



Deviation inequalities for sums of weakly dependent time series

Olivier Wintenberger

► To cite this version:

Olivier Wintenberger. Deviation inequalities for sums of weakly dependent time series. Electronic Communications in Probability, 2010, 15, pp.489-503. hal-00430608

HAL Id: hal-00430608

<https://hal.science/hal-00430608>

Submitted on 9 Nov 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DEVIATION INEQUALITIES FOR SUMS OF WEAKLY DEPENDENT TIME SERIES

OLIVIER WINTENBERGER

ABSTRACT. In this paper we give new deviation inequalities of Bernstein's type for the partial sums of weakly dependent time series. The loss from the independent case is studied carefully. We give non mixing examples such that dynamical systems and Bernoulli shifts for whom our deviation inequalities hold. The proofs are based on the blocks technique and different coupling arguments.

AMS 2000 subject classifications: Primary 60F99; Secondary 60G10

Keywords and phrases: Exponential inequality, weak dependence, coupling scheme, dynamical systems, Bernoulli shifts.

1. INTRODUCTION

The aim of this paper is to extend the deviation inequality of Bernstein's type from the independent case to some weakly dependent ones. We consider a sample $(X_1, \dots, X_n)$ of a stationary process (X_t) in a metric space $(\mathcal{X}, d)$. Considering the set $\mathcal{F}$ of 1-Lipschitz functions from $\mathcal{X}$ to $[-1/2, 1/2]$, we are interested by the deviation of the partial sum $S(f) = \sum_{i=1}^n f(X_i)$ for any $f \in \mathcal{F}$ assuming that $\mathbb{E}(f(X_i)) = 0$. If the X_i are independent and if $\sigma_k^2(f) = k^{-1} \text{Var}(\sum_{i=1}^k f(X_i))$, the classical Bernstein inequality gives the deviation estimate, see Bennett [3]:

$$(1.1) \quad \mathbb{P}\left(S(f) \geq \sqrt{2n\sigma_1^2(f)x} + x/6\right) \leq e^{-x} \quad \text{for all } x \geq 0.$$

This inequality reflects the gaussian approximation of the tail of $S(f)$ for small values x . And for large values of x , it reflects the exponential approximation of the tail of $S(f)$. This deviation inequality is very useful in statistics, see for example the monographs of Catoni [5] and of Massart [18].

To extend such deviation inequality to the dependent cases, a tradeoff between the sharpness of the estimates and the generality of the context has to be done. Estimates as sharp as in the independent cases (up to constants) are obtained for Markov chains in Lezaud [17], Joulin and Ollivier [16] under granularity. Bertail and Clemenon [4] obtain a deviation inequality for recurrent Markov chains. There exists $C > 0$ such that for all $M > 0$ and all $x > 0$:

$$\mathbb{P}(S(f) \geq C(\sqrt{n\sigma^2(f)_T x} + Mx)) \leq e^{-x} + n\mathbb{P}(T_1 \geq M),$$

where the T_i are the iid regeneration times and $\sigma_T^2(f) = \mathbb{E}(T)^{-1} \text{Var}(\sum_{i=1}^{T_1} f(X_i))$. Up to a constant, it is the limit variance in the CLT of $S(f)$, more natural than $\sigma_1^2(f)$ in (1.1). This primitive estimate of the tail is natural as, through the splitting technique of Nummelin [21], the partial sums $S(f)$ are sums of iid sequences of blocks of size T_i . If the regeneration times are bounded, then up to different constants the same estimate than in the iid case is obtained. If

the regeneration times admit finite exponential moments, fixing $M \approx \ln n$ Adamczak [1] obtains estimates of the deviations with a constant $C > 0$:

$$\mathbb{P}\left(S(f) \geq C\left(\sqrt{n\sigma_T^2(f)x} + \ln nx\right)\right) \leq e^{-x} \quad \text{for all } x \geq 0.$$

A loss of rate $\ln n$, that cannot be reduced, appears in the exponential approximation compared with the iid case, see Adamczak [1] for more details.

In all these works, the strong Markov property is crucial. To bypass the Markov assumption, one way is to use dependent coefficients. Ibragimov [14] introduced the uniformly ϕ -mixing coefficients. In this settings, Samson [24] achieves the deviation inequality (1.1) with different constants as soon as $\sum \sqrt{\phi_r} < \infty$. Less accurate results have been obtained for more general mixing coefficients than the ϕ -mixing ones: Viennet [25] for absolutely regular mixing and Merlevede *et al.* [19] for geometrically strongly mixing. Recently, mixing coefficients have been extended to weakly dependent ones, see Doukhan and Louhichi [11] and Dedecker and Prieur [8]. Under the exponential decrease of these coefficients, deviation inequalities for $S(f)$ with a loss in the exponential approximation are given in Doukhan and Neumann [12]. Merlevede *et al.* [20] extends these results for the partial sum $S(f)$ for unbounded functions f .

The dependence context of this paper is the one of the so-called φ -weakly dependent coefficients introduced by Rio in [22] to extend the uniformly ϕ -mixing coefficients. We provide new deviation inequalities for non mixing processes, such that dynamical systems called expanding maps, see Collet *et al.* [6] and continuous functions of Bernoulli shifts. The Bernstein's deviation inequality in these non mixing contexts sharpens the existing ones. The deviation inequality is obtained by dividing the sample $(X_1, \dots, X_n)$ in different blocks $(X_i, \dots, X_{i+k^*})$, where the length k^* must be carefully chosen and then by approximating non consecutive blocks by independent blocks using a coupling scheme.

The coupling scheme follows from a conditional Kantorovitch-Rubinstein duality due to Dedecker *et al.* [9] and detailed in Section 2. Using this coupling argument, a new deviation inequality can be stated in Section 3:

$$(1.2) \quad \mathbb{P}\left(S(f) \geq 5.8\sqrt{n\bar{\sigma}_{k^*}^2(f)x} + 1.5 k^* x\right) \leq e^{-x} \quad \text{for all } x \geq 0,$$

with $\bar{\sigma}_j^2(f) = \sup_{1 \leq k \leq n} \sigma_k^2(f)$ for all $1 \leq j \leq n$ and $k^* = \min\{k \geq 1; k\delta_k \leq \bar{\sigma}_k^2(f)\}$, where (δ_k) only depends of the φ -coefficients, see condition (3.1) for more details. Unlike $\sigma_1(f)$ in (1.1), the variance term $\bar{\sigma}_{k^*}^2(f)$ is natural as it tends to the limit variance in the CLT with k^* . When the TLC holds, i.e. $\bar{\sigma}_k^2(f) \rightarrow \sigma^2(f) > 0$, then the classical Bernstein's inequality (1.1) holds up to constants with $\sigma_1^2(f)$ replaced by $\sigma^2(f)$, see Subsection 3.3 for more details. On the opposite, if the functionals f_n are such that $\sigma_1^2(f_n) \rightarrow 0$, then for exponentially decreasing rate of the φ -coefficients, $k^* \approx -\ln(\bar{\sigma}_1^2(f_n))$ and a logarithmic loss appears. As in the recursive Markov chains case, the loss in the exponential approximation depends on the size of blocks k^* . We do not know if this loss in the exponential domain may be reduced for such non uniformly ϕ -mixing sequences.

