IMS Collections
High Dimensional Probability V: The Luminy Volume
Vol. 5 (2009) 273–292

DOI: 10.1214/09-IMSCOLL518

Bernstein inequality and moderate deviations under strong mixing conditions

Florence Merlevède[1] **, Magda Peligrad**[2,*] **and Emmanuel Rio**[3,†]

Université Paris Est, University of Cincinnati, Université de Versailles

Abstract: In this paper we obtain a Bernstein type inequality for a class of weakly dependent and bounded random variables. The proofs lead to a moderate deviations principle for sums of bounded random variables with exponential decay of the strong mixing coefficients that complements the large deviation result obtained by Bryc and Dembo (1998) under superexponential mixing rates.

1. Introduction

This paper has double scope. First we obtain a Bernstein's type bound on the tail probabilities of the partial sums S_n of a sequence of dependent and bounded random variables $(X_k, k \geq 1)$. Then we use the developed techniques to study the moderate deviations principle.

We recall the definition of strongly mixing sequences, introduced by Rosenblatt [19]: For any two σ-algebras $\mathcal{A}$ and $\mathcal{B}$, we define the α-mixing coefficient by

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

Let $(X_k, k \geq 1)$ be a sequence of real-valued random variables defined on $(\Omega, \mathcal{A}, \mathbb{P})$. This sequence will be called strongly mixing if

$$\alpha(n) := \sup_{k \geq 1} \alpha\left(\mathcal{M}_k, \mathcal{G}_{k+n}\right) \to 0 \text{ as } n \to \infty, \tag{1.1}$$

where $\mathcal{M}_j := \sigma(X_i, i \leq j)$ and $\mathcal{G}_j := \sigma(X_i, i \geq j)$ for $j \geq 1$.

Alternatively (see Bradley [5], Theorem 4.4)

$$4\alpha(n) := \sup\{\mathrm{Cov}(f, g)/||f||_\infty ||g||_\infty;\ f \in \mathbb{L}_\infty(\mathcal{M}_k),\ g \in \mathbb{L}_\infty(\mathcal{G}_{k+n})\}. \tag{1.2}$$

Establishing exponential inequalities for strongly mixing sequences is a very challenging problem. Some steps in this direction are results by Rio [18, Theorem 6.1],

*Supported in part by a Charles Phelps Taft Memorial Fund grant and NSA grant, H98230-07-1-0016 and H98230-09-1-0005.

†Supported in part by Centre INRIA Bordeaux Sud-Ouest & Institut de Mathématiques de Bordeaux

[1]Université Paris-Est-Marne-la-Vallée, LAMA and CNRS UMR 8050, e-mail: florence.merlevede@univ-mlv.fr

[2]Department of Mathematical Sciences, University of Cincinnati, e-mail: magda.peligrad@uc.edu

[3]Université de Versailles, Laboratoire de mathématiques, UMR 8100 CNRS, e-mail: rio@math.uvsq.fr

Keywords and phrases: deviation inequality, moderate deviations principle, weakly dependent sequences, strong mixing.

AMS 2000 subject classifications: 60E15, 60F10, 62G07.

who obtained a Fuk-Nagaev type inequality, by Dedecker and Prieur [10] who extended Theorem 6.1 in Rio [18] using coupling coefficients, by Doukhan and Neumann [13] who used combinatorics techniques. In a recent paper Merlevède, Peligrad and Rio [17] get exponential bounds for subexponential mixing rates, when the variables are not necessarily bounded, obtaining the same order of magnitude as in the independent case. More precisely they show that, if $\alpha(n) \leq \exp(-cn^{\gamma_1})$ and $\sup_{i>0} \mathbb{P}(|X_i| > t) \leq \exp(1 - t^{\gamma_2})$ with $\gamma_1 > 0$ and $\gamma_2 > 0$, such that $(1/\gamma_1) + (1/\gamma_2) = 1/\gamma > 1$, then there are positive constants C, C_1, C_2 and η depending only on c, γ_1 and γ_2, such that for all $n \geq 4$ and $\lambda \geq C(\log n)^\eta$

$$\mathbb{P}(|S_n| \geq \lambda) \leq (n+1)\exp(-\lambda^\gamma/C_1) + \exp(-\lambda^2/nC_2).$$

Here $S_n = \sum_{k=1}^n X_k$. The case not covered by that paper is the case of exponential mixing rates and bounded variables, that is $\gamma_1 = 1$ and $\gamma_2 = \infty$. The aim of this paper is to study this case, to point out several new recent techniques and ideas and to comment on the order of magnitude of the probabilities of large deviations. Our proofs will be based on estimations of the Laplace transform.

One of our results is that for a strongly mixing sequence of centered and bounded random variables satisfying, for a certain $c > 0$,

$$\alpha(n) \leq \exp(-2cn), \tag{1.3}$$

we can find two constants c_1 and c_2 depending only on c and on the uniform bound of the random variables, such that, for all $x > 0$,

$$\mathbb{P}(|S_n| > x) \leq \exp(-c_1 x^2/n) + \exp(-c_2 x/(\log n)(\log\log n)).$$

Then, we use this exponential inequality and the techniques that lead to this result to obtain moderate deviations asymptotic results, improving Proposition 2.4 in Merlevède and Peligrad [16]. Our results show that we can come close up to a logarithmic term to the moderate deviations asymptotics for independent random variables. Of course a kind of correction is needed since the traditional large deviations results do not hold for geometrically strongly mixing sequences. As a matter of fact the large deviations principle does not hold even in the context of uniformly mixing sequences with exponential rates. See Bryc and Dembo [6], Example 1, Proposition 5 and Example 2, that point out examples of empirical processes of Doeblin recurrent Markov chains, that are therefore ϕ-mixing with exponential mixing rate and that do not satisfy the large deviations principle.

We mention that strongly mixing sequences form a larger class than absolutely regular sequences. The problem of the moderate deviations principle was studied for absolutely regular Markov chains with exponential rates in de Acosta [9] and also in Chen and de Acosta [8], when the transition probabilities are stationary and there is a certain restriction on the class of initial distributions. A class of processes satisfying a splitting condition closely related to absolutely regular processes was considered by Tsirelson [21]. Recently, Dedecker, Merlevède, Peligrad and Utev [11] considered projective conditions with applications to ϕ-mixing.

Notice that we do not require any degree of stationarity for obtaining Bernstein inequality except for a uniform bound for the variables.

The strong mixing coefficient used in this paper can be generalized by using smaller classes of functions than those used in Definition (1.2) to include even more examples. Such examples include function of linear processes with absolutely regular innovations and Arch models. In this paper, we give an application to the moderate deviations principle for Kernel estimators of the common marginal density of a certain class of continuous time processes.

For the clarity of the proofs it will be more convenient to embed the initial sequence into a continuous time process; namely, $(X_t, t \geq 0)$ is defined from the original sequence $(X_n, n \geq 1)$ by $X_t = X_{[t+1]}$. For a Borel set A, define

$$S_A = \int_A X_t \, dt. \tag{1.4}$$

Then $S_{[0,n]} = \sum_{k=1}^{n} X_k$. The strong mixing coefficient of $(X_t, t \geq 0)$ is defined as

$$\tilde{\alpha}(u) := \sup_{t \geq 0} \alpha\left(\mathcal{M}_t, \mathcal{G}_{t+u}\right) \to 0 \text{ as } n \to \infty,$$

where $\mathcal{M}_t := \sigma(X_v, v \leq t)$ and $\mathcal{G}_w := \sigma(X_v, v \geq w)$.

Notice that, since $\tilde{\alpha}(u) \leq \alpha([u])$, if $(X_n, n \geq 1)$ satisfies (1.3), the continuous type mixing coefficients still satisfy a geometrically mixing condition; namely for any $u \geq 2$,

$$\tilde{\alpha}(u) \leq \exp(-cu). \tag{1.5}$$

In the rest of the paper, $\|Y\|_\infty$ stands for the essential supremum of a random variable Y.

