DEVIATION INEQUALITIES FOR SUMS OF WEAKLY DEPENDENT TIME SERIES

 WINTENBERGER OLIVIER
CEREMADE, UMR CNRS 7534
Université de PARIS - DAUPHINE
Place du Maréchal De Lattre De Tassigny
75775 PARIS CEDEX 16 - FRANCE
email: `wintenberger@ceremade.dauphine.fr`

Abstract

In this paper we extend the classical Bernstein inequality for partial sums from the independent case to two cases of weakly dependent time series. The losses compared with the independent case are studied carefully. We give several examples that satisfy our deviation inequalities. The proofs are based on the blocks technique and two different coupling arguments.

1 Introduction

The aim of this paper is to extend the Bernstein inequality from the independent case to some weakly dependent ones. We consider a sample $(X_1, \ldots, X_n) = (X_1^{(n)}, \ldots, X_n^{(n)})$ of a stationary process $(X_t^{(n)})$ with values in a metric space $(\mathcal{X}, d)$. Let $\mathcal{F}$ be the set 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 $\mathbb{E}(f(X_1)) = 0$. Denoting $\sigma_k^2(f) = k^{-1} \text{Var}(\sum_{i=1}^k f(X_i))$, the Bernstein inequality in the indenpendent case writes as, see Bennett [2]:

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

This inequality reflects the gaussian approximation of the tail of $S(f)$ for small values x and the exponential approximation of the tail of $S(f)$ for large values of x.

When one extends the Bernstein 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 [16], Joulin and Ollivier [15] under granularity. For Harris recurrent Markov chains, Bertail and Clemençon [3]

obtain a deviation inequality of a different nature. They prove the existence of $C > 0$ such that for all $M > 0$ and all $x > 0$:

$$\mathbb{P}(S(f) \geq C(\sqrt{n\sigma^2(f)x} + Mx)) \leq e^{-x} + n\mathbb{P}(\tau_1 \geq M),$$

where the T_is are the return times to the atom, $\tau_i = T_i - T_{i-1}$ and, if T_0 is the first return time to the atom after time 0, $\sigma^2(f) = \mathbb{E}(\tau_1)^{-1} \text{Var}(\sum_{i=T_0+1}^{T_1} f(X_i))$. Remark that $\sigma^2(f)$ is the sum of the series of the covariances attached to the sequence $(f(X_i))$ and thus the asymptotic variance in the CLT satisfied by $S(f)$. This approximation of the deviation of $S(f)$ is natural as, through the splitting technique of Nummelin [20], the partial sum $S(f)$ is the sum of iid sums with τ_i number of summands. If the regeneration times admit finite exponential moments, fixing $M \approx \ln n$ Adamczak [1] obtains the existence of a constant $C > 0$ satisfying

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

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 such as uniformly ϕ-mixing coefficients introduced by Ibragimov [13]. When $\sum \sqrt{\phi}_r < \infty$, Samson [23] achieves the deviation inequality (1.1) with different constants. Less accurate results have been obtained for more general mixing coefficients: Viennet [24] for absolutely regular β-mixing coefficients and Merlevède et al. [18] for geometrically strongly mixing coefficients. Recently, mixing coefficients have been extended to weakly dependent ones, see Doukhan and Louhichi [10] and Dedecker and Prieur [7]. If these coefficients decay geometrically, deviation inequalities for $S(f)$ are given in Doukhan and Neumann [11] and Merlevède et al. [19] extend these results for unbounded functions f.

The dependence context of this paper is the one of the φ-weakly dependent coefficients introduced by Rio in [21] to extend the uniformly ϕ-mixing coefficients. These coefficients are used to deal with non mixing processes, such as dynamical systems called expanding maps, see Collet et al. [7] and continuous functions of Bernoulli shifts, see Rio [21]. The Bernstein's deviation inequality given in Theorem 3.1 sharpens the existing ones in φ-weak dependence context. The deviation inequality is obtained by dividing the sample $(X_1, \ldots, X_n)$ in different blocks $(X_i, \ldots, X_{i+k^*})$, 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. [8] and detailed in Section 2.