In many practical examples such that chains with infinite memory introduced by Doukhan and Wintenberger [13], Bernoulli shifts and Markov kernels, an L^∞ coupling scheme is tractable, see Section 5 for a detailed definition. In these specific cases of φ -weakly dependent sequences, an

improved version of the deviation inequality (1.2) is given in Theorem 5.1:

$$\mathbb{P}\left(S(f) \geq 2\sqrt{n\sigma_{k^*}'^2(f)x} + 1.34 k^* x\right) \leq e^{-x} \quad \text{for all } x \geq 0,$$

with $k^* = \min\{1 \leq k \leq n / n\delta_k' \leq kx\}$, where (δ_k') only depends of the L^∞ coupling scheme, see condition (5.1) for more details. The paper finishes with the proofs collected in Section 6.

2. PRELIMINARIES: COUPLING AND WEAK DEPENDENCE COEFFICIENTS

Let $(X_1, \dots, X_n)$ with $n \geq 1$ be a sample of random variables on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with value in a metric space $(\mathcal{X}, d)$. We assume in all the sequel that for any $n \geq 1$ there exists a strictly stationary process $(X_t^{(n)})$ such that $(X_1, \dots, X_n) = (X_1^{(n)}, \dots, X_n^{(n)})$. Let us consider $\mathcal{F}$ the set of measurable functions $f : \mathcal{X} \mapsto \mathbb{R}$ satisfying:

$$(2.1) \quad |f(x) - f(y)| \leq d(x, y), \quad \forall (x, y) \in \mathcal{X} \times \mathcal{X} \quad \text{and} \quad \sup_{x \in \mathcal{X}} |f(x)| \leq 1/2.$$

We denote the partial sum $S(f) = \sum_{i=1}^n f(X_i)$ and $\mathcal{M}_j = \sigma(X_t; 1 \leq t \leq j)$ for all $1 \leq j \leq n$.

2.1. Kantorovitch-Rubinstein duality. The technique of coupling is related with the Kantorovitch-Rubinstein duality. The duality states that given two distribution P and Q on $\mathcal{X}$ there exists a random couple $Y = (Y_1, Y_2)$ with $Y_1 \sim P$ and $Y_2 \sim Q$ satisfying

$$\mathbb{E}(d(Y_1, Y_2)) = \sup_{f \in \Lambda_1} \mathbb{E}|f(dP - dQ)| = \inf_{Y'} \mathbb{E}(d(Y_1', Y_2')),$$

where Y' have the same margins than Y and Λ_1 denotes the set of 1-Lipschitz functions such that $|f(x) - f(y)| \leq d(x, y)$.

Dedecker, Prieur and Raynaud de Fitte [9] extend the classical Kantorovitch-Rubinstein duality in the time series framework by considering it conditionally on some event $\mathcal{M} \in \mathcal{A}$. Assuming that the original space Ω is rich enough, i.e. it exists a random variable U uniformly distributed over $[0, 1]$ and independent of $\mathcal{M}$, for any $Y_1 \sim P$ with values in a Polish space it exists a random variable $Y_2 \sim P$ independent of $\mathcal{M}$ satisfying

$$(2.2) \quad \mathbb{E}(d(Y_1, Y_2) \mid \mathcal{M}) = \sup\{|\mathbb{E}(f(Y_1) \mid \mathcal{M}) - \mathbb{E}(f(Y_1))|, f \in \Lambda_1\} \quad a.s.$$

2.2. φ -weak dependence coefficients and coupling schemes. Let us recall the weak dependence coefficient φ introduced in Rio [22]

Definition 2.1. For any $X \in \mathcal{X}$, for any σ -algebra $\mathcal{M}$ of $\mathcal{A}$ then

$$\varphi(\mathcal{M}, X) = \sup\{\|\mathbb{E}(f(X) \mid \mathcal{M}) - \mathbb{E}(f(X))\|_\infty, f \in \mathcal{F}\}.$$

Another equivalent definition is

$$(2.3) \quad \varphi(\mathcal{M}, X_r) = \sup\{|\text{Cov}(Y, f(X_r))|, f \in \mathcal{F} \text{ and } Y \text{ is } \mathcal{M}\text{-measurable and } \mathbb{E}|Y| = 1\},$$

see [8].

We will denote by **(A)** the specific case where $\mathcal{X}$ is a Polish space with $\sup_{(x,y) \in \mathcal{X}^2} d(x, y) \leq 1$. In the case **(A)** we have $\varphi(\mathcal{M}, X) = \tau_\infty(\mathcal{M}, X)$ where τ_∞ is the coupling coefficient defined in [7] by the relation

$$\tau_\infty(\mathcal{M}, X) = \sup\{\|\mathbb{E}(f(X) \mid \mathcal{M}) - \mathbb{E}(f(X))\|_\infty, f \in \Lambda_1\}.$$

This last coefficients is the essential supremum of the right hand side term of the conditional Kantorovitch-Rubinstein duality (2.2). Thus in the case **(A)** we get a coupling scheme directly on the variable X via the Kantorovitch-Rubinstein duality (2.2): it exists a version $X^* \sim X$ independent of $\mathcal{M}$ such that

$$\|\mathbb{E}(d(X, X^*) \mid \mathcal{M})\|_\infty = \tau_\infty(\mathcal{M}, X) = \varphi(\mathcal{M}, X).$$

When d is the Hamming distance $d(x, y) = \mathbb{1}_{x \neq y}$ the coefficient $\varphi(\mathcal{M}, X)$ coincides with the uniform mixing coefficient $\phi(\mathcal{M}, \sigma(X))$ of Ibragimov defined for 2 σ -algebra $\mathcal{M}$ and $\mathcal{M}'$ as:

$$\phi(\mathcal{M}, \mathcal{M}') = \sup_{M \in \mathcal{M}, M' \in \mathcal{M}'} |\mathbb{P}(M' \mid M) - \mathbb{P}(M')|.$$

In more general context than **(A)**, we have $\varphi(\mathcal{M}, X) \leq \tau_\infty(\mathcal{M}, X)$ and coupling scheme directly on X is not tractable. Then we do coupling scheme on the variables $f(X_i)$ for some function f : If the sample $(X_1, \dots, X_n)$ is such that the coefficients $\varphi(\mathcal{M}_j, X_i)$ are finite for $1 \leq j < i \leq n$ and if $f \in \mathcal{F}$ then the coupling scheme for $f(X_i)$ follows from the conditional Kantorovitch-Rubinstein duality (2.2) and the relation

$$\tau_\infty(\mathcal{M}_j, f(X_i)) \leq \varphi(\mathcal{M}_j, X_i) :$$

There exists $f(X_i)^*$ such that $f(X_i)^* \sim f(X_i)$ is independent of $\mathcal{M}_j$ and

$$\|\mathbb{E}(|f(X_i)^* - f(X_i)| \mid \mathcal{M}_j)\|_\infty = \tau_\infty(\mathcal{M}_j, f(X_i)) \leq \varphi(\mathcal{M}_j, X_i).$$

