \left( n^{1/2} \left[ \frac{l_n}{n} - \frac{1}{\mathbb{E}_A[\tau_A]} \right] > N \right) + C_1 \exp \left[ -\frac{\lambda_0(x-2\epsilon)n}{3} \right] \\ & \quad \left. + C_2 \exp \left[ -\frac{\lambda_1(x-2\epsilon)n}{3} \right] \right\}, \end{aligned}$$

where

$$C_1 = \mathbb{E}_\nu \left[ \exp \left| 2\lambda_0 \sum_{i=1}^{\tau_A} F(X_i) \right| \right] \quad \text{and} \quad C_2 = \mathbb{E}_A [\exp[2\lambda_1 |F|(B_1)]] .$$

*Proof:* The proof is a combination of the proofs of Theorem 3.8 and Theorem 3.12. We deal with the supremum over  $\mathcal{F}$  the same way as in Theorem 3.12. Indeed, for any  $0 < \epsilon < x/2$  we have

$$\begin{aligned} & \mathbb{P}_\nu \left[ \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n (f(X_i) - \mu(f)) \right| \geq x \right] \\ & \leq N_1(\epsilon, \mathcal{F}) \max_{j \in \{1, \dots, N_1(\epsilon, \mathcal{F})\}} \mathbb{P}_\nu \left\{ \left| \frac{1}{n} \sum_{i=1}^n g_j(X_i) - \mu(g_j) \right| \geq x - 2\epsilon \right\} \\ & = N_1(\epsilon, \mathcal{F}) \max_{j \in \{1, \dots, N_1(\epsilon, \mathcal{F})\}} \mathbb{P}_\nu \left[ \left| \sum_{i=1}^n \bar{g}_j(B_i) \right| \geq x \right] \end{aligned}$$

for the same collection of functions  $g_1, \dots, g_{N_1(\epsilon, \mathcal{F})}$  as in Theorem 3.12. Now we can do the following decomposition

$$\mathbb{P}_\nu \left[ \left| \sum_{i=1}^n \bar{g}_j(B_i) \right| \geq x \right] \leq \mathbb{P}_A \left[ \left| \sum_{i=1}^{l_n} \bar{g}_j(B_i) \right| \geq x/3 \right] + \mathbb{P}_\nu \left[ \left| \sum_{i=1}^{\tau_A} \bar{g}_j(X_i) \right| \geq x/3 \right] \quad (3.13)$$

$$+ \mathbb{P}_A \left[ \left| \sum_{i=\tau_A}^n \bar{g}_j(X_i) \right| \geq x/3 \right] \quad (3.14)$$

and control each term on the right hand side of the above inequality in an analogous way as in Theorem 3.8.  $\square$

We can obtain even sharper upper bound when class  $\mathcal{F}$  is uniformly bounded. In the following, we will show that it is possible to get a Hoeffding type inequality and have a stronger control of moments of the sum  $S_n(f)$  which is a natural consequence of uniform boundedness assumption imposed on  $\mathcal{F}$ .

A6. The class of functions  $\mathcal{F}$  is uniformly bounded, i.e. there exists a constant  $D$  such that  $\forall f \in \mathcal{F} |f| < D$ .

**Theorem 3.15.** *Assume that  $X = (X_n)_{n \in \mathbb{N}}$  is a regenerative positive recurrent Markov chain. Then, under assumptions A1, A2', A3', A5 – A6 and for any  $x > 0$ , any  $0 < \epsilon < x/2$  and for all  $n \geq 1$  we have*

$$\mathbb{P}_\nu \left[ \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^n f(X_i) - \mu(f) \right| \geq x \right] \leq N_1(\epsilon, \mathcal{F}) \left\{ 18 \exp \left[ -\frac{(x - 2\epsilon)^2 n}{2 \times 90^2 D^2} \right] + C_1 \exp \left[ -\frac{\lambda_0(x - 2\epsilon)n}{3} \right] + C_2 \exp \left[ -\frac{\lambda_1(x - 2\epsilon)n}{3} \right] \right\}, \tag{3.15}$$

where

$$C_1 = \mathbb{E}_\nu \exp |2\lambda_0 \tau_A D| \quad \text{and} \quad C_2 = \mathbb{E}_A \exp |2\lambda_1 l(B_1) D|.$$