2. Results

Our first result is the following exponential inequality:

Theorem 1. *Let $(X_j)_{j \geq 1}$ be a sequence of centered real-valued random variables. Suppose that the sequence satisfies (1.3) and that there exists a positive M such that $\sup_{i \geq 1} \|X_i\|_\infty \leq M$. Then there are positive constants C_1 and C_2 depending only on c such that for all $n \geq 4$ and t satisfying $0 < t < \frac{1}{C_1 M (\log n)(\log\log n)}$, we have*

$$\log \mathbb{E}\big(\exp(tS_n)\big) \leq \frac{C_2 t^2 n M^2}{1 - C_1 t M (\log n)(\log\log n)}.$$

In terms of probabilities, there is a constant C_3 depending on c such that for all $n \geq 4$ and $x \geq 0$

$$\mathbb{P}(|S_n| \geq x) \leq \exp\bigg(-\frac{C_3 x^2}{nM^2 + Mx(\log n)(\log\log n)}\bigg). \tag{2.1}$$

As a counterpart, the following Bernstein type inequality holds.

Theorem 2. *Under conditions of Theorem 1, there are positive constants C_1 and C_2 depending only on c such that for all $n \geq 2$ and any positive t such that $t < \frac{1}{C_1 M (\log n)^2}$, the following inequality holds:*

$$\log \mathbb{E}\big(\exp(tS_n)\big) \leq \frac{C_2 t^2 (nv^2 + M^2)}{1 - C_1 t M (\log n)^2},$$

where v^2 is defined by

$$v^2 = \sup_{i > 0}\Big(\mathrm{Var}(X_i) + 2\sum_{j > i} |\mathrm{Cov}(X_i, X_j)|\Big). \tag{2.2}$$

In terms of probabilities, there is a constant C_3 depending only on c such that for all $n \geq 2$,

$$\mathbb{P}(|S_n| \geq x) \leq \exp\Big(-\frac{C_3 x^2}{v^2 n + M^2 + xM(\log n)^2}\Big). \tag{2.3}$$

To compare these two results, we notice that the coefficient of x in the inequality (2.1) has a smaller order than the corresponding in (2.3). However, the term $v^2 n$ can be considerably smaller than nM^2, which is an advantage in some applications when the variables are not uniformly bounded. Notice also that if stationarity is assumed, v^2 can be taken as

$$v^2 = \mathrm{Var}(X_1) + 4\sum\nolimits_{i\geq 1} \mathbb{E}X_1^2 I(|X_1| \geq Q(2\alpha_i)),$$

where $Q(u) = \inf\{t > 0, \mathbb{P}(|X_1| > t) \leq u\}$ for u in $]0,1]$.

In the context of bounded functions f of stationary geometrically strongly mixing Markov chains, Theorem 6 in Adamczak [1] provides a Bernstein type inequality for $S_n(f) = f(X_1) + \cdots + f(X_n)$ with the factor $\log n$ instead of $(\log n)^2$, which appears in (2.3). To be more precise, under the centering condition $\mathbb{E}(f(X_1)) = 0$, he proves that

$$\mathbb{P}(|S_n(f)| \geq x) \leq C \exp\Big(-\frac{1}{C}\min\Big(\frac{x^2}{n\sigma^2}, \frac{x}{\log n}\Big)\Big),$$

where $\sigma^2 = \lim_n n^{-1}\mathrm{Var} S_n(f)$.

The two previous results are useful to study the moderate deviations principle (MDP) for the partial sums of the underlying sequences. In our terminology the moderate deviations principle (MDP) signifies the following type of behavior.

Definition 3. We say that the *MDP* holds for a sequence $(T_n)_n$ of random variables with the speed $a_n \to 0$ and rate function $I(t)$ if for each Borel set A,

$$\begin{aligned} -\inf_{t\in A^o} I(t) &\leq \liminf_n a_n \log \mathbb{P}(\sqrt{a_n} T_n \in A) \\ &\leq \limsup_n a_n \log \mathbb{P}(\sqrt{a_n} T_n \in A) \leq -\inf_{t\in \bar{A}} I(t) \end{aligned} \tag{2.4}$$

where $\bar{A}$ denotes the closure of A and A^o the interior of A.

Notice that the moderate deviations principle for $(S_n/\sqrt{n})$ is an intermediate behavior between CLT, $\mathbb{P}(S_n/\sqrt{n} \in A)$ and large deviation, $\mathbf{P}(S_n/n \in A)$. Our moderate deviations results are the following:

Theorem 4. *Let $(X_j)_{j\geq 1}$ be a sequence of centered real valued random variables satisfying the assumptions of Theorem 1. Let $S_n = \sum_{i=1}^n X_i$, $\sigma_n^2 = \mathrm{Var}(S_n)$ and assume in addition that* $\liminf_{n\to\infty} \sigma_n^2/n > 0$. *Then for all positive sequences a_n with*

$$a_n \to 0 \textit{ and } \frac{n a_n}{(\log n)^2 (\log\log n)^2} \to \infty \tag{2.5}$$

the sequence $(\sigma_n^{-1} S_n)_{n\geq 1}$ satisfies (2.4) with the good rate function $I(t) = t^2/2$.

If we assume that the sequence is $\mathbb{L}_2$-stationary, then by Lemma 1 in Bradley [4], we get the following corollary:

Corollary 5. *Let $(X_j)_{j\geq 1}$ be as in Theorem 4. Suppose in addition that the sequence is $\mathbb{L}_2-$stationary and $\sigma_n^2 \to \infty$. Then, $\lim_{n\to\infty} \sigma_n^2/n = \sigma^2 > 0$ and for all positive sequences a_n satisfying (2.5), $(n^{-1/2}S_n)_{n\geq 1}$ satisfies (2.4) with the good rate function $I(t) = t^2/(2\sigma^2)$.*

In the next result, we derive conditions ensuring that the MDP holds for the partial sums of triangular arrays of strongly mixing sequences. For a double indexed sequence $(X_{j,n}, j \geq 1)_{n\geq 1}$ of real valued random variables, we define for any $k \geq 0$,

$$\alpha_n(k) = \sup_{j\geq 1} \alpha(\sigma(X_{n,i}, i \leq j), \sigma(X_{n,i}, i \geq k+j)). \tag{2.6}$$

Theorem 6. *For all $n \geq 1$, let $(X_{j,n}, j \geq 1)_{n\geq 1}$ be a double indexed sequence of centered real valued random variables such that for every $j \geq 1$ and every $n \geq 1$, $\|X_{j,n}\|_\infty \leq M_n$ where M_n is a positive number. For all $n \geq 1$ and all $k \geq 0$, let $\alpha_n(k)$ be defined by (2.6) and assume that $\alpha(k) = \sup_{n\geq 1} \alpha_n(k)$ satisfies (1.3). Define v^2 by*

$$v^2 = \sup_{n\geq 1} \sup_{i>0} \Big(\mathrm{Var}(X_{i,n}) + 2 \sum_{j>i} |\mathrm{Cov}(X_{i,n}, X_{j,n})| \Big). \tag{2.7}$$

and suppose $v^2 < \infty$. Let $S_n = \sum_{i=1}^n X_{i,n}$, $\sigma_n^2 = \mathrm{Var}(S_n)$ and assume in addition that $\liminf_{n\to\infty} \sigma_n^2/n > 0$. Then for all positive sequences a_n with

$$a_n \to 0 \text{ and } na_n/M_n^2(\log n)^4 \to \infty \tag{2.8}$$

the sequence $(\sigma_n^{-1}S_n)_{n\geq 1}$ satisfies (2.4) with the good rate function $I(t) = t^2/2$.

3. Discussion and Examples

1. The first comment is on Theorems 1 and 2. Notice that compared to the traditional Bernstein inequality for independent random variables there is a logarithmic correction in the linear term in x appearing in the inequalities (2.1) and (2.3). We includ in our paper another bound. Corollary 12 gives better results than the other exponential bound results in the large deviation range, that is when x is close to n. As a matter of fact the tail probability $\mathbb{P}(|S_n| \geq x)$ can be bounded with the minimum of the right hand sides of inequalities (2.1), (2.3) and (4.16). Among these inequalities, (2.1) provides the best condition leading to a moderate deviations principle when the random variables are uniformly bounded.