Using this coupling scheme, we provide a new deviation inequality in Section 3:

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

with $\overline{\sigma}^2_j(f) = \sup_{j \leq k \leq n} \sigma^2_k(f)$ for all $1 \leq j \leq n$ and $k^* = \min\{k \geq 1; \ k\delta_k \leq \overline{\sigma}^2_k(f)\}$, where (δ_k) only depends on the φ-coefficients, see condition (3.1) for more details. Under dependence, the variance term $\overline{\sigma}^2_{k^*}(f)$ is more natural than $\sigma_1(f)$ as it tends to the limit variance in the CLT with k^*. When the non degenerate CLT holds, i.e. $\overline{\sigma}^2_k(f) \to \sigma^2(f) > 0$, then the size k^* of the blocks is bounded uniformly over n and the classical Bernstein's inequality (1.1) holds with different constants. But if the functionals f_n are such as $\overline{\sigma}^2_k(f_n) \to 0$ for all k (it is the case when studying

the risk of estimators, see [17] and [4] for more details), then a loss appears in the exponential approximation. As for Harris recurrent Markov chains, the loss is due to the size k^* of the blocks that goes to infinity with n. In contrast with the Harris recurrent Markov chains context, it is still an open question wether the loss in the exponential approximation can be reduced or not.

However, in many bounded φ-weakly dependent examples (such as chains with infinite memory, Bernoulli shifts and Markov chains), an L^∞ coupling scheme for (X_t) is directly tractable, see Section 5 for details. Using this different coupling scheme, an improved version of the deviation inequality (1.2) is given in Theorem 5.1:

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

with $k^{*\prime} = \min\{1 \leq k \leq n \ / \ n\delta'_k \leq kx\}$, where (δ'_k) only depends on the L^∞ coupling scheme, see condition (5.1) for more details. The refinement is due to a smaller size $k^{*\prime}$ of blocks than the size k^* used in (1.2), see Remark 5.1. Moreover, the size $k^{*\prime}$ is fixed independently of $\overline{\sigma}^2_k(f)$. Then we provide a convenient dependent context for the study of the risk of estimators: the limiting behavior $\overline{\sigma}^2_k(f_n) \to 0$ does not affect the accuracy of the deviation inequality when an L^∞ coupling scheme exists.

2 Preliminaries: coupling and weak dependence coefficients

Let $(X_1, \ldots, X_n)$ 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 $(X_1, \ldots, X_n) = (X_1^{(n)}, \ldots, X_n^{(n)})$. Let us consider $\mathcal{F}$ the set of measurable functions $f : \mathcal{X} \mapsto \mathbb{R}$ satisfying

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

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 classical Kantorovitch-Rubinstein 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 *et al.* [8] extend the classical Kantorovitch-Rubinstein duality 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 there exists a random variable $Y_2 \sim P$ independent of $\mathcal{M}$ satisfying

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

2.2 φ-weak dependence coefficients and coupling schemes

Rio [21] defines the weak dependence coefficient φ as follows:

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

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

Another equivalent definition is given in [7]:

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

Let the condition **(A)** be satisfied when $\mathscr{X}$ is a metric Polish space and $\sup_{(x,y)\in\mathscr{X}^2} d(x,y) \leq 1$. Under **(A)** we have $\varphi(\mathscr{M}, X) = \tau_\infty(\mathscr{M}, X)$ where τ_∞ is defined in [6] by the relation

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

This coefficient is the essential supremum of the right hand side term of the identity (2.2). Thus the conditional Kantorovitch-Rubinstein duality (2.2) provides a first coupling scheme directly on the variable X: it exists a version $X^* \sim X$ independent of $\mathscr{M}$ such that

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

A well known case corresponds to the Hamming distance $d(x, y) = \mathbb{1}_{x \neq y}$ that satisfies **(A)** for any $\mathscr{X}$. The coefficient $\varphi(\mathscr{M}, X)$ then coincides with the uniform mixing coefficient $\phi(\mathscr{M}, \sigma(X))$ of Ibragimov defined for 2 σ-algebras $\mathscr{M}$ and $\mathscr{M}'$ by the relation