In the case **(A)** we also have another possible coupling scheme for $f(X_i)$, see Section 5 for practical examples: $f(X_i^*) \sim f(X_i)$ is independent of $\mathcal{M}_j$ and

$$\|\mathbb{E}(|f(X_i^*) - f(X_i)| \mid \mathcal{M}_j)\|_\infty \leq \|\mathbb{E}(d(X_i^*, X_i) \mid \mathcal{M}_j)\|_\infty = \tau_\infty(\mathcal{M}_j, f(X_i)) = \varphi(\mathcal{M}_j, X_i).$$

2.3. Extensions on the product space $\mathcal{X}^q$, $q > 1$. To consider conditional coupling schemes of length $q > 1$ we need to extend the notions of weak dependence coefficients on $X = (X_t)_{r \leq t < r+q} \in \mathcal{X}^q$. It depends on the metric d_q chosen for $\mathcal{X}^q$:

Definition 2.2. For any $q \geq 1$, any $X \in \mathcal{X}^q$ and any σ -algebra $\mathcal{M}$ of $\mathcal{A}$ let us define the coefficients

$$\varphi(\mathcal{M}, X) = \sup\{\|\mathbb{E}(f(X) \mid \mathcal{M}) - \mathbb{E}(f(X))\|_\infty, f \in \mathcal{F}_q\},$$

where $\mathcal{F}_q$ is the set of 1-Lipschitz functions with values in $[-1/2, 1/2]$ of $\mathcal{X}^q$ equipped with the metric $d_q(x, y) = q^{-1} \sum_{i=1}^q d(x_i, y_i)$.

Let us discuss the consequences of the choice of the metric d_q :

- The τ_∞ coupling coefficients on $\mathcal{X}^q$ are defined for the metric d_q , see [7], and for all $f \in \mathcal{F}$:

$$\tau_\infty(\mathcal{M}, (f(X_1), \dots, f(X_q))) \leq \varphi(\mathcal{M}, (X_1, \dots, X_q)).$$

Moreover, in the case **(A)** it holds $\tau_\infty(\mathcal{M}, X) = \varphi(\mathcal{M}, X)$.

- If d is the Hamming metric, as $d_q(x, y) \leq \mathbb{1}_{x \neq y}$ then $\varphi(\mathcal{M}, X) \leq \phi(\mathcal{M}, \sigma(X))$. Thus the definition of the weakly dependent coefficients φ differs here from the one of Rio in [22] where $\mathcal{X}^q$ is equipped with $d_\infty(x, y) = \max_{1 \leq i \leq q} d(x_i, y_i)$.

2.4. First application: deviation inequality of Hoeffding type. This application is due to Dedecker and Prieur [8]. Assume that $\mathcal{X}$ is a Polish space such that $\sup_{x,y} d(x,y) \leq 1$, i.e. we are in the case **(A)**. Assume that the coefficients $\varphi(\mathcal{M}_j, (X_{j+1}, \dots, X_n))$ are finite for all $1 \leq j \leq n-1$ and that $g : \mathcal{X}^n \rightarrow \mathbb{R}$ satisfies

$$|g(x_1, \dots, x_n) - g(y_1, \dots, y_n)| \leq \sum_{i=1}^n d(x_i, y_i).$$

As we are in the case **(A)** it exists a coupling scheme $(X_{j+1}^*, \dots, X_n^*)$ of $(X_{j+1}, \dots, X_n)$ for any $1 \leq j \leq n-1$ such that, keeping the same notation than in [23]:

$$\begin{aligned} \Gamma(g) &= \|\mathbb{E}(g(X_{j+1}, \dots, X_n) \mid \mathcal{M}_j) - \mathbb{E}(g(X_{j+1}, \dots, X_n))\|_\infty \\ &= \|\mathbb{E}(g(X_{j+1}, \dots, X_n) - g(X_{j+1}^*, \dots, X_n^*) \mid \mathcal{M}_j)\|_\infty \leq (n-j)\varphi(\mathcal{M}_j, (X_{j+1}, \dots, X_n)). \end{aligned}$$

Applying Theorem 1 of [23], if $\mathbb{E}(g(X_1, \dots, X_n)) = 0$ then for all $x \geq 0$ it holds:

$$\mathbb{P} \left(f(X_1, \dots, X_n) \geq \sqrt{2^{-1} \sum_{j=1}^n (1 + 2(n-j)\varphi(\mathcal{M}_j, (X_{j+1}, \dots, X_n)))^2 x} \right) \leq e^{-x}.$$

This deviation inequality of Hoeffding type only differs from the one for independence by a constant. However, such inequalities are not as satisfactory as Bernstein ones for statistical applications.

3. DEVIATION INEQUALITY AROUND THE MEAN INEQUALITY

Let us give an inequality for the deviation around the mean of $S(f) = \sum_{i=1}^n f(X_i)$ for $f \in \mathcal{F}$, with $(X_1, \dots, X_n)$ on the metric space $(\mathcal{X}, d)$ and such that there exists a non increasing sequence (δ_r) that satisfies

$$(3.1) \quad \sup_{1 \leq j \leq n-2r+1} \varphi(\mathcal{M}_j, (X_{r+j}, \dots, X_{2r+j-1})) \leq \delta_r \text{ for all } r \geq 1.$$

3.1. A deviation inequality of Bernstein type. Assume with no loss of generality that $\mathbb{E}(f(X_1)) = 0$.

Theorem 3.1. *For any integer n , if there exists (δ_r) as in (3.1) then*

$$\mathbb{P} \left(S(f) \geq 5.8 \sqrt{n \bar{\sigma}_{k^*}^2(f)} x + 1.5 k^* x \right) \leq e^{-x},$$

where $k^* = \min\{1 \leq k \leq n / k \delta_k \leq \bar{\sigma}_k^2(f)\}$ and $\bar{\sigma}_{k^*}^2(f) = \max\{\sigma_k^2(f) / k^* \leq k \leq n\}$.

The proof of this Theorem is given in Subsection 6.1. We adopt the convention $\min \emptyset = +\infty$ and the estimate is non trivial when $r \delta_r \rightarrow 0$ and $n \delta_n \geq \bar{\sigma}_n(f)$, i.e. for not too small values of n .

Remark that the variance term $\bar{\sigma}_{k^*}^2(f)$ is more natural than $\sigma_1^2(f)$ in (1.1) as in the central limit theorem $\bar{\sigma}_{k^*}^2(f)$ converges to the limit variance as k^* goes to infinity. Before giving some remarks on this Theorem, the next proposition give estimates of the quantity $\sigma_k^2(f) = k^{-1} \text{Var} \left(\sum_{i=1}^k f(X_i) \right)$.

3.2. The variance terms $\sigma_k^2(f)$. Under suitable assumptions on (δ_r) , it is always possible to obtain rough estimates of $\sigma_k^2(f)$ in function of $\sigma_1^2(f)$ and $\mathbb{E}|f(X_1)|$:

Proposition 3.2. *If the condition (3.1) is satisfied then we have for all $k \leq n$ the inequality:*

$$\sigma_k^2(f) \leq \left(\sigma_1^2(f) + 2\mathbb{E}|f(X_1)| \sum_{r=1}^{k-1} \delta_r \right).$$

See Subsection for a straightforward proof 6.2 of this Proposition. The estimate given in Proposition 3.2 can be rough, for example in the degenerate cases when $\sigma_k^2(f)$ tends to 0 with k .