*Proof:* The proof bears resemblance to the proof of Theorem 3.12, with a few natural modifications which are a consequence of the uniform boundedness of  $\mathcal{F}$ .  $\square$

Under additional condition A4 we can obtain easily the bound with smaller constants, we follow the analogous way as in Theorem 3.14.

**3.2. General Harris recurrent case.** It is noteworthy that Theorems 3.4, 3.12, 3.15 are also valid in a Harris recurrent case under slightly modified assumptions. It is well known that it is possible to retrieve all regeneration techniques also in a Harris recurrent case via the Nummelin splitting technique which allows to extend the probabilistic structure of any chain in order to artificially construct a regeneration set. The Nummelin splitting technique relies heavily on the notion of small set. For the clarity of exposition we recall the definition.

**Definition 3.16.** We say that a set  $S \in \mathcal{E}$  is small if there exists a parameter  $\delta > 0$ , a positive probability measure  $\Phi$  supported by  $S$  and an integer  $m \in \mathbb{N}^*$  such that

$$\forall x \in S, B \in \mathcal{E} \quad \Pi^m(x, B) \geq \delta \Phi(B), \tag{3.16}$$

where  $\Pi^m$  denotes the  $m$ -th iterate of the transition probability  $\Pi$ .

We expand the sample space in order to define a sequence  $(Y_n)_{n \in \mathbb{N}}$  of independent r.v.'s with parameter  $\delta$ . We define a joint distribution  $\mathbb{P}_{\nu, \mathcal{M}}$  of  $X^{\mathcal{M}} = (X_n, Y_n)_{n \in \mathbb{N}}$ . The construction relies on the mixture representation of  $\Pi$  on  $S$ , namely

$$\Pi(x, B) = \delta \Phi(B) + (1 - \delta) \frac{\Pi(x, B) - \delta \Phi(B)}{1 - \delta}.$$

It can be retrieved by the following randomization of the transition probability  $\Pi$  each time the chain  $X$  visits the set  $S$ . If  $X_n \in S$  and

- if  $Y_n = 1$  (which happens with probability  $\delta \in ]0, 1[$ ), then  $X_{n+1}$  is distributed according to the probability measure  $\Phi$ ,
- if  $Y_n = 0$  (that happens with probability  $1 - \delta$ ), then  $X_{n+1}$  is distributed according to the probability measure  $(1 - \delta)^{-1}(\Pi(X_n, \cdot) - \delta \Phi(\cdot))$ .

This bivariate Markov chain  $X^{\mathcal{M}}$  is called the *split chain*. It takes its values in  $E \times \{0, 1\}$  and possesses an atom, namely  $A = S \times \{1\}$ . The split chain  $X^{\mathcal{M}}$  inherits all the stability and communication properties of the chain  $X$ . The regenerative blocks of the split chain are i.i.d. (in case  $m = 1$  in (3.16)) (see Meyn and Tweedie, 1993 for further details).