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

In a more general context than **(A)**, we have $\varphi(\mathscr{M}, X) \leq \tau_\infty(\mathscr{M}, X)$ and coupling scheme directly on the variable X does not follow from the Eqn.(2.2). However, there exists a new coupling scheme on the variables $f(X_i)$ for $f \in \mathscr{F}$. If the sample $(X_1, \ldots X_n)$ satisfies $\varphi(\mathscr{M}_j, X_i) < \infty$ for all $1 \leq j < i \leq n$ then a coupling scheme for $f(X_i)$ follows from the identity (2.2) and the relation

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

there exists $f(X_i)^*$ independent of $\mathscr{M}_j$ such that $f(X_i)^*$ is distributed as $f(X_i)$ and

$$\|\mathbb{E}(|f(X_i)^* - f(X_i)| \mid \mathscr{M}_j)\|_\infty = \tau_\infty(\mathscr{M}_j, f(X_i)) \leq \varphi(\mathscr{M}_j, X_i). \qquad (2.4)$$

In this paper we use this second coupling scheme as it is more general than the first one.

For some examples satisfying **(A)**, we use a third coupling scheme in Section 5. There, the dynamic of $(X_t^{(n)})$ provides the existence of X_i^* distributed as X_i and independent of $\mathscr{M}_j$ such that $d(X_i, X_i^*)$ is estimated a.s. by a sequence u_{i-j} decreasing to 0 when $i - j \to \infty$. Such coupling scheme in L^∞ provides immediately another L^∞ coupling scheme for $f(X_i)$ ($f \in \mathscr{F}$): $f(X_i^*)$ independent of $\mathscr{M}_j$ is distributed as $f(X_i)$ and

$$|f(X_i^*) - f(X_i)| \leq d(X_i^*, X_i) \leq u_{i-j} \qquad a.s.$$

2.3 Extensions to the product space $\mathscr{X}^q$, $q > 1$.

To consider conditional coupling schemes of length $q > 1$, we define the notion of φ-coefficients for $X = (X_t)_{r \leq t < r+q} \in \mathscr{X}^q$.

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

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

where $\mathscr{F}_q$ is the set of 1-Lipschitz functions with values in $[-1/2, 1/2]$ of $\mathscr{X}^d$ 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 coefficient τ_∞ is defined for $\mathscr{X}^q$ equipped with the same metric d_q, see [6]. Thus we have $\tau_\infty(\mathscr{M}, X) = \varphi(\mathscr{M}, X)$ under **(A)** and $\tau_\infty(\mathscr{M}, (f(X_1), \ldots, f(X_q))) \leq \varphi(\mathscr{M}, (X_1, \ldots, X_q))$ for all $f \in \mathscr{F}$.

- As $d_q(x, y) \leq \mathbb{1}_{x \neq y}$ then $\varphi(\mathscr{M}, X) \leq \phi(\mathscr{M}, \sigma(X))$; our definition of the coefficient φ differs from the one of Rio in [21] where $\mathscr{X}^q$ is equipped with $d_\infty(x, y) = \max_{1 \leq i \leq q} d(x_i, y_i)$.

3 A deviation inequality under φ-weak dependence

Assume that there exists a non increasing sequence $(\delta_r)_{r>0}$ satisfying

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

3.1 A Bernstein type inequality

Theorem 3.1. *If condition (3.1) is satisfied, for any $f \in \mathscr{F}$ such that $\mathbb{E}(f(X_1)) = 0$ we have*

$$\mathbb{P}\left(S(f) \geq 5.8\sqrt{n\overline{\sigma}_{k^*}^2(f)x} + 1.5 k^* x\right) \leq e^{-x}, \tag{3.2}$$

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

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 $r\delta_r \to 0$ and $n\delta_n \geq \overline{\sigma}_n(f)$, i.e. for not too small values of n.