2. The strong mixing coefficients are not used in all their strength. For obtaining our Bernstein type inequalities we can considerably restrict the class of functions used to define the strong mixing coefficients to those functions that are coordinatewise nondecreasing, and one sided relations. Assume that for any index sets Q and Q^* (sets of natural numbers) such that $Q \subset (0, p]$ and $Q^* \subset [n + p, \infty)$, where n and p are arbitrary integers, there exists a decreasing sequence $\alpha^*(n)$ such that

$$\mathrm{Cov}(f(S_Q), g(S_{Q^*})) \leq \alpha^*(n)||f(S_Q)||_\infty||f(S_{Q^*})||_\infty,$$

where f and g are bounded functions coordinatewise nondecreasing. Here $S_Q = \sum_{i\in Q} X_i$. Clearly the families of functions $\exp(\sum_{i\in Q} tx_i)$ are coordinatewise nondecreasing for $t > 0$ and then, for bounded random variables we have for all $t > 0$

$$\mathrm{Cov}(\exp(tS_Q), \exp(tS_{Q^*})) \leq \alpha^*(n)|| \exp(tS_Q)||_\infty|| \exp(tS_{Q^*}))||_\infty. \tag{3.1}$$

Also by using the functions $f(x) = g(x) = x$

$$\operatorname{Cov}(X_j, X_{j+n}) \leq \alpha^*(n)||X_j||_\infty ||X_{j+n}||_\infty. \tag{3.2}$$

As a matter of fact these are the only functions we use in the proof of our Bernstein inequality. So if $\alpha^*(n)$ decreases geometrically our results still hold. Inequality (3.1) is used to bound the Laplace transform of partial sums, and Inequality (3.2) is used to bound their variance. Since both of these inequalities (3.1) and (3.2) are stable under convolution we can obtain Bernstein type inequality for example for sequences of the type $X_n = Y_n + Z_n$, where Y_n is strongly mixing as in Theorem 1 and is Z_n a noise, independent on Y_n, negatively associated, such as a truncated Gaussian sequence with negative correlations.

We point out that similar results can be obtained by using alternative mixing coefficients such as the τ-mixing coefficient introduced by Dedecker and Prieur [10]. Consequently we can treat all the examples in Merlevède, Peligrad and Rio [17], namely: instantaneous functions of absolutely regular processes, functions of linear processes with absolutely regular innovations and ARCH(∞) models.

3. We now give an application to the moderate deviations principle behavior for kernel estimators of the density of a continuous time process.

Let $X = (X_t, t \geq 0)$ be a real valued continuous time process with an unknown common marginal density f. We wish to estimate f from the data $(X_t, 0 \leq t \leq T)$. In what follows, we will call a kernel a function K from $\mathbb{R}$ to $\mathbb{R}$ which is a bounded continuous symmetric density with respect to Lebesgue measure and such that

$$\lim_{|u|\to\infty} uK(u) = 0, \text{ and } \int_{\mathbb{R}} u^2 K(u)du < \infty.$$

The kernel density estimator is defined as

$$f_T(x) = \frac{1}{Th_T} \int_0^T K\Big(\frac{x - X_t}{h_T}\Big)dt,$$

where $h_T \to 0^+$ and K is a kernel. In order to derive sufficient conditions ensuring that the MDP holds for the sequence $\sqrt{T}(f_T(x) - \mathbb{E}f_T(x))$, we assume that there exists a constant $c > 0$ such that for any $u \geq 0$,

$$\alpha_u = \sup_{t\geq 0} \alpha(\sigma(X_s, s \leq t), \sigma(X_s, i \geq u + t)) \leq e^{-cu}. \tag{3.3}$$

In addition, we assume that the joint distribution f_{X_s,X_t} between X_s and X_t exists and that $f_{X_s,X_t} = f_{X_0,X_{|t-s|}}$. Applying Theorem 6, we obtain the following result:

Corollary 7. *Suppose that $g_u = f_{X_0,X_u} - f \otimes f$ exists for $u \neq 0$, and that the function $u \mapsto \sup_{x,y} |g_u(x, y)|$ is integrable on $]0, \infty[$ and $g_u(\cdot, \cdot)$ is continuous at (x, x) for each $u > 0$. In addition assume that the strong mixing coefficients of the process satisfy (3.3). Then for all positive sequences a_T with*

$$a_T \to 0, \ \frac{a_{[T]}}{a_T} \to 1, \ \textit{and} \ \frac{a_T T h_T^2}{(\log T)^4} \to \infty \ ,$$

the sequence $\sqrt{T}(f_T(x) - \mathbb{E}f_T(x))$ satisfies (2.4) with speed a_T and the good rate function

$$I(t) = t^2 \Big/ \Big(4 \int_0^\infty g_u(x, x)du\Big)^{-1}. \tag{3.4}$$

Furthermore if f is differentiable and such that f' is l-Lipschitz for a positive constant l, and if $a_T T h_T^4 \to 0$, then the sequence $\sqrt{T}(f_T(x) - f(x))$ satisfies (2.4) with speed a_T and the good rate function defined by (3.4).

Some examples of diffusion processes satisfying Condition (3.3) may be found in Veretennikov [20] (see also Leblanc, [15]).

4. Proofs

First let us comment on the variance of partial sums. By using the notation (1.4), for any compact set K_A included in $[a, a+A]$ where $A > 0$ and $a \geq 0$, we have that

$$\operatorname{Var}(S_{K_A}) \leq A \sup_{i>0} \Big(\operatorname{Var}(X_i) + 2 \sum_{j>i} |\operatorname{Cov}(X_i, X_j)| \Big).$$

If the variables are bounded by M, then by using the definition (1.2), we get that

$$\operatorname{Var}(S_{K_A}) \leq A \Big(1 + 8 \sum_{i \geq 1} \alpha_i \Big) M^2 \leq KAM^2.$$

If some degrees of stationarity are available we can have better upper bounds. For instance if $\mathbb{P}(|X_n| > x) \leq \mathbb{P}(|X_0| > x)$, then by Theorem 1.1 in Rio [18],

$$\operatorname{Var}(S_{K_A}) \leq A \Big(\operatorname{Var}(X_0) + 4 \sum_{i \geq 1} \mathbb{E} X_0^2 I(|X_0| \geq Q(2\alpha_i)) \Big),$$

where $Q(u) = \inf\{t > 0, \mathbb{P}(|X_0| > t) \leq u\}$ for u in $]0, 1]$.

4.1. Preliminary lemmas

The first step is to prove an upper bound on the Laplace transform, valid for small values of t. Without restricting the generality it is more convenient to embed the index set into continuous time. In the following we shall use the notation (1.4).

Lemma 8. *Let $(X_n)_{n \geq 1}$ be as in Theorem 1. Let $B \geq 2$ and $a \geq 0$. Then for any subset K_B of $(a, a+B]$ which is a finite union of intervals, and for any positive t with $tM \leq (\frac{1}{2}) \wedge (\frac{c}{2B})^{1/2}$, we have*

$$\log \mathbb{E} \exp(t S_{K_B}) \leq B \Big(6.2 t^2 v^2 + \frac{Mt}{2} \exp\Big(-\frac{c}{2tM} \Big) \Big), \tag{4.1}$$

where v^2 is defined by (2.2).

Remark 9. Notice that under our conditions $v^2 \leq KM^2$ where $K = 1 + 8 \sum_{i \geq 1} \alpha_i$.

Proof of Lemma 8. If $tM \leq 4/B$, then $tS_{K_B} \leq 4$, which ensures that

$$\exp(tS_{K_B}) \leq 1 + tS_{K_B} + \frac{e^4 - 5}{16} t^2 (S_{K_B})^2,$$

since the function $x \mapsto x^{-2}(e^x - x - 1)$ is increasing. Now $(e^4 - 5)/16 \leq 3.1$. Hence

$$\mathbb{E} \exp(tS_{K_B}) \leq 1 + 3.1 B v^2 t^2, \tag{4.2}$$

which implies Lemma 8 by taking into account that $\log(1+x) \le x$.