3.3. Remarks on Theorem 3.1. The gaussian behavior around the mean is, up to a universal constant, the same than in the iid case with the more natural variance $\bar{\sigma}_{k^*}^2(f)$ instead of $\sigma_1^2(f)$ in (1.1). However in the exponential domain the estimates given in Theorem 3.1 is sometimes less sharp than the one obtained for ϕ -mixing in Samson [24].

In the non degenerate case $\sigma_k^2(f) \rightarrow \sigma^2(f) > 0$ then k^* is finite as soon as $r\delta_r \downarrow 0$. The deviation inequality of Theorem 3.1 becomes similar than the one in the iid case (1.1) with the variance term $\sigma^2(f)$ instead of $\sigma_1^2(f)$: there exists $C > 0$ such that for n sufficiently large we have

$$\mathbb{P}(S(f) \geq C(\sqrt{n\sigma^2(f)x} + x)) \leq e^{-x} \text{ for all } x \geq 0.$$

However, the estimate of the exponential behavior in Theorem 3.1 may differ from the one of the iid case. For example, for statistical issues it is often assumed that f is chosen depending on n such that $\sigma_1^2(f_n) \rightarrow 0$. Assume that $r\delta_r$ is summable. Using Proposition 3.2 and Jensen's inequality we have the estimate $\bar{\sigma}_1^2(f_n) \lesssim \sigma_1^2(f_n)^{1/2}$. If $\sigma_1^2(f_n)^{-1/2}n\delta_n \downarrow 0$ then for n sufficiently large such that $k_n^* = \min\{k \leq n / k\delta_k \leq \sigma_1^2(f_n)^{1/2}\}$ exists, it holds

$$\mathbb{P}(S(f) \geq C(\sqrt{\bar{\sigma}_{k_n^*}^2(f_n)nx} + k_n^*x)) \leq e^{-x} \text{ for all } x \geq 0, \text{ with } C > 0.$$

As $k_n^* \uparrow \infty$ there is a loss compare with the iid case (1.1). We do not know if this loss may be reduced outside the cases of uniformly mixing processes where (1.1) holds, see Samson [24].

This loss may be reduced when the autocorrelations are controlled, choosing a smaller size of blocks k_n^* . Assume that $\bar{\sigma}_1^2(f_n) \lesssim \sigma_1^2(f_n)$ (such relation is satisfied in the uniformly ϕ -mixing context). If $\sigma_1^2(f_n)^{-1}n\delta_n \downarrow 0$ then for n sufficiently large such that $k_n^* = \min\{k \leq n / k\delta_k \leq \sigma_1^2(f_n)\}$ exists, it holds

$$\mathbb{P}\left(S(f_n) \geq C\left(\sqrt{\sigma_1^2(f_n)nx} + k_n^*x\right)\right) \leq e^{-x} \text{ for all } x \geq 0, \text{ with } C > 0.$$

The loss compare with the iid case is due to $k_n^* \uparrow \infty$. More precisely

- If $\delta_r = C\delta^r$ for $C > 0$ and $0 < \delta < 1$ then $k_n^* \approx -\ln(\sigma_1^2(f_n))$,
- If $\delta_r = Cr^\delta$ for $C > 0$ and $\delta > 1$ then $k_n^* \approx \sigma_1^2(f_n)^{1/(1-\delta)}$.

4. EXAMPLES

We focus on non ϕ -mixing examples as for them the inequality (1.1) holds up to constants, see Samson [24]. We present dynamical systems that are known to be non ϕ -mixing processes but they satisfy (3.1) in the case **(A)**. Other examples in the case **(A)** are presented in the Section 5 as a sharpened deviation inequality holds for them, see Theorem 5.1. We also present in this

Section continuous functions of Bernoulli shifts that are examples not ϕ -mixing and not in the case **(A)** and thus cannot be treated by the approach of Section 5 and of [24].

4.1. Dynamical systems. Here we are in the case $\mathcal{X} = [0; 1]$ and $d(x, y) = |x - y|$, i.e. in the **(A)** and then $\varphi = \tau_\infty$. Since Andrews [2], dynamical systems, defined as stationary solutions of $X_t = T(X_{t+1})$ for all t are classical examples of non-mixing processes. Let us consider X_t the stationary solution of

$$X_t = \frac{1}{2}(X_{t-1} + \xi_t)$$

where (ξ_t) is an iid sequence distributed as a Bernoulli(1/2). Then $X_t = T(X_{t+1})$ where $T(x) = 2x$ modulo 1. Even if it is not mixing, easy computation shows that (X_t) satisfies (3.1) with $r\delta_r = (4/9)2^{-r}$ (in fact this specific case satisfies also $r\delta'_r = (4/9)2^{-r}$, see Section 5 for more details).

More general examples of dynamical systems are studied in Collet *et al.* [6]. They obtain estimates of covariances terms, multivariate versions of (2.3), for dynamical systems called expanding maps. Then it follows the existence of $C > 0$ and $0 < \rho < 1$ such that (3.1) is satisfied with $r\delta_r = C\rho^r$, see Dedecker and Prieur [8] for more details.

4.2. Continuous functions of Bernoulli shifts. Let us consider a ϕ -mixing stationary process (ξ_t) in some measurable space $\mathcal{Y}$ and a sequence (U_t) in the metric space $\mathcal{X}$ defined as

$$U_t = F(\xi_{t-j}; j \in \mathbb{N}),$$

where F is a measurable function. Assume that the original state space is large enough such that it exists (ξ'_t) distributed as (ξ_t) but independent of it. As in [22], assume that there exists a non increasing sequence (v_k) satisfying almost surely

$$d(F(\xi_j; j \in \mathbb{N}), F(\xi'_j; j \in \mathbb{N})) \leq v_k,$$

with the sequence (ξ'_t) satisfying $\xi_t = \xi'_t$ for $0 \leq t \leq k$ and for $t > k$, $\xi'_t = \xi_t$. Finally set $X_t = H(U_t)$ for some measurable function $H : \mathcal{X} \rightarrow \mathcal{X}$ and $t = \{1, \dots, n\}$ and denote

$$w_H(x, \eta) = \sup_{d(x, y) \leq \eta} d(H(x), H(y)).$$

Proposition 4.1. *The sample $(X_1, \dots, X_n)$ satisfies (3.1) with*

$$\delta_r = \inf_{1 \leq k \leq r-1} \{2\phi_{r-k} + \mathbb{E}(3w_H(U_0, 2v_k)) \wedge 1\}.$$

See the Subsection 6.3 for the proof of this Proposition. Remark that by construction the process (X_t) is non necessarily in the case **(A)**

5. IN THE CASE **(A)** WITH A COUPLING SCHEME IN L^∞ .

In all this section we place us in the case **(A)** where $\mathcal{X}$ is a Polish metric space and $d(x, y) \leq 1$ for all $x, y \in \mathcal{X}$. For all $r \geq 1$ a coupling scheme in L^∞ for $(X_i)_{r+j \leq i < 2r+j-1}$, $j \geq 1$, exists when we can construct $(X_i^*)_{r+j \leq i < 2r+j-1}$ distributed as $(X_i)_{r+j \leq i < 2r+j-1}$ and independent of $\mathcal{M}_j$ such that

$$(5.1) \quad \sup_{1 \leq j \leq n-2r+1} \sum_{i=r+j}^{2r+j-1} d(X_i, X_i^*) \leq r\delta'_r \quad a.s. \quad \text{for all } r \geq 1.$$