We will formulate a Bernstein type inequality for unbounded classes of functions in the Harris recurrent case (equivalent of Theorem 3.4). Theorems 3.12 and 3.15 can be reformulated for Harris chains in similar way. We impose the following conditions:

AH1. (Bernstein’s block moment condition) There exists a positive constant  $M_1$  such that for any  $p \geq 2$  and for every  $f \in \mathcal{F}$

$$\sup_{y \in S} \mathbb{E}_y |\bar{f}(B_1)|^p \leq \frac{1}{2} p! \sigma^2(f) M_1^{p-2}. \tag{3.17}$$

AH2. (Non-regenerative block exponential moment assumption) There exists a constant  $\lambda_0 > 0$  such that for every  $f \in \mathcal{F}$  we have

$$\mathbb{E}_\nu \left[ \exp \left| \lambda_0 \sum_{i=1}^{\tau_S} \bar{f}(X_i) \right| \right] < \infty.$$

AH3. (Exponential block moment assumption) There exists a constant  $\lambda_1 > 0$  such that for every  $f \in \mathcal{F}$  we have

$$\sup_{y \in S} \mathbb{E}_y [\exp[\lambda_1 |\bar{f}|(B_1)]] < \infty.$$

Let  $\sup_{y \in S} \mathbb{E}_y [\tau_S] = \alpha_{\mathcal{M}} < \infty$ . We are ready to formulate a Bernstein type inequality for Harris recurrent Markov chains.

**Theorem 3.17.** *Assume that  $X_{\mathcal{M}}$  is a Harris recurrent, strongly aperiodic Markov chain. Then, under assumptions AH1-AH3, we have*

$$\begin{aligned} \mathbb{P}_\nu \left[ \left| \sum_{i=1}^n f(X_i) - \mu(f) \right| \geq x \right] &\leq 18 \exp \left[ -\frac{x^2}{2 \times 90^2 (n\sigma^2(f) + M_1 x/90)} \right] \\ &+ C_1 \exp \left[ -\frac{\lambda_0 x}{3} \right] + C_2 \exp \left[ -\frac{\lambda_1 x}{3} \right], \end{aligned} \tag{3.18}$$

where

$$\begin{aligned} C_1 &= \mathbb{E}_\nu \left[ \exp \left| \lambda_0 \sum_{i=1}^{\tau_S} \bar{f}(X_i) \right| \right], \\ C_2 &= \sup_{y \in S} \mathbb{E}_y [\exp[\lambda_1 |\bar{f}|(B_1)]]. \end{aligned}$$

The proof of Theorem 3.17 is analogous to the proof of Theorem 3.4. We can obtain a bound with much smaller constants under an extra block moment condition.

AH4. (Block length moment assumption) There exists a positive constant  $M_2$  such that for any  $p \geq 2$

$$\begin{aligned} \sup_{y \in S} \mathbb{E}_y [\tau_S]^p &\leq p! M_2^{p-2} E_A \tau_A^2, \\ E_\nu [\tau_S]^p &\leq p! M_2^{p-2} E_\nu \tau_S^2. \end{aligned}$$

**Theorem 3.18.** *Assume that  $X_{\mathcal{M}}$  is a Harris recurrent, strongly aperiodic Markov chain. Then, under assumptions AH1-AH4, we have for any  $x > 0$  and for  $N \in \mathbb{R}$*

$$\begin{aligned} \mathbb{P}_{\nu} \left[ \left| \sum_{i=1}^n f(X_i) - \mu(f) \right| \geq x \right] &\leq 2 \exp \left[ \frac{-x^2}{2 \times 3^2(1+a)^2 \left( \lfloor \frac{n}{\alpha} \rfloor \sigma^2(f) + \frac{M_1}{3} \frac{x}{1+a} \right)} \right] \\ &+ 18 \exp \left[ \frac{-a^2 x^2}{2 \times 90^2(1+a)^2 \left( N\sqrt{n}\sigma^2(f) + \frac{M_1}{90} \frac{ax}{1+a} \right)} \right] \\ &+ \mathbb{P}_{\nu} \left[ n^{1/2} \left( \frac{l_n}{n} - \frac{1}{\alpha} \right) > N \right] \\ &+ C_1 \exp \left[ -\frac{\lambda_0 x}{3} \right] + C_2 \exp \left[ -\frac{\lambda_1 x}{3} \right], \end{aligned}$$

where

$$\begin{aligned} C_1 &= \mathbb{E}_{\nu} \left[ \exp \left| \lambda_0 \sum_{i=1}^{\tau_S} \bar{f}(X_i) \right| \right], \\ C_2 &= \sup_{y \in S} \mathbb{E}_y \left[ \exp[\lambda_1 |\bar{f}|(B_1)] \right]. \end{aligned}$$