If $tM > 4/B$, it will be convenient to apply Lemma 15 in Appendix, to get the result. Let p be a positive real to be chosen later on. Let $k = [B/2p]$, square brackets denoting the integer part. We divide the interval $(a, a+B]$ into $2k$ consecutive intervals of equal size $B/(2k)$. Denote these subintervals by $\{I_j; 1 \le j \le 2k\}$ and let

$$\tilde{S}_1 = \sum_{j=1}^{k} S_{K_B \cap I_{2j-1}} \text{ and } \tilde{S}_2 = \sum_{j=1}^{k} S_{K_B \cap I_{2j}}.$$

By the Cauchy-Schwarz inequality,

$$2 \log \mathbb{E} \exp(tS_{K_B}) \le \log \mathbb{E}(2t\tilde{S}_1) + \log \mathbb{E} \exp(2t\tilde{S}_2). \tag{4.3}$$

Now let $p = 1/(tM)$. Since $(X_n)_{n \ge 1}$ satisfies Condition (1.3), and since $B/(2k) \ge p \ge 2$, by applying Lemma 15 in Appendix, we obtain

$$\mathbb{E} \exp(2t\tilde{S}_2) \le k \exp\Big(\frac{MBt}{2} - \frac{cB}{2k}\Big) + \prod_{j=1}^{k} \mathbb{E} \exp(2tS_{K_B \cap I_{2j}}).$$

Notice that we are in the case $tM > 4/B$ implying that $p \le B/4$ and then $k \ge 2$. Now, under the assumptions of Lemma 8, we have $tM \le (c/(2B))^{1/2}$ which ensures that

$$MBt - \frac{cB}{2k} \le MBt - cp \le MBt - \frac{c}{tM} \le -\frac{c}{2Mt}.$$

Therefore,

$$\mathbb{E}(\exp(2t\tilde{S}_2)) \le \frac{BMt}{2} \exp(-c/(2tM)) + \prod_{j=1}^{k} \mathbb{E} \exp(2tS_{K_B \cap I_{2j}}).$$

Since the random variables $(X_i)_{i \ge 1}$ are centered, the Laplace transforms of $\tilde{S}_2$ and each of $(S_{K_B \cap I_{2j}})_{j \ge 1}$ are greater than one. Hence applying the inequality

$$|\log x - \log y| \le |x - y| \text{ for } x \ge 1 \text{ and } y \ge 1, \tag{4.4}$$

we derive that

$$\log \mathbb{E}(\exp(2t\tilde{S}_2)) \le \sum_{j=1}^{k} \log \mathbb{E} \exp(2tS_{K_B \cap I_{2j}}) + \frac{BMt}{2} \exp(-c/(2tM)).$$

Next $|2tS_{K_B \cap I_{2j}}| \le 2tMB/(2k)$. Since $p \le B/4$, $k \ge B/(4p)$ implying that $|2tS_{K_B \cap I_{2j}}| \le 4$, and consequently we may repeat the arguments of the proof of (4.2), so that

$$\sum_{j=1}^{k} \log \mathbb{E} \exp(2tS_{K_B \cap I_{2j}}) \le 6.2Bt^2v^2.$$

It follows that

$$\log \mathbb{E} \exp(2t\tilde{S}_2) \le 6.2Bt^2v^2 + (BMt/2) \exp(-c/(2tM)).$$

Clearly the same inequality holds true for the log-Laplace transform of $\tilde{S}_1$ which, together with relation (4.3), gives the result. $\square$

The key lemma for proving our theorems is a new factorization lemma. Its proof combines the ideas of Bernstein big and small type argument with a twist, diadic recurrence and Cantor set construction.

Lemma 10. *Let* $(X_i)_{i\geq 1}$ *be as in Theorem 1. Then, for every* $A \geq 2(c \vee 10)$ *there exists a subset* K_A *of* $[0, A]$*, with Lebesgue measure larger than* $A/2$ *(not depending on the random process) such that for all* t*,* $0 \leq tM \leq c_0/(\log A)$ *where* $c_0 = \frac{c}{8} \wedge \sqrt{\frac{c \log 2}{8}}$

$$\log(\mathbb{E}\exp(tS_{K_A})) \leq 6.2v^2t^2A + (c+1)A^{-1}\exp(-c/4tM). \tag{4.5}$$

where v^2 *is defined by (2.2). Moreover, if* $A \geq 4 \vee (2c)$ *for all* $0 \leq tM < \frac{c\wedge 1}{2}$*, we can find a constant* C *depending only* c *such that*

$$\log(\mathbb{E}\exp(tS_{(0,A]}) \leq Ct^2M^2A\log A. \tag{4.6}$$

Proof of Lemma 10. The proof is inspired by the construction of a "Cantor set" and has several steps.

Step 1. A "Cantor set" construction. Let A be a strictly positive real number strictly more than one. Let $\delta \in (0, 1)$ that will be selected later, and let k_A be the largest integer k such that $((1-\delta)/2)^k \geq 1/A$. We divide the interval $[0, A]$ in three parts and delete the middle one of size $A\delta$. The remaining ordered sets are denoted $K_{1,1}$, $K_{1,2}$ and each has the Lebesgue measure $A(1-\delta)/2$. We repeat the procedure. Each of the remaining two intervals $K_{1,1}$, $K_{1,2}$, are divided in three parts and the central one of length $A\delta(1-\delta)/2$ is deleted. After j steps ($j \leq k_A$), we are left with a disjoint union of 2^j intervals denoted by $K_{j,i}$, $1 \leq i \leq 2^j$, each of size $A((1-\delta)/2)^j$ and we deleted a total length $\sum_{i=0}^{j-1} A\delta(1-\delta)^i = A(1-(1-\delta)^j)$. We use the first index of sets $K_{j,i}$ to denote the step, and second one to denote its order. Set $k = k_A$ when no confusion is allowed, and define

$$K_A = \bigcup_{i=1}^{2^k} K_{k,i}. \tag{4.7}$$

We shall use also the following notation: for any ℓ in $\{0, 1, \ldots, k_A\}$,

$$K_{A,\ell,j} = \bigcup_{i=(j-1)2^{k-\ell}+1}^{j2^{k-\ell}} K_{k,i},$$

implying that for any ℓ in $\{0, 1, \ldots, k\}$: $K_A = \bigcup_{j=1}^{2^\ell} K_{A,\ell,j}$.

Step 2. Proof of Inequality (4.5). Here we consider K_A as constructed in step 1, with

$$\delta = \frac{\log 2}{2\log A}.$$

Since $A \geq 2$, with this selection of δ we get that $\delta \leq 1/2$. Since $k(A) \leq \log A/\log 2$, it follows that

$$\lambda\big([0, A] \setminus K_A\big) \leq A\delta k(A) \leq A/2 \text{ whence } \lambda(K_A) \geq A/2.$$

We estimate now the Laplace transform of S_{K_A}. We first notice that since K_A is included in $[0, A]$, then if $tM \leq (\frac{c_0}{\log A}) \wedge \sqrt{\frac{c}{2A}}$, by applying Lemma 8, we derive that

$$\log\mathbb{E}\exp(tS_{K_A}) \leq 6.2At^2v^2 + A\frac{tM}{2}\exp(-c/(2tM)).$$

Since $tM \le c/(8\log A)$, we have

$$\exp(-c/(2tM)) \le A^{-2}\exp(-c/(4tM)). \tag{4.8}$$

Consequently

$$\log \mathbb{E}\exp(tS_{K_A}) \le 6.2At^2v^2 + A^{-1}\frac{tM}{2}\exp(-c/(4tM)),$$

proving (4.5) since $tM \le 1/2$. Then we assume in the rest of the proof that $(c/(2A))^{1/2} < tM \le c_0/(\log A)$, and we shall then estimate the Laplace transform of S_{K_A} by the diadic recurrence. Let t be a positive real. Since $K_{A,1,1}$ and $K_{A,1,2}$ are spaced by an interval of size $A\delta$ and $A\delta \ge 2$ (since $A \ge 20$), by using Lemma 15 below and condition (1.3), we derive that

$$\begin{aligned}\mathbb{E}\exp(tS_{K_A}) &= \mathbb{E}\exp(tS_{K_{A,1,1}})\exp(tS_{K_{A,1,2}})\\ &\le \mathbb{E}\exp(tS_{K_{A,1,1}})\mathbb{E}\exp(tS_{K_{A,1,2}}) + \exp(-cA\delta + A(1-\delta)tM).\end{aligned}$$