5.1. A sharper deviation inequality of Bernstein's type. Remark that condition (5.1) with (δ'_r) implies condition (3.1) with $\delta_r = \delta'_r$. Then we obtain a slightly sharper deviation inequality than in Theorem 3.1:

Theorem 5.1. *For $f \in \mathcal{F}$ such that $\mathbb{E}(f(X_1)) = 0$ then we have for all $x \geq n\delta'_k$ and all $1 \leq k \leq n$:*

$$\mathbb{P}(S(f) \geq x) \leq \exp \left(-\frac{2n\sigma_k^2(f)}{k^2} h \left(\frac{k(x - n\delta'_k)}{2n\sigma_k^2(f)} \right) \right)$$

where $h(x) = (1+x)\ln(1+x) - x$ for all $x \geq 0$. Then it holds for all $x \geq 0$:

$$\mathbb{P} \left(S(f) \geq 2\sqrt{n\sigma_{k^{*'}}^2(f)x} + 1.34 k^{*'} x \right) \leq \exp(-x)$$

with $k^{*'} = \min\{1 \leq k \leq n / n\delta'_k \leq kx\}$.

The proof of this Theorem is given in Subsection 6.4.

Let us compare this deviation inequality with the result of Theorem 3.1. In Theorem 5.1 the variance term $\sigma_k^2(f)$ sharpens $\bar{\sigma}_k^2(f)$ and the normal approximation is better here. For the exponential approximation, in both Theorems losses are due to the chosen blocks sizes. As $k^{*'} = \min\{1 \leq k \leq n / k\delta'_k \leq xk^2/n\}$, if $k\delta'_k$ is decreasing as $k\delta_k$ then $k^{*'} \leq k^*$ as soon as $n\sigma_k^2(f) \leq k^2x$ or equivalently $\sqrt{n\sigma_k^2(f)x} \leq kx$, i.e. as soon as x is in the domain of the exponential approximation. Thus for the normal and the exponential approximations, the deviation inequality in Theorem 5.1 improves the one of Theorem 3.1.

A tradeoff between the generality of the context and the sharpness of the deviation inequalities is done. Even if (5.1) is less general than (3.1), it is satisfied for many examples, see below.

5.2. Bounded Markov Chains. Following Dedecker and Prieur [8], let us consider a stationary Markov chain (X_t) with transition kernel P satisfying, for all $f \in \Lambda_1$, that $P(f) = \int f(y)P(x, dy)$ is a κ -Lipschitz function with $\kappa < 1$. Then

$$r\delta'_r = \kappa^r(1 + \dots + \kappa^r),$$

see [8] for more details.

5.3. Bounded chains with infinite memory. Let the sequence of the innovations $(\xi_t)_{t \in \mathbb{Z}}$ be an iid process on a measurable space $\mathcal{Y}$. We define $X = (X_t)_{t \in \mathbb{Z}}$ as the solution of the equation

$$(5.2) \quad X_t = F(X_{t-1}, X_{t-2}, \dots; \xi_t) \quad a.s.,$$

for some bounded function $F : \mathcal{X}^{(N \setminus \{0\})} \times \mathcal{Y} \rightarrow \mathcal{X}$ satisfying the condition

$$(5.3) \quad d(F((x_k)_{k \in N \setminus \{0\}}; \xi_0), F((y_k)_{k \in N \setminus \{0\}}; \xi_0)) \leq \sum_{j=1}^{\infty} a_j(F) d(x_j, y_j), \quad a.s.$$

for all $(x_k)_{k \in N \setminus \{0\}}, (y_k)_{k \in N \setminus \{0\}} \in \mathcal{X}^{N \setminus \{0\}}$ such that there exists $N > 0$ as $x_k = y_k = 0$ for all $k > N$ and with $a_j(F) \geq 0$ satisfying

$$(5.4) \quad \sum_{j=1}^{\infty} a_j(F) := a(F) < 1.$$

Let $(\xi_t^*)_{t \in \mathbb{Z}}$ be a stationary sequence distributed as $(\xi_t)_{t \in \mathbb{Z}}$, independent of $(\xi_t)_{t \leq 0}$ and such that $\xi_t = \xi_t^*$ for $t > 0$. Let $(X_t^*)_{t \in \mathbb{Z}}$ be the solution of the equation

$$X_t^* = F(X_{t-1}^*, X_{t-2}^*, \dots; \xi_t^*), \quad a.s.$$

Using similar arguments than in Doukhan and Wintenberger [13] we have the following result,

Lemma 5.2. *Under condition (5.4) there exists some bounded (by $1/2$) stationary process X solution of the equation (5.2). Moreover, this solution satisfies (5.1) with*

$$r\delta'_r = \sum_{j=r}^{2r-1} \inf_{0 < p \leq j} \left\{ a(F)^{r/p} + \sum_{j=p}^{\infty} a_j(F) \right\}.$$

As the proof of this Lemma is similar than the one in [13], it is omitted here.

Many solutions of econometrical models may be written as chains with infinite memory. However, the assumption of boundedness is very restrictive for practical models.

5.4. Bernoulli shifts. Solutions of the recurrence equation (5.2) may always be written as $X_t = H((\xi_j)_{j \leq t})$ for some measurable function $H : \mathcal{Y}^{\mathbb{N}} \mapsto \mathcal{X}$ where (ξ_t) is an iid process called the innovations. In this very general framework, a coupling version X_t^* is given by $X_t^* = H((\xi_t^*))$ where (ξ_t^*) is a stationary sequence distributed as (ξ_t) , independent of $(\xi_t)_{t \leq 0}$ and such that $\xi_t = \xi_t^*$ for $t > 0$. If there exist $a_i \geq 0$ such that

$$d(H(x), H(y)) \leq \sum_{i \geq 1} a_i d(x_i, y_i) \quad \text{with} \quad \sum_{i \geq 1} a_i < \infty,$$

and if $\mathcal{Y}$ is a metric space such that it exists $y \in \mathcal{Y}$ with $d(\xi_1, y)$ bounded a.s., then (X_t) satisfies (5.1) with

$$r\delta'_r = C \sum_{i \geq r} a_r$$

for some $C > 0$.

6. PROOFS

This Section contains the proofs.

6.1. Proofs of the Theorems 3.1. This section contains the proofs of the Bernstein's type estimates on the partial sums $S(f)$ for $f \in \mathcal{F}$. As in the independent case, the proofs follow the Chernoff device. We will proceed using Bernstein's block technique as in [10]. Let us denote by I_j the j -th block of indices of size k , i.e. $\{(j-1)k+1, jk\}$ except the last blocks and let p be an integer such that $2p-1 \leq k^{-1}n \leq 2p$.