3.2 The variance terms $\sigma_k^2(f)$

Before giving some remarks on Theorem 3.1, the next proposition give estimates of the quantity $\sigma_k^2(f) = k^{-1} \text{Var}(\sum_{i=1}^k f(X_i))$:

Proposition 3.2. *Under the assumption of Theorem 3.1, for any $1 \leq k \leq n$ we have*

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

See Subsection 6.2 for a proof of Proposition 3.2. The estimate given in Proposition 3.2 can be rough, for example in the degenerate case when $\sigma_k^2(f)$ tends to 0 with k. Note also that this estimate is often useless when the correlations terms are summable as the inequality $\sigma_k^2(f) \leq \sigma_1^2(f) \sum_{k \geq 1} |\text{Corr}(f(X_0), f(X_k))|$ may lead to better estimates.

3.3   Remarks on Theorem 3.1

Remark that the variance term $\overline{\sigma}^2_{k^*}(f)$ is more natural than $\sigma^2_1(f)$ in (1.1) as it converges to the asymptotic variance in the CLT as k^* goes to infinity. In the non degenerate case $\sigma^2_k(f) \to \sigma^2(f) > 0$ then k^* is finite as soon as $r\delta_r \downarrow 0$. The deviation inequality (3.1) becomes similar to (1.1), except its variance term $\overline{\sigma}^2_1(f)$ instead of $\sigma^2_1(f)$: there exists $C > 0$ satisfying

$$\mathbb{P}(S(f) \geq 5.8\sqrt{n\overline{\sigma}^2_1(f)x} + Cx) \leq e^{-x} \quad \text{for all} \quad x \geq 0.$$

Remark that $\overline{\sigma}^2_1(f)$ may be replaced by the asymptotic variance $\sigma^2(f)$ (or the marginal variance $\sigma^2_1(f)$) at the cost of a constant larger than 5.8.

In the degenerate case the size of blocks k^* might depend on n and the deviation inequality (3.1) is no longer of the same type than (1.1). Consider the statistical example of the deviation of the risk of an estimator, see [17] and [4] for more details; there f depends on n and $\overline{\sigma}^2_k(f_n) \to 0$ with n for all k. Assume that $k^*_n = \min\{k \leq n \,/\, k\delta_k \leq \overline{\sigma}^2_k(f_n)\}$ exists, then we obtain

$$\mathbb{P}(S(f) \geq 5.8\sqrt{\overline{\sigma}^2_{k^*_n}(f_n)nx} + 1.5\,k^*_n x) \leq e^{-x} \text{ for all } x \geq 0.$$

As for the Harris recurrent Markov chains, the loss in the exponential approximation is due to the size k^*_n of the blocks that tends to infinity with n. More precisely, we can fix:

- $k^*_n \approx -\ln(\overline{\sigma}^2_1(f_n))$ if $\delta_r = C\delta^r$ for $C > 0$ and $0 < \delta < 1$,

- $k^*_n \approx \overline{\sigma}^2_1(f_n)^{1/(1-\delta)}$ if $\delta_r = Cr^\delta$ for $C > 0$ and $\delta > 0$.

Thus the exponential approximation given by the classical Bernstein inequality (1.1) is no longer valid. It is an open question wether this loss may be reduced, as for uniformly mixing processes, see Samson [23], or not, as for Harris recurrent Markov Chains, see Adamczak [1].

4   Continuous functions of Bernoulli shifts and expanding maps

Let us focus on φ-weakly dependent examples that are not uniformly ϕ-mixing as, in the latter context, the classical Bernstein inequality (1.1) holds up to constants, see Samson [23]. Thus we present continuous functions of Bernoulli shifts that are not ϕ-mixing nor under **(A)** and expanding maps that are not ϕ-mixing under **(A)**. For other examples under **(A)**, we refer the reader to the Section 5 where an L^∞ coupling scheme and a sharper deviation inequality are given.

4.1   Continuous functions of Bernoulli shifts

Let us consider a ϕ-mixing stationary process (ξ_t) in some measurable space $\mathscr{Y}$ and the sequence (U_t) with value in $\mathscr{X}$ 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 [21], assume that there exists a non increasing sequence (ν_k) satisfying almost surely

$$d(F(\xi_j; j \in \mathbb{N}), F(\xi^k_j; j \in \mathbb{N})) \leq \nu_k,$$

where the sequence (ξ_t^k) satisfies $\xi_t = \xi_t^k$ for $0 \leq t \leq k$ and for $t > k$, $\xi_t^k = \xi_t'$. Finally set $X_t = H(U_t)$ for some measurable function $H : \mathscr{X} \to \mathscr{X}$ 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 ϕ-mixing nor under (**A**).