Since the variables are centered, $\mathbb{E}\exp(tS_{K_{A,1,i}}) \ge 1$ for $i = 1, 2$. Hence by taking into account (4.4), we obtain that

$$\log \mathbb{E}\exp(tS_{K_A}) \le \sum\nolimits_{i=1}^{2} \log \mathbb{E}\exp(tS_{K_{A,1,i}}) + \exp(-cA\delta + A(1-\delta)tM). \tag{4.9}$$

Now, let

$$\ell = \ell(t) = \inf\Big\{k \in \mathbb{Z} : A((1-\delta)/2)^k \le \frac{c}{2(tM)^2}\Big\}. \tag{4.10}$$

Notice that $\ell(t) \ge 1$ since $t^2M^2 > c/(2A)$. In addition by the selection of k_A and since $\delta \le 1/2$ and $tM \le \sqrt{\frac{c}{8}}$, it follows that $\ell(t) \le k_A$. Notice also, by the bound on tM and since $A \ge 4$, we have

$$A\delta\frac{(1-\delta)^{\ell(t)-1}}{2^{\ell(t)-1}} > \frac{c\delta}{2(tM)^2} \ge 2.$$

Using the homogeneity properties of K_A, the decomposition (4.9) and iterating until $\ell(t)$, we get that

$$\begin{aligned}\log \mathbb{E}\exp(tS_{K_A}) &\le \sum_{j=1}^{2^\ell} \log \mathbb{E}\exp(tS_{K_{A,\ell,j}})\\ &\quad + \sum_{j=0}^{\ell-1} 2^j \exp\Big(-cA\delta\frac{(1-\delta)^j}{2^j} + 2tMA\frac{(1-\delta)^{j+1}}{2^{j+1}}\Big).\end{aligned} \tag{4.11}$$

Consequently, for any $t \le c\delta/(2M)$,

$$\log \mathbb{E}\exp(tS_{K_A}) \le \sum_{j=1}^{2^\ell} \log \mathbb{E}\exp(tS_{K_{A,\ell,j}}) + \sum_{j=0}^{\ell-1} 2^j \exp\Big(-\frac{cA\delta}{2}\frac{(1-\delta)^j}{2^j}\Big). \tag{4.12}$$

Whence, since $2^{\ell(t)} \le A$ and $tM \le c\delta/2$ we obtain

$$\sum_{j=0}^{\ell(t)-1} 2^j \exp\Big(-\frac{cA\delta}{2}\frac{(1-\delta)^j}{2^j}\Big) \le 2^{\ell(t)}\exp\big(-c^2\delta/(2tM)^2\big) \le A\exp(-c/(2tM)).$$

Now we estimate each of the terms $\mathbb{E}\exp(tS_{K_{A,\ell,j}})$. By the definition of $\ell(t)$ the conditions of Lemma 8 are satisfied for $S_{K_{A,\ell,j}}$ with $B = A((1-\delta)/2)^{\ell(t)}$. Consequently,

$$\log \mathbb{E}\exp(tS_{K_{A,\ell,j}}) \le B\big(6.2v^2t^2 + tM\exp(-c/(2tM))\big).$$

Therefore, by using (4.8), we derive that

$$\begin{aligned}\log \mathbb{E}\exp(tS_{K_A}) &\le 6.2v^2t^2A + tMA^{-1}\exp(-c/4tM) + A^{-1}\exp(-c/(4tM))\\ &\le 6.2v^2t^2A + (c+1)A^{-1}\exp(-c/4tM).\end{aligned}$$

This ends the proof of Inequality (4.5).

Proof of Inequality (4.6). The proof of this part uses the same construction with the difference that we do not remove the holes from the set and we use instead their upper bound. Once again if $tM \le (c/(2A))^{1/2}$, applying Lemma 8 together with the fact that $\exp(-c/(2tM)) \le 2tM/c$ and $A \ge 4$, we derive that

$$\log \mathbb{E}\exp(tS_{(0,A]}) \le A\log A(6.2t^2v^2 + (tM)^2/c),$$

Taking into account that $v^2 \le KM^2$ with $K = 1 + 8\sum_{i\ge 1}\alpha_i$, the inequality (4.6) holds true with $C \ge 6.2K + 1/c$. Then we can assume without loss of generality in the rest of the proof that $(c/(2A))^{1/2} < tM < \frac{c\wedge 1}{2}$. We start by selecting $\delta = 2tM/c < 1$. For this δ, we select k_A as before and $\ell = \ell(t)$ as in relation (4.10). At first stage we divide as before the interval $[0, A]$ in 3 parts, the central one having a Lebesgue measure $A\delta$. Notice that $A\delta \ge 2$ since $tM > \sqrt{c/(2A)} \ge c/A$ by the fact that $A \ge 2c$. Consequently, since the variables are bounded by M, by condition (1.3),

$$\begin{aligned}\mathbb{E}\exp(tS_{(0,A]}) &\le [\mathbb{E}\exp(tS_{K_{A,1,1}})\exp(tS_{K_{A,1,2}})]e^{tAM\delta}\\ &\le [\mathbb{E}\exp(tS_{K_{A,1,1}})\mathbb{E}\exp(tS_{K_{A,1,2}}) + \exp(-A\delta c + AtM)]e^{tAM\delta}.\end{aligned}$$

Since the variables are centered, $\mathbb{E}\exp(tS_{K_{A,1,i}})) \ge 1$ for $i = 1, 2$. Hence applying (4.4) and recalling that $\delta = 2tM/c$, we obtain

$$\log \mathbb{E}\exp(tS_{[0,A]}) \le \sum\nolimits_{i=1}^{2} \log \mathbb{E}\exp(tS_{K_{A,1,i}})) + \exp(-A\delta c/2) + tAM\delta.$$

Then, we repeat the same procedure starting with $K_{A,1,1}$ and $K_{A,1,2}$, and after $\ell = \ell(t)$ iterations we obtain

$$\begin{aligned}\log \mathbb{E}\exp(tS_{(0,A]}) &\le \sum_{i=1}^{2^\ell} \log \mathbb{E}\exp(tS_{K_{A,\ell,i}})\\ &\quad + \sum_{i=0}^{\ell-1} 2^i\bigg(\exp\bigg(-\delta c\frac{A(1-\delta)^i}{2^{i+1}}\bigg) + tM\delta\frac{A(1-\delta)^i}{2^i}\bigg).\end{aligned}$$

The above computation is valid since by the definition of $\ell(t)$

$$A\delta\frac{(1-\delta)^{\ell(t)-1}}{2^{\ell(t)-1}} > \frac{c\delta}{2(tM)^2} = \frac{2tM}{2(tM)^2} \ge 2.$$

By the above considerations and the selection of ℓ, proceeding as in the proof of Inequality (4.5), we obtain

$$\log \mathbb{E}\exp(tS_{(0,A]}) \le 6.2v^2t^2A + A\bigg(\frac{tM}{2} + 1\bigg)\exp(-c/(2tM)) + tM\delta A\sum_{i=0}^{\ell\ -1}(1-\delta)^j.$$

Now notice that for the selection $\delta = 2tM/c$ and since $\ell \le k_A$, we have

$$tAM\delta\ell \le 2t^2M^2A\log A/(c\log 2).$$

Also since for any $x \ge 0$, $\exp(x) \ge x \vee (x^2/2)$, we get

$$\exp(-c/(2tM)) \le (2tM/c) \wedge (8t^2M^2/c^2).$$

Overall

$$\log \mathbb{E}\exp(tS_{(0,A]}) \le 6.2v^2t^2A + t^2M^2A(1/c + 8/c^2) + 2t^2M^2A\log A/(c\log 2).$$

By taking into account that $v^2 \le KM^2$ we obtain the desired result with the constant

$$C = 6.2K + (1/c + 8/c^2) + 2/(c\log 2), \tag{4.13}$$

where $K = (1 + 8\sum_{i\ge 1}\alpha_i)$. □

To prepare for the proof of Theorem 1 we shall reformulate the conclusions of Lemma 10 in an alternative form. Keeping the same notations as in Lemma 10, the following corollary holds.