Let us denote by S_1 and S_2 the sums of even and odd blocks defined as

$$S_1 = \sum_{i \in I_{2j}, 1 \leq j \leq p} f(X_i) \quad \text{and} \quad S_2 = \sum_{i \in I_{2j-1}, 1 \leq j \leq p} f(X_i).$$

From Cauchy-Schwartz inequality, it holds:

$$\ln \mathbb{E}[\exp(tS(f))] \leq \frac{1}{2} (\ln \mathbb{E} \exp(2tS_1) + \ln \mathbb{E} \exp(2tS_2)).$$

Now let us treat in detail the term depending on S_1 , the same argument applies identically to S_2 . We want to prove that for any $0 \leq t \leq 1$, choosing $k = [1/t] \wedge n$ as in [10] it holds:

$$(6.1) \quad \ln \mathbb{E}(\exp(tS(f))) \leq 4nt^2(2(e-2)\sigma_k^2(f) + ek\delta_k).$$

Denoting $L_m = \ln \mathbb{E}(\exp(2t \sum_{i \in I_{2j}, 1 \leq j \leq m} f(X_i)))$ for any $1 \leq m \leq p$, we do a recurrence on m remarking that $\ln \mathbb{E}(\exp(2tS_1)) = L_p$. From Holder inequality, we have for any $2 \leq m \leq p-1$ the inequalities:

$$\begin{aligned} & \exp(L_{m+1}) - \exp(L_m) \exp(L_1) \\ & \leq \exp(L_m) \left\| \mathbb{E} \left(\exp \left(2t \sum_{i \in I_{2(m+1)}} f(X_i) \right) \mid \mathcal{M}_{2mk} \right) - \mathbb{E} \left(\exp \left(2t \sum_{i \in I_{2(m+1)}} f(X_i) \right) \right) \right\|_{\infty} \\ & \leq \exp(L_m) \left\| \mathbb{E} \left(\exp \left(2t \sum_{i \in I_{2(m+1)}} f(X_i) \right) - \exp \left(2t \sum_{i \in I_{2(m+1)}} f(X_i) \right)^* \mid \mathcal{M}_{2mk} \right) \right\|_{\infty}, \end{aligned}$$

where $\exp \left(2t \sum_{i \in I_{2(m+1)}} f(X_i) \right)^*$ is a coupling version of the variable $\exp \left(2t \sum_{i \in I_{2(m+1)}} f(X_i) \right)$, independent of $\mathcal{M}_{2mk}$. From the definition of the coupling coefficients τ_{∞} , we know that

$$\begin{aligned} & \left\| \mathbb{E} \left(\exp \left(2t \sum_{i \in I_{2(m+1)}} f(X_i) \right) - \exp \left(2t \sum_{i \in I_{2(m+1)}} f(X_i) \right)^* \mid \mathcal{M}_{2mk} \right) \right\|_{\infty} \\ & \leq \tau_{\infty} \left(\mathcal{M}_{2mk}, \exp \left(2t \sum_{i \in I_{2(m+1)}} f(X_i) \right) \right). \end{aligned}$$

As $\sum_{i \in I_{2(m+1)}} f(X_i)$ is bounded with $k/2$, then $u \rightarrow \exp(2tu)$ is a Lipschitz function with constant $2kt \exp(kt)$ with respect to d_k and bounded with $\exp(kt)$ for all $t \geq 0$. We then deduce that for $n^{-1} < t \leq 1$, choosing $k = [1/t] \wedge (n-1)$ and under condition (3.1) we have

$$\tau_{\infty} \left(\mathcal{M}_{2mk}, \exp \left(2t \sum_{i \in I_{2(m+1)}} f(X_i) \right) \right) \leq 2kte^{kt} \varphi(\mathcal{M}_{2mk}, (X_i)_{i \in I_{2(m+1)}}) \leq 2e\delta_k.$$

Collecting this inequalities, we achieve that

$$\exp(L_{m+1}) \leq \exp(L_m)(\exp(L_1) + 2e\delta_k).$$

The classical Bennett's inequality on $\sum_{i \in I_2} f(X_i)$ gives the estimates $\exp(L_1) \leq 1 + 4\sigma_k^2(f)/k(e^{kt} - kt - 1)$ and as $kt \leq 1$ we obtain

$$L_{m+1} \leq L_m + \ln \left(1 + \frac{4(e-2)\sigma_k^2(f) + 2ek\delta_k}{k} \right) \leq L_m + \frac{4(e-2)\sigma_k^2(f) + 2ek\delta_k}{k}.$$

The p steps of the recurrence leads to the desired inequality

$$\ln \mathbb{E}(\exp(2tS_1)) \leq 2p \frac{2(e-2)\sigma_k^2(f) + ek\delta_k}{k}.$$

As the same inequality holds for S_2 we obtain (6.1) for $n^{-1} < t \leq 1$ remarking that $2pk^{-1} \leq 4nt^2$. For $t \leq n^{-1}$, classical Bennett inequality on S_1 gives

$$\ln \mathbb{E}(\exp(2tS_1)) \leq 4\sigma_n^2(f)/n(e^{nt} - nt - 1).$$

Remarking that $e^{nt} - nt - 1 \leq (nt)^2 \sum_{k \geq 0} (nt)^k / (k+2)!$ and $(k+2)! \geq 23^k$ we derive that $e^{nt} - nt - 1 \leq 2^{-1}(nt)^2 \sum_{k \geq 0} 3^{-k} \leq 3/4(nt)^2$ for $nt \leq 1$. Then collecting these bounds, for $t \leq n^{-1}$ it holds

$$\ln \mathbb{E}(\exp(2tS_1)) \leq 3n\bar{\sigma}_n^2(f)t^2 \leq 4nt^2(2(e-2)\sigma_n^2(f) + en\delta_n).$$

The same holds for S_2 and then (6.1) follows for $0 \leq t \leq n^{-1}$ and then for all $0 \leq t \leq 1$.

Note that for $k \geq k^*$ we have $\sigma_k^2(f) \leq \bar{\sigma}_{k^*}^2(f)$ and $k\delta_k \leq \bar{\sigma}_{k^*}^2(f)$ by definition. From (6.1) we achieve

$$\ln \mathbb{E}(\exp(tS(f))) \leq Kn\bar{\sigma}_{k^*}^2(f)t^2, \text{ for } 0 \leq t \leq k^{*-1},$$

with $K = 4(3e-4)$. Follow the Chernoff's device, i.e. using $\ln \mathbb{P}(S(f) \geq x) \leq \ln \mathbb{E}(\exp(tS(f))) - tx$ and optimizing in $0 \leq t \leq k^{*-1}$, we obtain

$$\mathbb{P}(S(f) \geq x) \leq \exp\left(-\frac{x^2}{2Kn\bar{\sigma}_{k^*}^2(f)}\right) \mathbb{1}_{k^*x \leq 2Kn\bar{\sigma}_{k^*}^2(f)} + \exp\left(\frac{Kn\bar{\sigma}_{k^*}^2(f)}{k^{*2}} - \frac{x}{k^*}\right) \mathbb{1}_{k^*x > 2Kn\bar{\sigma}_{k^*}^2(f)}.$$

Easy calculation yields for all $x \geq 0$

$$\mathbb{P}(S(f) \geq x) \leq \sqrt{2Kn\bar{\sigma}_{k^*}^2(f)} \mathbb{1}_{k^{*2}x \leq 2Kn\bar{\sigma}_{k^*}^2(f)} + (k^*t + k^{*-1}Kn\bar{\sigma}_{k^*}^2(f)) \mathbb{1}_{k^{*2}x > 2Kn\bar{\sigma}_{k^*}^2(f)} \leq e - x.$$

A rough bound $k^*t + k^{*-1}Kn\bar{\sigma}_{k^*}^2(f) \leq 3k^*x/2$ for $k^{*2}x > 2Kn\bar{\sigma}_{k^*}^2(f)$ leads to the result of the Theorem.