*Remark 3.19.* In the Theorem 3.17 we assumed that  $X_{\mathcal{M}}$  is strongly aperiodic. It is easy, however, to relax this assumption and impose only the aperiodicity condition on Harris chain by using the same trick as in Levental (1988). Note that if  $X_{\mathcal{M}}$  satisfies  $\mathcal{M}(m, S, \delta, \Phi)$  for  $m > 1$ , then the blocks of data are 1-dependent. Denote by  $\bar{S} = S \cup \{*\}$ , where  $\{*\}$  is an ideal point which is not in  $S$ . Next, we define a pseudo-atom  $\alpha_{\bar{\mathcal{M}}} = \bar{S} \times \{1\}$ . In order to impose only aperiodicity in this case it is sufficient to consider two processes  $\{E_i\}$  and  $\{O_i\}$  such that  $O_i = f(X_i)$  if  $\tau_{\alpha_{\bar{\mathcal{M}}}}(2k+1) < i \leq \tau_{\alpha_{\bar{\mathcal{M}}}}(2k+2)$  and  $O_i = *$  otherwise  $E_i = f(X_i)$  if  $\tau_{\alpha_{\bar{\mathcal{M}}}}(2k) < i \leq \tau_{\alpha_{\bar{\mathcal{M}}}}(2k+1)$ , for some  $k \geq 0$  and  $E_i = *$ . Every function  $f : S \rightarrow \mathbb{R}$  will be considered as defined on  $\bar{S}$  with identification  $f(*) = 0$  (see also Levental, 1988 for more details concerning those two processes). Then, we prove Bernstein type of inequality similarly as we prove Theorems 3.4 and 3.17 applying all the reasoning to  $\{E_i\}$  and  $\{O_i\}$  separately, yielding to a similar inequality up to an additional multiplicative constant 2.

**Acknowledgements**

The authors would like to thank a referee and the associate editor for their insightful comments, corrections and references.

**Appendix**

*Proof of Lemma 3.6:* Let  $\tau_k$  be the time of the  $k$ -th visit to the atom A ( $S \times \{1\}$  in the general case).



In the following we make use of the argument from [Dedecker and Gouëzel \(2015\)](#) and observe that we have for any  $k \leq n$

$$\begin{aligned} \mathbb{P}_\nu(l_n \geq k) &= \mathbb{P}_\nu(\tau_k \leq n) = \mathbb{P}_\nu\left(\sum_{i=1}^k \Delta\tau_i \leq n\right) \\ &= \mathbb{P}\left((\Delta\tau_1 - E_\nu\tau_A) + \sum_{i=2}^k (\Delta\tau_i - \alpha) \leq n - (k-1)\alpha - E_\nu\tau_A\right). \end{aligned}$$

It follows that if  $x > \sqrt{n}(1 - \alpha^{-1})$  ( $\alpha = E_A\tau_A$ ), then

$$\mathbb{P}_\nu\left(n^{1/2}\left(\frac{l_n}{n} - \frac{1}{\alpha}\right) \geq x\right) = 0$$

and if  $0 < x \leq \sqrt{n}(1 - \alpha^{-1})$ , then

$$\begin{aligned} \mathbb{P}_\nu\left(n^{1/2}\left(\frac{l_n}{n} - \frac{1}{\alpha}\right) \geq x\right) &= \mathbb{P}_\nu\left(l_n \geq \frac{n}{\alpha} + x\sqrt{n}\right) \\ &\leq \mathbb{P}_\nu\left(l_n \geq \left[\frac{n}{\alpha} + x\sqrt{n}\right]\right) \\ &\leq \mathbb{P}\left((\Delta\tau_1 - E_\nu\tau_A) + \sum_{i=2}^{\lceil \frac{n}{\alpha} + x\sqrt{n} \rceil} (\Delta\tau_i - \alpha) \leq n - (\lceil \frac{n}{\alpha} + x\sqrt{n} \rceil - 1)\alpha - E_\nu\tau_A\right), \end{aligned}$$

where  $\lceil \cdot \rceil$  is the integer part.