Corollary 11. *Let $(X_i)_{i\ge 1}$ be as in Theorem 1. Assume that $A \ge 2(c \vee 10)$ and $0 \le tM \le c_0/(\log A)$ with $c_0 = \frac{c}{8} \wedge \sqrt{\frac{c\log 2}{8}}$. Then, there is a constant C' depending only on c such that*

$$\log \mathbb{E}(\exp(tS_{K_A})) \le \frac{C'^2A(v + M/A)^2}{1 - t(\log A)/c_0}. \tag{4.14}$$

Assume that $A \ge 2(c \vee 10)$ and $0 \le tM < (c \wedge 1)/2$, then for the constant C defined in (4.13),

$$\log \mathbb{E}(\exp(tS_{(0,A]}) \le \frac{Ct^2AM^2\log A}{1 - 2tM/(c \wedge 1)}. \tag{4.15}$$

Before proving Theorem 1 we remark that the second part of the above corollary already gives a bound on the tail probability with a correction in the quadratic term in x.

Corollary 12. *Under conditions of Theorem 1, for all $n \ge 2(c \vee 2)$ and $x \ge 0$,*

$$\mathbb{P}(|S_n| \ge x) \le \exp\Big(-\frac{x^2}{n(\log n)4CM^2 + 4Mx/(c \wedge 1)}\Big), \tag{4.16}$$

where C is defined in (4.13).

4.2. Proof of Theorem 1

If $n \le 16(c \vee 10)^2$ then for any positive t such that $tM < \frac{1}{4(c\vee 10)^2}$ we get that $|tS_n| \le 4$. Hence as in the beginning of the proof of Lemma 8, we derive that

$$\log \mathbb{E}(\exp(tS_n)) \le 3.1\frac{nv^2t^2}{1 - 4tM(c \vee 10)^2}.$$

We assume now that $n \geq 16(c \vee 10)^2$. Let us first introduce the following notation: for any positive real A, let K_A be the Cantor set as defined in step 1 of the proof of Lemma 8, let $\lambda(K_A)$ be the Lebesgue measure of K_A, and let F_A be the nondecreasing and continuous function from $[0, A]$ onto $[0, A - \lambda(K_A)]$ defined by

$$F_A(t) = \lambda([0,t] \cap K_A^c) \text{ for any } t \in [0, A], \tag{4.17}$$

where $K_A^c = [0, A] \setminus K_A$. Let F_A^{-1} be the inverse function of F_A. Let $A_0 = n$. Define then the real-valued process $(X_t^{(1)})_t$ from $(X_t)_{t \in [0, A_0]}$ by

$$X_t^{(1)} = X_{F_{A_0}^{-1}(t)} \text{ for any } t \in [0, A_0 - \lambda(K_{A_0})].$$

Let $A_1 = A_0 - \lambda(K_{A_0})$. Clearly, the random process $(X_t^{(1)})_{t \in [0, A_1]}$ is uniformly bounded by M and verifies (1.5) with the same constant. We now define inductively the sequence $(A_i)_{i \geq 0}$ and the random processes $(X_t^{(i)})_{i \in [0, A_i]}$ as follows. First $A_0 = n$ and $(X_t^{(0)}) = (X_t)$. And second for any nonnegative integer i, $A_{i+1} = A_i - \lambda(K_{A_i})$ and, for any t in $[0, A_{i+1}]$,

$$X_t^{(i+1)} = X^{(i)}_{F_{A_i}^{-1}(t)}. \tag{4.18}$$

Then, for any nonnegative integer j, the following decomposition holds

$$\int_0^n X_u \, du = \sum_{i=0}^{j-1} \int_{K_{A_i}} X_u^{(i)} \, du + \int_0^{A_j} X_u^{(j)} \, du. \tag{4.19}$$

Let

$$Y_i = \int_{K_{A_i}} X_u^{(i)} \, du \text{ for } 0 \leq i \leq j-1 \text{ and } Z_j = \int_0^{A_j} X_u^{(j)} \, du. \tag{4.20}$$

Now set

$$L = L_n = \inf\{j \in \mathbb{N}^*, \, A_j \leq n/(\log n)\}.$$

Notice that, since $A_j \leq n/2^j$,

$$L \leq [(\log \log n)/(\log 2)] + 1. \tag{4.21}$$

Also by the definition of L, $A_{L-1} \geq n/(\log n)$. Since $\log n \leq 2\sqrt{n}$, it follows that

$$A_{L-1} \geq \sqrt{n}/2 \geq 2(c \vee 10).$$

Hence, we can apply the inequality (4.14) to each Y_j for all $0 \leq j \leq L_n - 1$. Consequently for every $0 \leq j \leq L_n - 1$, and any positive t satisfying $tM < c_0/(\log(n/2^j))$,

$$\log \mathbb{E}(\exp(tY_j)) \leq \frac{C'^2 (v(n/2^j)^{1/2} + (n/2^j)^{-1/2} M)^2}{1 - Mt(\log(n/2^j))/c_0}. \tag{4.22}$$

To estimate Z_L, we first assume that $A_L \geq 2(c \vee 2)$. Applying inequality (4.15) we then obtain, for any positive t such that $tM < (c \wedge 1)/2$,

$$\log \mathbb{E}(\exp(tZ_L)) \leq \frac{Ct^2 M^2 n}{1 - 2tM/(c \wedge 1)}.$$

To aggregate all the contributions, we now apply Lemma 13 of Appendix with $\kappa_i = M(\log(n/2^j))/c_0$ and $\sigma_i^2 = C'^j)^{1/2} + (n/2^j)^{-1/2}M)^2$ for $0 \le i \le L-1$, $\sigma_L^2 = CnM^2$ and $\kappa_L = 2M/(c \wedge 1)$. Consequently, by (4.21), there exists C_1 depending only on c such that

$$\sum_{i=1}^{L} \kappa_i + \kappa_L \le C_1 M(\log n)(\log\log n).$$

Furthermore

$$\sum_{i=0}^{L-1} \sigma_i + \sigma_L \le \sqrt{C'}(4vn^{1/2} + 2M((\log n)/n)^{1/2}) + \sqrt{Cn}M.$$

Hence by Lemma 13, for any $n \ge 4$ and any positive $t < 1/(MC_1 \log n(\log\log n))$ there exists C_2 depending only on c such that

$$\log \mathbb{E}(\exp(tS_n)) \le \frac{C_2 n t^2 M^2}{1 - tMC_1(\log n)(\log\log n)},$$

and the result follows. If $A_L \le 2(c \vee 2)$, it suffices to notice that if $tM < 2/(c \vee 2)$, then $|tZ_L| \le 4$. Hence as in the proof of Lemma 8, we derive that

$$\log \mathbb{E}(\exp(tZ_L)) \le 3.1 \frac{4(c \vee 2)^2 t^2 M^2}{1 - tM(c \vee 2)/2},$$

and we proceed as before with $\kappa_L = M(c \vee 2)/2$ and $\sigma_L = 2(c \vee 2)M$. Inequality (2.1) follows from the Laplace transform estimate by standard computations.