6.2. Proof of Proposition 3.2. We have the classical decomposition

$$\text{Var}\left(\sum_{i=1}^k f(X_i)\right) = k \text{Var}(f(X_1)) + 2 \sum_{r=1}^{k-1} (k-r) \text{Cov}(f(X_1), f(X_{r+1})).$$

Now let us consider the coupling scheme $f(X_{r+1})^*$ distributed as $f(X_{r+1})$ but independent of $\mathcal{M}_1$. Then from Holder inequality it holds

$$\text{Cov}(f(X_1), f(X_{r+1})) = \mathbb{E}(\mathbb{E}(f(X_{r+1}) - f(X_{r+1})^* | \mathcal{M}_1) f(X_1)).$$

But as $f(X_{r+1}) - f(X_{r+1})^* \leq \delta_r$ conditionally to $\mathcal{M}_0$ we get the desired result.

6.3. Proof of Proposition 4.1. We adapt the proof of [22]. We are interested in estimated the coefficients $\varphi(\mathcal{M}_j, (X_{r+j}, \dots, X_{2r-1+j}))$ for any (j, r) satisfying $1 \leq j \leq j+r \leq 2r-1+j \leq n$. Let us fix (j, r) and denote (ξ_t^k) a sequence such that $\xi_t^k = \xi_t$ for all $t \geq r+j-k > j$ and $\xi_t^k = \xi_t'$ otherwise. Denote $U_t^k = F(\xi_{t-j}^k; j \in \mathbb{N})$ and $X_t^k = H(U_t^k)$. For any $f \in \mathcal{F}$, we have

$$(6.2) \quad f(X_{r+j}, \dots, X_{2r-1+j}) - f(X_{r+j}^k, \dots, X_{2r-1+j}^k) \leq \left(\frac{1}{r} \sum_{i=r+j}^{2r-1+j} d(X_i, X_i^k) \right) \wedge 1.$$

By definition of the modulus of continuity and as $d(U_i^k, U_i) \leq v_k$ for any $r+j \leq i \leq 2r-1+j$, we have

$$d(X_i, X_i^k) = d(H(U_i), H(U_i^k)) \leq w_H(U_i^k, v_k).$$

Remarking that $\left(r^{-1} \sum_{i=r+j}^{2r-1+j} w_H(U_i^k, v_k)\right) \wedge 1$ is a measurable function of $((\xi'_t)_{t < r+j-k}, (\xi_t)_{t \geq r+j-k})$ bounded by 1, it holds from the definition of the ϕ -mixing coefficients:

$$\mathbb{E} \left(\left(r^{-1} \sum_{i=r+j}^{2r-1+j} w_H(U_i^k, v_k) \right) \wedge 1 \mid \mathcal{M}_j \right) \leq \phi_{r-k} + \mathbb{E} \left(\left(r^{-1} \sum_{i=r+j}^{2r-1+j} w_H(U_i^k, v_k) \right) \wedge 1 \right).$$

Using again that $d(U_i^k, U_i) \leq v_k$, then $w_H(U_i^k, v_k) \leq 2w_H(U_i, 2v_k)$. By stationarity of (U_t) , we obtain

$$\mathbb{E} \left(\left(r^{-1} \sum_{i=r+j}^{2r-1+j} w_H(U_i^k, v_k) \right) \wedge 1 \right) \leq \mathbb{E}(2w_H(U_0, 2v_k)) \wedge 1.$$

So combining these inequalities we obtain for all $1 \leq k \leq r-1$:

$$(6.3) \quad \left\| \mathbb{E} \left(f(X_{r+j}, \dots, X_{2r-1+j}) - f(X_{r+j}^k, \dots, X_{2r-1+j}^k) \mid \mathcal{M}_j \right) \right\|_{\infty} \leq \phi_{r-k} + \mathbb{E}(2w_H(U_0, 2v_k)) \wedge 1.$$

Using again the definition of the ϕ -mixing coefficients, as f is bounded by 1 it holds

$$(6.4) \quad \left\| \mathbb{E} \left(f(X_{r+j}^k, \dots, X_{2r-1+j}^k) \mid \mathcal{M}_j \right) - \mathbb{E} \left(f(X_{r+j}^k, \dots, X_{2r-1+j}^k) \right) \right\|_{\infty} \leq \phi_{r-k}.$$

Finally, using again (6.2) and that $d(X_i, X_i^k) \leq w_H(U_i, v_k)$, by stationarity of (U_t) we obtain

$$(6.5) \quad \mathbb{E} f(X_{r+j}, \dots, X_{2r-1+j}) - \mathbb{E} f(X_{r+j}^k, \dots, X_{2r-1+j}^k) \leq \mathbb{E}(w_H(U_0, v_k)) \wedge 1.$$

The result of the Proposition 4.1 follow from the definition of the φ -coefficients, the inequalities (6.3), (6.4) and (6.5).

6.4. Proof of Theorem 5.1. Let us keep the same notation than in the proof of Theorem 3.1. The Bennett's type deviation inequality follows classically from the Chernoff device applies with the estimate:

$$(6.6) \quad \ln(\mathbb{E}(\exp(tS(f)))) \leq \frac{2n\sigma_k^2(f)}{k^2}(\exp(kt) - kt - 1) + n\delta'_k t \quad \text{for all } t \geq 0.$$

To prove (6.6), let us use the L^∞ -coupling scheme and (5.1) to derive for all $1 \leq m \leq p$:

$$\left\| \sum_{i \in I_{2m}} f(X_i) - \sum_{i \in I_{2m}} f(X_i^*) \right\|_{\infty} \leq \sum_{i \in I_{2(m+1)}} \|d(X_i, X_i^*)\|_{\infty} \leq k\delta'_k,$$

where, as in Subsection 6.1, $|I_j| = k$ for all $1 \leq j \leq 2p$ with $2p-1 \leq nk^{-1} \leq 2p$. Then, for all $t \geq 0$ we have:

$$\exp \left(2t \sum_{i \in I_{2m}} f(X_i) \right) \leq e^{2tk\delta'_k} \exp \left(2t \sum_{i \in I_{2m}} f(X_i^*) \right) \quad \text{a.s.}$$

for all $1 \leq m \leq p$. In particular, by independence of $(X_i^*)_{i \in I_{2m}}$ with $\mathcal{M}_{2i-1}$ and by stationary we deduce that

$$\mathbb{E} \left(\exp \left(2t \sum_{i \in I_{2m}} f(X_i) \right) \mid \mathcal{M}_{2(m-1)} \right) \leq e^{2tk\delta'_k} \mathbb{E} \left(\exp \left(2t \sum_{i \in I_1} f(X_i^*) \right) \right)$$

for all $1 \leq m \leq p$. Applying this inequality for $m = p$ we have

$$\begin{aligned} \mathbb{E} \exp(2tS_1) &= \mathbb{E} \left(\exp \left(2t \sum_{1 \leq m \leq p-1} \sum_{i \in I_{2m}} f(X_i) \right) \mathbb{E} \left(\exp \left(\sum_{i \in I_{2p}} f(X_i) \right) \mid \mathcal{M}_{2(p-1)} \right) \right) \\ &\leq e^{2tk\delta'_k} \mathbb{E} \left(\exp \left(2t \sum_{i \in I_1} f(X_i^*) \right) \right) \mathbb{E} \left(\exp \left(2t \sum_{1 \leq m \leq p-1} \sum_{i \in I_{2m}} f(X_i) \right) \right). \end{aligned}$$

Let us do the same reasoning recursively on $m = p-1, \dots, 2$ to obtain finally

$$\ln \mathbb{E} \exp(2tS_1) \leq 2(p-1)k\delta'_k t + p \ln \mathbb{E} \left(\exp \left(2t \sum_{i \in I_1} f(X_i^*) \right) \right).$$

The classical Bennett inequality gives

$$\ln \mathbb{E} \left(\exp \left(2t \sum_{i \in I_1} f(X_i^*) \right) \right) \leq \frac{4\sigma_k^2(f)}{k} (\exp(kt) - kt - 1)$$

and the inequality (6.6) follows remarking that $4pk^{-1} \leq 2nk^{-2}$ and $2(p-1)k \leq n$.