Since  $\frac{n}{\alpha} + x\sqrt{n} - 1 \leq \lceil \frac{n}{\alpha} + x\sqrt{n} \rceil \leq \frac{n}{\alpha} + x\sqrt{n}$ , we get

$$\begin{aligned} n - (\lceil \frac{n}{\alpha} + x\sqrt{n} \rceil - 1)\alpha - E_\nu\tau_A &\leq n - (\frac{n}{\alpha} + x\sqrt{n} - 2)\alpha - E_\nu\tau_A \\ &= -\alpha x\sqrt{n} + 2\alpha - E_\nu\tau_A. \end{aligned}$$

It follows that

$$\begin{aligned} \mathbb{P}_\nu\left(n^{1/2}\left(\frac{l_n}{n} - \frac{1}{\alpha}\right) \geq x\right) \\ \leq \mathbb{P}\left((\Delta\tau_1 - E_\nu\tau_A) + \sum_{i=2}^{\lceil \frac{n}{\alpha} + x\sqrt{n} \rceil} (\Delta\tau_i - \alpha) \leq -\alpha x\sqrt{n} + 2\alpha - E_\nu\tau_A\right), \end{aligned}$$

where  $\lceil \cdot \rceil$  is the integer part.

Now, we can apply any Bennett’s or Bernstein’s inequality on these centered i.i.d. random variables to get an exponential bound. This can be done since we assumed A4. Note that other bounds (polynomial for instance) can be obtained under appropriate modifications of A4. In our case we get

$$\begin{aligned} \mathbb{P}\left((\Delta\tau_1 - E_\nu\tau_A) + \sum_{i=2}^{\lceil \frac{n}{\alpha} + x\sqrt{n} \rceil} (\Delta\tau_i - \alpha) \leq -\alpha x\sqrt{n} + 2\alpha - E_\nu\tau_A\right) \\ \leq \exp\left(-\frac{\frac{1}{2}(\alpha x\sqrt{n} - 2\alpha + E_\nu\tau_A^2)/S_n^2}{1 + (\alpha x\sqrt{n} - 2\alpha + E_\nu\tau_A)M_2/S_n}\right), \end{aligned}$$

where

$$S_n^2 = E_\nu\tau_A^2 + (\lceil \frac{n}{\alpha} + x\sqrt{n} \rceil - 1)E_A\tau_A^2.$$

The above bound can be reduced to

$$\exp\left(\frac{-\frac{1}{2}(\alpha x\sqrt{n} - 2\alpha)^2}{(E_\nu\tau_A^2 + (\frac{n}{\alpha} + x\sqrt{n})E_A\tau_A^2) + (\alpha x\sqrt{n} + E_\nu\tau_A)M_2(E_\nu\tau_A^2 + (\frac{n}{\alpha} + x\sqrt{n})E_A\tau_A^2)^{1/2}}\right).$$