4.3. Proof of Theorem 2

We proceed as in the proof of Theorem 1 with the difference that for $n \ge 2(c \vee 10)$, we choose

$$L = L_n = \inf\{j \in \mathbb{N}^*,\ A_j \le 2(c \vee 10)\}.$$

Consequently,

$$L \le \Big[\frac{\log(n) - \log(2(c \vee 10))}{\log 2}\Big] + 1.$$

4.4. Proof of Theorem 4

The proof is based on the construction of the Cantor-like sets as described in the proof of Lemma 10. Let $(\varepsilon_n)_{n \ge 1}$ be a sequence converging to 0 that will be constructed later. Without loss of generality we assume $\varepsilon_n < \log 2$ and define

$$\delta_n = \frac{\varepsilon_n}{\log n}. \tag{4.23}$$

We impose for the moment that ε_n has to satisfy

$$\delta_n \sqrt{n a_n} \to \infty. \tag{4.24}$$

(It is always possible to choose such an ε_n since (2.5) is assumed). Select in addition

$$k_n = \inf\Big\{ j \in \mathbb{N}^* : n\frac{(1-\delta_n)^j}{2^j} \le \sqrt{n a_n} \Big\}.$$

Construct the intervals $K_{k_n,i}$, $1 \le i \le 2^{k_n}$, as in the step 1 of the proof of Lemma 10. Then $R_n = (0,n]\backslash K_n$, with $K_n = \cup_{i=1}^{2^{k_n}} K_{k_n,i}$, has a Lebesgue measure smaller than $\delta_n n k_n = o(n)$ and by inequality (2.1) there exists a constant C depending only on c such that

$$a_n \log \mathbb{P}(|S_{R_n}| \ge x\sigma_n/\sqrt{a_n}) \le -\frac{Cx^2\sigma_n^2}{\delta_n k_n n M^2 + Mx\sqrt{\sigma_n^2/a_n}(\log n)(\log\log n)}.$$

Taking into account that $\liminf_{n\to\infty} \sigma_n^2/n > 0$, Condition (2.5) ensures that

$$\lim_{n\to\infty} a_n \log \mathbb{P}(|S_{R_n}| \ge x\sigma_n/\sqrt{a_n}) = -\infty.$$

According to Theorem 4.2.13 in Dembo and Zeitouni [12], S_{R_n} is negligible for the moderate deviations type of behavior. To treat the main part we rewrite the inequality (4.11) by using both sides of Lemma 15 from Appendix and the fact that for n large enough, (4.24) entails that for any t

$$\frac{|tM|}{\sigma_n\sqrt{a_n}} \le \delta_n/2 \text{ and } n\delta_n \frac{(1-\delta_n)^{k_n-1}}{2^{k_n-1}} > \delta_n\sqrt{na_n} \ge 2,$$

(by using also the fact that $\liminf_n \sigma_n^2/n > 0$). So, by definition of the sets $K_{k_n,i}$ from relation (4.7), we get for any real t and n large enough,

$$\Big|\log \mathbb{E}\exp\Big(\frac{tS_{K_n}}{\sigma_n\sqrt{a_n}}\Big) - \sum_{i=1}^{2^{k_n}} \log \mathbb{E}\exp\Big(\frac{tS_{K_{k_n,i}}}{\sigma_n\sqrt{a_n}}\Big)\Big| \le \sum_{j=0}^{k_n-1} 2^j \exp\Big(-\frac{cn\delta_n}{2}\frac{(1-\delta_n)^j}{2^j}\Big).$$

Now, by the definition of k_n notice that $2^{k_n-1} \le \sqrt{n/a_n}$. Whence

$$\sum_{j=0}^{k_n-1} 2^j \exp\Big(-n\delta_n\frac{(1-\delta_n)^j}{2^j}\Big) \le 2^{k_n}\exp\Big(-\frac{c}{2}\delta_n\sqrt{na_n}\Big) \le 2\sqrt{\frac{n}{a_n}}\exp\Big(-\frac{c}{2}\delta_n\sqrt{na_n}\Big)$$
$$\le 2\frac{\sqrt{na_n}}{a_n}\exp(-\delta_n\sqrt{na_n}) \le \frac{2}{a_n}\exp\Big(\log(\sqrt{na_n}) - \frac{c}{2}\delta_n\sqrt{na_n}\Big).$$

We select now the sequence ε_n used in the construction of δ_n, in such a way that

$$\log(\sqrt{na_n}) - \frac{c}{2}\delta_n\sqrt{na_n} \to \infty. \tag{4.25}$$

Denote

$$A_n = \frac{\sqrt{na_n}}{(\log n)^2} \text{ and } B_n = \frac{\sqrt{na_n}}{(\log n)(\log\log n)}.$$

If $na_n \ge (\log n)^5$ we select $\varepsilon_n = A_n^{-1/2}$. If $na_n < (\log n)^5$ we take $\varepsilon_n = B_n^{-1/2}$. Notice that by construction and by (2.5), $\varepsilon_n \to 0$ as $n \to \infty$ and (4.24) is satisfied. It is easy to see that

$$\log(\sqrt{na_n}) - c\delta_n\sqrt{na_n}/2 \to -\infty \text{ when } n \to \infty.$$

Indeed, if $na_n \ge (\log n)^5$, then $\varepsilon_n = A_n^{-1/2}$, so that

$$\log(\sqrt{na_n}) - c\delta_n\sqrt{na_n}/2 \le (\log n)/2 - \frac{c}{\log n}\frac{A_n}{2A_n^{1/2}}(\log n)^2$$
$$= (\log n)(1 - cA_n^{1/2})/2.$$

If $na_n < (\log n)^5$, then $\varepsilon_n = B_n^{-1/2}$, so that

$$\log(\sqrt{na_n}) - c\delta_n\sqrt{na_n}/2 = \log(\frac{\sqrt{na_n}}{(\log n)^{5/2}}) + \log[(\log n)^{5/2}]$$
$$-c\varepsilon_n B_n(\log\log n)/2 \le (5/2)\log\log n - cB_n^{1/2}(\log\log n)/2.$$

Consequently, for any real t, we get

$$a_n \Big| \log \mathbb{E}\exp\Big(\frac{tS_{K_n}}{\sigma_n\sqrt{a_n}}\Big) - \sum_{i=1}^{2^{k_n}} \log \mathbb{E}\exp\Big(\frac{tS_{K_{k_n,i}}}{\sigma_n\sqrt{a_n}}\Big)\Big| \to 0 \text{ as } n \to \infty.$$

Therefore the proof is reduced to proving the MDP for a triangular array of independent random variables $S^*_{K_{k_n,i}}$, $1 \le i \le 2^{k_n}$, each having the same law as $S_{K_{k_n,i}}$. By the selection of k_n, for any $1 \le i \le 2^{k_n}$, $\|S_{K_{k_n,i}}\|_\infty \le M\sqrt{na_n}$. In addition, for each $\epsilon > 0$

$$\lim_{n\to\infty} \frac{1}{\sigma_n^2} \sum\nolimits_{j=1}^{2^{k_n}} E(S^2_{K_{k_n,j}} I(|S_{K_{k_n,j}}| > \epsilon\sigma_n\sqrt{a_n})) = 0,$$

by using inequality (2.1) to give an upper bound of $\mathbb{P}(|S_{K_{k_n,i}}| \ge x)$. Hence, by Lemma 14, we just have to prove that

$$\lim_{n\to\infty} \frac{1}{\sigma_n^2} \sum_{j=1}^{2^{k_n}} \mathbb{E}(S_{K_{k_n,j}})^2 = 1. \tag{4.26}$$

With this aim we first notice that since $\mathrm{Card}R_n = o(n)$, it follows that when $n \to \infty$, $\mathrm{Var}(S_{R_n})/n \to 0$. Hence to prove (4.26), it suffices to prove that

$$\mathrm{Var}\Bigg(\sum_{i=1}^{2^{k_n}} S_{K_{k_n,i}}\Bigg) \Big/ \sum_{i=1}^{2^{k_n}} \mathrm{Var}(S_{K_{k_n,i}}) \to 1 \text{ as } n \to \infty. \tag{4.27}$$