For the Bernstein's type inequality, we use (6.6), the series expansion of the function $\exp(x) - x - 1$ and that $k! \geq 23^{k-2}$ for $k \geq 2$ to derive:

$$\ln(\mathbb{E}(\exp(tS(f)))) \leq \frac{n\sigma_k^2(f)t^2}{1 - (k/3)t} + n\delta'_k t \quad \text{for all } t \geq 0.$$

With the same notation than in [18], for $x \geq n\delta'_k$ the Chernoff device leads to:

$$\mathbb{P}(S(f) \geq x) \leq \exp \left(\frac{2n\sigma_k^2(f)}{(k/3)^2} h_1 \left(\frac{(k/3)(x - n\delta'_k)}{2n\sigma_k^2(f)} \right) \right),$$

where $h_1(x) = 1 + x - \sqrt{1 + 2x}$ for all $x \geq 0$. Then for all $x \geq 0$ we have

$$\mathbb{P}(S(f) \geq x + n\delta'_k) \leq \exp \left(\frac{2n\sigma_k^2(f)}{(k/3)^2} h_1 \left(\frac{(k/3)x}{2n\sigma_k^2(f)} \right) \right)$$

and the desired result follows as $h_1^{-1}(x) = \sqrt{2x} + x$ for all $x \geq 0$.

Acknowledgments. The author is grateful to Jérôme Dedecker for his helpful comments.

REFERENCES

- [1] ADAMCZAK, R. A tail inequality for suprema of unbounded empirical processes with applications to markov chains. *Electron. J. Probab.* 13 (2008), 1000–1034.
- [2] ANDREWS, D. W. K. Nonstrong mixing autoregressive processes. *J. Appl. Probab.* 21, 4 (1984), 930–934.
- [3] BENNETT, G. Probability inequalities for the sum of independant random variables. *Journal of the American Statistician Association* 57 (1962), 33–45.
- [4] BERTAIL, P., AND CLMENON, S. Sharp bounds for the tail of functionals of markov chains. to appear *Probability Theory and its applications*.
- [5] CATONI, O. *Statistical Learning Theory and Stochastic Optimization, Lecture Notes in Mathematics (Saint-Flour Summer School on Probability Theory 2001)*. Springer, 2004.

- [6] COLLET, P., MARTINEZ, S., AND SCHMITT, B. Exponential inequalities for dynamical measures of expanding maps of the interval. *Probability Theory and Related Fields* 123 (2002), 301–322.
- [7] DEDECKER, J., DOUKHAN, P., LANG, G., LEÓN, J. R., LOUHICHI, S., AND PRIEUR, C. *Weak Dependence, Examples and Applications*, vol. 190 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin, 2007.
- [8] DEDECKER, J., AND PRIEUR, C. New dependence coefficients: Examples and applications to statistics. *Probability Theory and Related Fields* 132 (2005), 203–235.
- [9] DEDECKER, J., PRIEUR, C., AND RAYNAUD DE FITTE, P. Parametrized kantorovich-rubinstein theorem and application to the coupling of random variables. In *Dependence in Probability and Statistics* (New York, 2006), P. Bertail, P. Doukhan, and P. Soulier, Eds., vol. 187 of *Lecture Notes in Statistics*, Springer, pp. 105–121.
- [10] DOUKHAN, P. *Mixing*, vol. 85 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1994.
- [11] DOUKHAN, P., AND LOUHICHI, S. A new weak dependence condition and applications to moment inequalities. *Stochastic Process. Appl.* 84, 2 (1999), 313–342.
- [12] DOUKHAN, P., AND NEUMANN, M. A bernstein type inequality for times series. *Stoch. Proc. Appl.* 117-7 (2007), 878–903.
- [13] DOUKHAN, P., AND WINTENBERGER, O. Weakly dependent chains with infinite memory. *Stochastic Process. Appl.* 118 (2008), 1997–2013.
- [14] IBRAGIMOV, I. Some limit theorems for stationary processes. *Theory of Probab. Appl.* 7 (1962), 349–382.
- [15] IBRAGIMOV, I. A., AND LINNIK, Y. V. *Independent and stationary sequences of random variables*. Wolters-Noordhoff Publishing, 1971.
- [16] JOULIN, A., AND OLLIVIER, Y. Curvature, concentration, and error estimates for markov chain monte carlo. arXiv:0904.1312v1.
- [17] LEZAUD, P. Chernoff-type bound for finite markov chains. *Ann. Appl. Probab.* 8 (1998), 849–867.
- [18] MASSART, P. *Concentration Inequalities and Model Selection, Lecture Notes in Mathematics (Saint-Flour Summer School on Probability Theory 2003)*. Springer, 2006.
- [19] MERLEVEDE, F., PELIGRAD, M., AND RIO, E. Bernstein inequality and moderate deviations under strong mixing conditions. available at <http://hal.inria.fr/inria-00360856/fr/>.
- [20] MERLEVEDE, F., PELIGRAD, M., AND RIO, E. A bernstein type inequality and moderate deviations for weakly dependent sequences. available at <http://hal.inria.fr/inria-00358525/fr/>.
- [21] NUMMELIN, E. A splitting technique for harris recurrent markov chains. *Z. Wahrsch. Verw. Gebiete* 43 (1978), 309–318.
- [22] RIO, E. The berry-esseen theorem for weakly dependent sequences. *Probab. Theory Related Fields* 104 (1996), 255–282.
- [23] RIO, E. Ingalités de hoeffding pour les fonctions lipschitziennes de suites dpendantes. *Comptes Rendus de l’Académie des Sciences de Paris, Série I* 330 (2000), 905–908.
- [24] SAMSON, P.-M. Concentration of measure inequalities for markov chains and ϕ -mixing processes. *Ann. Probab.* 28 (2000), 416–461.
- [25] VIENNET, G. Inequalities for absolutely regular sequences: application to density estimation. *Probab. Theory Relat. Fields* 107 (1997), 467–492.

CEREMADE, UMR CNRS 7534, UNIVERSITÉ DE PARIS - DAUPHINE, PLACE DU MARÉCHAL DE LATTRE DE TASSIGNY, 75775 PARIS CEDEX 16 - FRANCE, WINTENBERGER@CEREMADE.DAUPHINE.FR