4.2 Expanding maps

Consider stationary expanding maps as in Collet *et al.* [5] where the authors prove a covariance inequality similar to (2.3). 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 [7] for more details.

5 Under (A) with a coupling scheme in L^∞.

Assume that the condition (**A**) holds: $\mathscr{X}$ is a metric Polish space and $\sup_{(x,y) \in \mathscr{X}^2} d(x, y) \leq 1$. We say that an L^∞ coupling scheme exists for $(X_1, \dots, X_n)$ when for any r, j 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 $\mathscr{M}_j$) and a sequence $(\delta_r')_{r \geq 1}$ satisfying the relation

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

5.1 A sharper deviation inequality

Remark that condition (5.1) with (δ_r') implies condition (3.1) with $\delta_r = \delta_r'$. Under condition (5.1), we can refine Eqn. (3.2):

Theorem 5.1. *If condition* (5.1) *is satisfied, for any* $f \in \mathscr{F}$ *such that* $\mathbb{E}(f(X_1)) = 0$, *any* $x \geq n\delta_k'$ *with* $1 \leq k \leq n$ *we have*

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

where $h(u) = (1+u)\ln(1+u) - u$ *for all* $u \geq 0$. *Then for any* $x \geq 0$

$$\mathbb{P}\left(S(f) \geq 2\sqrt{n\sigma_{k^{*\prime}}^2(f)x} + 1.34\, k^{*\prime}x\right) \leq \exp(-x) \tag{5.2}$$

with $k^{*\prime} = \min\{1 \leq k \leq n \mid n\delta_k' \leq kx\}$.

The proof of this Theorem is given in Subsection 6.4. In the theorem 5.1, the first deviation inequality is of Bennett's type. It refines the exponential approximation of Bernstein's type inequalities with a poisson approximation.

Remark 5.1. *To compare the two Bernstein's type inequalities* (3.2) *and* (5.2), *we compare the blocks sizes k^* and $k^{*'}$ involve only in the exponential approximation. As $k^{*'} = \min\{1 \le k \le n \ / \ k\delta_k' \le xk^2/n\}$, if $\delta_k' = \delta_k$ then $k^{*'} \le k^*$ as soon as $n\sigma_k^2(f) \le k^2 x$ or equivalently $\sqrt{n\sigma_k^2(f)x} \le kx$, i.e. as soon as x is in the domain of the exponential approximation. Thus the exponential approximation given in inequality* (5.2) *sharpens the one in* (3.2) *and as their normal approximations are similar, the latter inequality is always the sharpest.*

A tradeoff between the generality of the context and the sharpness of the deviation inequalities is done. However, for the deviation study of the risk of an estimator where $\sigma_k^2(f_n) \to 0$ for all k, the latter context is more convenient as the blocks size $k^{*'}$ is independent of $\sigma_k^2(f_n)$. The negative effects, due to the blocks size tending to infinity with n and described in the remarks of Subsection 3.3, are no longer valid here. Moreover, condition (5.1) is satisfied in many practical examples:

5.2   Bounded Markov Chains

Following Dedecker and Prieur [7], let us consider a stationary Markov chain (X_t) with its transition kernel P that is a κ-Lipschitz operator on Λ_1 (the set of 1-Lipschitz functions) with $\kappa < 1$. Then

$$r\delta_r' = \kappa^r(1 + \cdots + \kappa^{r-1}),$$

see [7] for more details.