□

## References

- R. Adamczak. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.* **13**, no. 34, 1000–1034 (2008). [MR2424985](#).
- R. Adamczak and W. Bednorz. Exponential concentration inequalities for additive functionals of Markov chains. *ESAIM Probab. Stat.* **19**, 440–481 (2015). [MR3423302](#).
- P. Bertail and S. Cléménçon. Sharp bounds for the tails of functionals of Markov chains. *Teor. Veroyatn. Primen.* **54** (3), 609–619 (2009). [MR2766354](#).
- S. Boucheron, G. Lugosi and P. Massart. *Concentration inequalities*. Oxford University Press, Oxford (2013). ISBN 978-0-19-953525-5. [MR3185193](#).
- J.-R. Chazottes and F. Redig. Concentration inequalities for Markov processes via coupling. *Electron. J. Probab.* **14**, no. 40, 1162–1180 (2009). [MR2511280](#).
- S. J. M. Cléménçon. Moment and probability inequalities for sums of bounded additive functionals of regular Markov chains via the Nummelin splitting technique. *Statist. Probab. Lett.* **55** (3), 227–238 (2001). [MR1867526](#).
- F. Cucker and S. Smale. On the mathematical foundations of learning. *Bull. Amer. Math. Soc. (N.S.)* **39** (1), 1–49 (2002). [MR1864085](#).
- J. Dedecker and S. Gouëzel. Subgaussian concentration inequalities for geometrically ergodic Markov chains. *Electron. Commun. Probab.* **20**, no. 64, 12 (2015). [MR3407208](#).
- F. Gao, A. Guillin and L. Wu. Bernstein-type concentration inequalities for symmetric Markov processes. *Theory Probab. Appl.* **58** (3), 358–382 (2014). [MR3403002](#).
- P. W. Glynn and D. Ormoneit. Hoeffding’s inequality for uniformly ergodic Markov chains. *Statist. Probab. Lett.* **56** (2), 143–146 (2002). [MR1881167](#).
- W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13–30 (1963). [MR0144363](#).
- I. Kontoyiannis and S. P. Meyn. Geometric ergodicity and the spectral gap of non-reversible Markov chains. *Probab. Theory Related Fields* **154** (1-2), 327–339 (2012). [MR2981426](#).
- S. Levental. Uniform limit theorems for Harris recurrent Markov chains. *Probab. Theory Related Fields* **80** (1), 101–118 (1988). [MR970473](#).
- P. Lezaud. Chernoff and Berry-Esséen inequalities for Markov processes. *ESAIM Probab. Statist.* **5**, 183–201 (2001). [MR1875670](#).
- Z. Lin and Z. Bai. *Probability inequalities*. Science Press Beijing, Beijing; Springer, Heidelberg (2010). ISBN 978-7-03-025562-4; 978-3-642-05260-6. [MR2789096](#).
- S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Communications and Control Engineering Series. Springer-Verlag London, Ltd., London (1993). ISBN 3-540-19832-6. [MR1287609](#).

- S. J. Montgomery-Smith. Comparison of sums of independent identically distributed random vectors. *Probab. Math. Statist.* **14** (2), 281–285 (1994) (1993). [MR1321767](#).
- E. Nummelin. A splitting technique for Harris recurrent Markov chains. *Z. Wahrsch. Verw. Gebiete* **43** (4), 309–318 (1978). [MR0501353](#).
- D. Paulin. Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electron. J. Probab.* **20**, no. 79, 32 (2015). [MR3383563](#).
- V. H. de la Peña and E. Giné. *Decoupling*. Probability and its Applications (New York). Springer-Verlag, New York (1999). ISBN 0-387-98616-2. [MR1666908](#).
- W. L. Smith. Regenerative stochastic processes. *Proc. Roy. Soc. London. Ser. A.* **232**, 6–31 (1955). [MR0073877](#).
- A. W. van der Vaart and J. A. Wellner. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York (1996). ISBN 0-387-94640-3. [MR1385671](#).
- O. Wintenberger. Exponential inequalities for unbounded functions of geometrically ergodic Markov chains: applications to quantitative error bounds for regenerative Metropolis algorithms. *Statistics* **51** (1), 222–234 (2017). [MR3600471](#).
- B. Zou, H. Zhang and Z. Xu. Learning from uniformly ergodic Markov chains. *J. Complexity* **25** (2), 188–200 (2009). [MR2513615](#).