5.3   Bounded chains with infinite memory

Let the sequence of the innovations (ξ_t) be an iid process on a measurable space $\mathscr{Y}$. We define (X_t) as the solution of the equation

$$X_t = F(X_{t-1}, X_{t-2}, \ldots; \xi_t) \qquad a.s., \tag{5.3}$$

for some bounded function $F : \mathscr{X}^{\mathbb{N}\backslash\{0\}} \times \mathscr{Y} \to \mathscr{X}$. Assume that F satisfies the condition

$$d(F((x_k)_{k\ge1}; \xi_0), F((y_k)_{k\ge1}; \xi_0)) \le \sum_{j=1}^{\infty} a_j(F)d(x_j, y_j), \ a.s. \tag{5.4}$$

for all $(x_k)_{k\ge1}, (y_k)_{k\ge1} \in \mathscr{X}^{\mathbb{N}\backslash\{0\}}$ such that there exists $N > 0$ as $x_k = y_k = 0$ for all $k > N$ with $a_j(F) \ge 0$ satisfying

$$\sum_{j=1}^{\infty} a_j(F) := a(F) < 1. \tag{5.5}$$

Let (ξ_t^*) be a stationary sequence distributed as $(\xi_t)_{t\in Z}$, independent of $(\xi_t)_{t\le0}$ and such that $\xi_t = \xi_t^*$ for all $t > 0$. Let $(X_t^*)_{t\in\mathbb{Z}}$ be the solution of the equation

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

Using similar arguments as in Doukhan and Wintenberger [12] we get

Lemma 5.2. *Under condition* (5.5) *there exists some bounded (by $1/2$) stationary process X solution of the equation* (5.3). *Moreover, this solution satisfies* (5.1) *with*

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

As the proof of this lemma is similar to the proof given in [12], 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 in econometrics.

5.4 Bernoulli shifts

Solutions of the recurrence equation (5.3) may always be written as $X_t = H((\xi_j)_{j \leq t})$ for some measurable function $H : \mathcal{Y}^{\mathbb{N}} \mapsto \mathcal{X}$. A coupling version (X_t^*) is thus given by $X_t^* = H((\xi_t^*))$ for all $t \in \mathbb{Z}$ where (ξ_t^*) is 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 $d(\xi_1, y)$ is bounded a.s. for some $y \in \mathcal{Y}$, then there exists $C > 0$ satisfying (5.1) with

$$r \delta_r' = C \sum_{i \geq r} a_r.$$

6 Proofs

6.1 Proof of the Theorem 3.1

To deal with the dependence, we first use the Bernstein's blocks technique as in [9] and then the proof ends with the Chernoff device as in the iid case. 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 $2p - 1 \leq k^{-1}n \leq 2p$. Denote S_1 and S_2 the sums of even and odd blocks defined as

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

We want to prove that for any $0 \leq t \leq 1$, choosing $k = [1/t] \wedge n$ as in [9] it holds:

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

From the Schwarz inequality, we get

$$\ln \mathbb{E}[\exp(tS(f))] \leq \frac{1}{2} \left(\ln \mathbb{E} \exp\left(2tS_1\right) + \ln \mathbb{E} \exp\left(2tS_2\right) \right). \tag{6.2}$$

Now let us treat in detail the term depending on S_1, the same argument applies identically to S_2. 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 the Holder inequality, for any $2 \leq m \leq p-1$ we get

$$\exp(L_{m+1}) - \exp(L_m)\exp(L_1)$$

$$\leq \exp(L_m) \left\| \mathbb{E}\left(\exp\left(2t \sum_{i \in I_{2(m+1)}} f(X_i) \right) \mid \mathcal{M}_{2mk} \right) - \mathbb{E}\left(\exp\left(2t \sum_{i \in I_{2(m+1)}} f(X_i) \right) \right) \right\|_\infty$$

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

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 $\mathcal{M}_{2mk}$. The definition of the coupling coefficients τ_∞ provides that

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

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

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

Collecting these inequalities, we achieve that

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

The classical Bennett inequality applied on $\sum_{i \in I_2} f(X_i)$ gives the estimate $\exp(L_1) \leq 1 + 4\sigma_k^2(f)(e^{kt} - kt - 1)/k$. 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 last step 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) from (6.2) 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 $e^{nt} - nt - 1 \leq (nt)^2 \sum_{j \geq 0} (nt)^j/(j+2)!$ and $(j+2)! \geq 2 \cdot 3^j$, we obtain the inequality $e^{nt} - nt - 1 \leq 2^{-1}(nt)^2 \sum_{j \geq 0} 3^{-j} \leq 3(nt)^2/4$ for $nt \leq 1$. Using it, we derive for any $t \leq n^{-1}$ that

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

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

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

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

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

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

Easy calculation yields for all $x \geq 0$

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

A rough bound $k^*t + k^{*-1}Kn\sigma_{k^*}^2(f) \leq 3k^*x/2$ for $k^{*2}x > 2Kn\overline{\sigma}_{k^*}^2(f)$ then leads to the result of the Theorem 3.1.

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

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

and as $|\mathbb{E}(f(X_{r+1}) - f(X_{r+1})^* | \mathcal{M}_1)| \leq \delta_r$ by ineq. (2.4), we get the desired result.

6.3 Proof of Proposition 4.1

We sketch the proof of [21]. We are interested in estimating the coefficients $\varphi(\mathcal{M}_j, (X_{r+j}, \ldots, X_{2r-1+j}))$ 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. Let $U_t^k = F(\xi_{t-j}^k; j \in \mathbb{N})$ and $X_t^k = H(U_t^k)$. For any $f \in \mathcal{F}$, we have

$$f(X_{r+j}, \ldots, X_{2r-1+j}) - f(X_{r+j}^k, \ldots, X_{2r-1+j}^k) \leq \left(\frac{1}{r}\sum_{i=r+j}^{2r-1+j} d(X_i, X_i^k)\right) \wedge 1. \qquad (6.3)$$

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

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

Noting 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, we get, from the definition of the ϕ-mixing coefficients,

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

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 $1 \leq k \leq r - 1$:

$$\left\| \mathbb{E} \left(f(X_{r+j}, \ldots, X_{2r-1+j}) - f(X_{r+j}^k, \ldots, X_{2r-1+j}^k) \middle/ \mathcal{M}_j \right) \right\|_\infty \leq \phi_{r-k} + \mathbb{E}(2w_H(U_0, 2v_k)) \wedge 1. \quad (6.4)$$

Using again the definition of the ϕ-mixing coefficients we get, since f is bounded by 1,

$$\left\| \mathbb{E} \left(f(X_{r+j}^k, \ldots, X_{2r-1+j}^k) \middle/ \mathcal{M}_j \right) - \mathbb{E} \left(f(X_{r+j}^k, \ldots, X_{2r-1+j}^k) \right) \right\|_\infty \leq \phi_{r-k}. \quad (6.5)$$

Finally, using again (6.3) and that $d(X_i, X_i^k) \leq w_H(U_i, v_k)$, by stationarity of (U_t) we obtain

$$\mathbb{E}f(X_{r+j}, \ldots, X_{2r-1+j}) - \mathbb{E}f(X_{r+j}^k, \ldots, X_{2r-1+j}^k) \leq \mathbb{E}(w_H(U_0, v_k)) \wedge 1. \quad (6.6)$$

The result of the Proposition 4.1 follow from the definition of the φ-coefficients, the inequalities (6.4), (6.5) and (6.6).

6.4   Proof of Theorem 5.1

Let us keep the same notation as in the proof of Theorem 3.1. The Bennett's type deviation inequality follows classically from the Chernoff device together with the estimate:

$$\ln(\mathbb{E}(\exp(tS(f)))) \leq \frac{2n\sigma_k^2(f)}{k^2}(\exp(kt) - kt - 1) + n\delta_k' t \quad \text{for all } t \geq 0. \quad (6.7)$$

To prove (6.7), let us use the L^∞-coupling scheme and (5.1) to derive that, for any $1 \leq m \leq p$:

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

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 any $t \geq 0$ 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 stationarity we deduce that

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

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

$$\mathbb{E}\exp(2tS_1) = \mathbb{E}\left(\exp\left(2t \sum_{1 \leq m \leq p-1} \sum_{i \in I_{2m}} f(X_i) \right) \mathbb{E}\left(\exp\left(\sum_{i \in I_{2p}} f(X_i) \right) \middle| \mathcal{M}_{2(p-1)} \right) \right)$$

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

Using the same argument recursively on $m = p-1,\ldots,2$ we then get that

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

Now 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 (6.7) follows remarking that $4pk^{-1} \leq 2nk^{-2}$ and $2(p-1)k \leq n$.

To prove the Bernstein's type inequality, from (6.7), the series expansion of the function $\exp(x)-x-1$ and the fact that $j! \geq 2\cdot 3^{j-2}$ for all $j \geq 2$ we first obtain that

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

With the same notation than in [17], 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

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. MR2424985

[2] BENNETT, G. Probability inequalities for the sum of independant random variables. *Journal of the American Statistician Association 57* (1962), 33–45.

[3] BERTAIL, P., AND CLÉMENÇON, S. Sharp bounds for the tail of functionals of Markov chains. *Theory Probab. Appl. 54* (2010) 505–515.

[4] CATONI, O. *Statistical Learning Theory and Stochastic Optimization, Lecture Notes in Mathematics (Saint-Flour Summer School on Probability Theory 2001)*. Springer, 2004. MR2163920

[5] 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. MR1918536

[6] 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. MR2338725

[7] DEDECKER, J., AND PRIEUR, C. New dependence coefficients: Examples and applications to statistics. *Probability Theory and Related Fields 132* (2005), 203–235. MR2199291

[8] 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 *Lectures Notes in Statistics*, Springer, pp. 105–121. MR2283252

[9] DOUKHAN, P. *Mixing*, vol. 85 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1994. MR1312160

[10] DOUKHAN, P., AND LOUHICHI, S. A new weak dependence condition and applications to moment inequalities. *Stochastic Process. Appl. 84*, 2 (1999), 313–342. MR1719345

[11] DOUKHAN, P., AND NEUMANN, M. A Bernstein type inequality for times series. *Stoch. Proc. Appl. 117-7* (2007), 878–903.

[12] DOUKHAN, P., AND WINTENBERGER, O. Weakly dependent chains with infinite memory. *Stochastic Process. Appl. 118* (2008), 1997–2013. MR2462284

[13] IBRAGIMOV, I. Some limit theorems for stationary processes. *Theory of Probab. Appl. 7* (1962), 349–382. MR0148125

[14] IBRAGIMOV, I. A., AND LINNIK, Y. V. *Independent and stationary sequences of random variables*. Wolters-Noordhoff Publishing, 1971. MR0322926

[15] JOULIN, A., AND OLLIVIER, Y. Curvature, concentration, and error estimates for Markov chain Monte Carlo. *Ann. Probab. 38* (2010), 2418–2442.

[16] LEZAUD, P. Chernoff-type bound for finite Markov chains. *Ann. Appl. Probab. 8* (1998), 849–867. MR1627795

[17] MASSART, P. *Concentration Inequalities and Model Selection, Lecture Notes in Mathematics (Saint-Flour Summer School on Probability Theory 2003)*. Springer, 2006. MR2319879

[18] MERLEVÈDE, F., PELIGRAD, M., AND RIO, E. Bernstein inequality and moderate deviations under strong mixing conditions. In *High Dimensional Probability V: The Luminy Volume* (Beachwood, Ohio, USA: IMS, 2009) C. Houdré, V. Koltchinskii, D.M. Mason and M. Peligrad, Eds., 273–292..

[19] MERLEVÈDE, F., PELIGRAD, M., AND RIO, E. A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probab. Theory Related Fields* (2010) DOI: 10.1007/s00440-010-0304-9.

[20] NUMMELIN, E. A splitting technique for Harris recurrent Markov chains. *Z. Wahrsch. Verw. Gebiete 43* (1978), 309–318. MR0501353

[21] RIO, E. The Berry-Esseen theorem for weakly dependent sequences. *Probab. Theory Related Fields 104* (1996), 255–282. MR1373378

[22] RIO, E. Inégalités de Hoeffding pour les fonctions lipschitziennes de suites dépendantes. *Comptes Rendus de l'Acamédie des Sciences de Paris, Série I 330* (2000), 905–908. MR1771956

[23] SAMSON, P.-M. Concentration of measure inequalities for Markov chains and ϕ-mixing processes. *Ann. Probab. 28* (2000), 416–461. MR1756011

[24] VIENNET, G. Inequalities for absolutely regular sequences: application to density estimation. *Probab. Theory Relat. Fields 107* (1997), 467-492. MR1440142