Between the two sets $K_{k_n,i}$ and $K_{k_n,j}$, for $i \ne j$, there is a gap at least equal to $n\delta_n 2^{1-k_n}(1-\delta_n)^{k_n-1} > \delta_n\sqrt{na_n}$, by the selection of k_n. Consequently, since for all j, $\|S_{K_{k_n,j}}\|_\infty \le M\sqrt{na_n}$, by (1.2) we get

$$\sum\nolimits_{i=1}^{2^{k_n}-1} \sum\nolimits_{j=i+1}^{2^{k_n}-1} \mathrm{cov}(S_{K_{k_n,i}}, S_{K_{k_n,j}}) \le 4 \times 2^{k_n} M^2 na_n \sum_{j\ge 1} \exp(-cj\delta_n\sqrt{na_n}). \tag{4.28}$$

Since $2^{k_n} \le 2\sqrt{n/a_n}$, by using (4.25), it follows that

$$\sum\nolimits_{i=1}^{2^{k_n}-1} \sum\nolimits_{j=i+1}^{2^{k_n}-1} \mathrm{Cov}(S_{K_{k_n,i}}, S_{K_{k_n,j}}) = o(n), \tag{4.29}$$

which together with the fact that $C_1 n \le \mathrm{Var}(\sum_{i=1}^{2^{k_n}} S_{K_{k,i}}) \le C_2 n$ implies (4.27).

4.5. Proof of Theorem 6

The proof is similar to that of Theorem 4 with the following modifications. Inequality (2.3) is used instead of Inequality (2.1) (notice that (2.8) implies $n/M_n^2 \to \infty$), and the sequences ε_n (defining δ_n) and k_n are selected as follows:

$$\varepsilon_n \to 0 \text{ and } \varepsilon_n \frac{\sqrt{na_n}}{M_n(\log n)^2} \to \infty, \tag{4.30}$$

and

$$k_n = \inf\Big\{ j \in \mathbb{N}^* : n(M_n \vee 1)\frac{(1-\delta_n)^j}{2^j} \le \sqrt{na_n} \Big\}.$$

Notice that with this selection, $2^{k_n} \le 2(M_n \vee 1)\sqrt{n/a_n}$. By the selection of ε_n, all the steps of the previous theorem can be done similarly. Also to prove that

$$\mathrm{Var}\bigg(\sum_{i=1}^{2^{k_n}} S_{K_{k_n,i}} \bigg) \Big/ \sum_{i=1}^{2^{k_n}} \mathrm{Var}(S_{K_{k_n,i}}) \to 1 \text{ as } n \to \infty,$$

we make use of Condition (2.7) together with the fact that by the selection of k_n, for all j, $\|S_{K_{k_n,j}}\|_\infty \le \sqrt{na_n}$. The inequality (4.28) becomes

$$\begin{aligned}\sum_{i=1}^{2^{k_n}-1} \sum_{j=i+1}^{2^{k_n}-1} &\mathrm{Cov}(S_{K_{k_n,i}}, S_{K_{k_n,j}}) \\ &\le 8(M_n \vee 1) n a_n \sqrt{n/a_n} \sum_{j\ge 1} \exp(-cj\delta_n (M_n \vee 1)^{-1}\sqrt{na_n}),\end{aligned}$$

which implies (4.29) by the selection of δ_n and the fact that $M_n = o(\sqrt{n})$.

4.6. Proof of Corollary 7

For each $n \ge 1$, let us construct the following sequence of triangular arrays: for any $i \in \mathbb{Z}$,

$$X_{n,i} = \frac{1}{\sqrt{\delta} h_T} \Big\{ \int_{(i-1)\delta}^{i\delta} K\Big(\frac{x - X_t}{h_T}\Big) dt - \mathbb{E} \int_{(i-1)\delta}^{i\delta} K\Big(\frac{x - X_t}{h_T}\Big) dt \Big\},$$

where $n\delta = T$, $n = [T]$, $(T \ge 1)$ and consequently $2 > \delta \ge 1$. Notice that

$$\sum_{i=1}^{n} X_{n,i} = T\big(f_T(x) - \mathbb{E} f_T(x)\big).$$

Now for any $k \ge 1$, the strong mixing coefficients $\alpha_n(k)$ of the processes $(X_{n,i})_{i\in\mathbb{Z}}$ are uniformly bounded by the strong mixing coefficient α_{k-1} of the process $(X_t, t \in \mathbb{R})$. Hence to apply Theorem 6, it suffices to show (2.7) and to prove that

$$T^{-1}\mathrm{Var}\bigg(\sum_{i=1}^{n} X_{n,i} \bigg) \to 2 \int_0^\infty g_u(x,x) du \text{ as } n \to \infty.$$

The above convergence was proved by Castellana and Leadbetter [7] under assumptions on g_u. To prove (2.7), we first notice that for all $j > i$,

$$\mathrm{Cov}(X_{i,n}, X_{j,n}) = \frac{1}{\delta h_T^2} \int_{\mathbb{R}^2} K\Big(\frac{x-y}{h_T}\Big) K\Big(\frac{x-z}{h_T}\Big) \int_{i\delta-\delta}^{i\delta} \int_{j\delta-\delta}^{j\delta} g_{t-s}(y,z) ds dt dy dz.$$

Consequently, since K is a kernel, for all $j > i$,

$$\big|\mathrm{Cov}(X_{i,n}, X_{j,n})\big| \le \int_{(j-i-1)\delta}^{(j-i+1)\delta} \sup_{x,y} |g_u(x,y)| du.$$

Similarly

$$\operatorname{Var}(X_{i,n}) \leq 2 \int_0^\delta \sup_{x,y} |g_u(x,y)| du.$$

Hence (2.7) holds with

$$v^2 \leq 2 \int_0^\delta \sup_{x,y} |g_u(x,y)| du + 4 \int_0^\infty \sup_{x,y} |g_u(x,y)| du.$$

To finish the proof, it remains to notice that if f is differentiable and such that f' is l-Lipschitz for a positive constant l then, since K is a kernel,

$$|\mathbb{E} f_T(x) - f(x)| = O(h_T^2)$$

(see for instance relation 4.15 in Bosq [3]).

Appendix

One of our tools is the technical lemma below, which provides bounds for the log-Laplace transform of any sum of real-valued random variables. It comes from Lemma 3 in Merlevède, Peligrad and Rio [17].

Lemma 13. *Let $Z_0, Z_1, \dots$ be a sequence of real valued random variables. Assume that there exists positive constants $\sigma_0, \sigma_1, \dots$ and $\kappa_0, \kappa_1, \dots$ such that, for any $i \geq 0$ and any t in $[0, 1/c_i[$,*

$$\log \mathbb{E} \exp(tZ_i) \leq (\sigma_i t)^2 / (1 - \kappa_i t).$$

Then, for any positive n and any t in $[0, 1/(\kappa_0 + \kappa_1 + \cdots + \kappa_n)[$,

$$\log \mathbb{E} \exp(t(Z_0 + Z_1 + \cdots + Z_n)) \leq (\sigma t)^2 / (1 - \kappa t),$$

where $\sigma = \sigma_0 + \sigma_1 + \cdots + \sigma_n$ and $\kappa = \kappa_0 + \kappa_1 + \cdots + \kappa_n$.

The next lemma is due to Arcones [2, Lemma 2.3] and it permits us to derive the MDP for triangular array of independent r.v.'s.

Lemma 14 (Arcones [2]). *Let $\{X_{n,j}; 1 \leq j \leq k_n\}$ be a triangular array of independent r.v.'s with mean zero. Let $\{a_n\}_{n \geq 1}$ be a sequence of real numbers converging to 0. Suppose that:*
(i) The following limit exists and is finite:

$$\lim_{n \to \infty} \sum_{j=1}^{k_n} \mathbb{E}(X_{n,j}^2) = \sigma^2,$$

(ii) There exists a constant C such that for each $1 \leq j \leq k_n$,

$$|X_{n,j}| \leq C \sqrt{a_n},$$

(iii) For each $\epsilon > 0$

$$\lim_{n \to \infty} \sum_{j=1}^{k_n} \mathbb{E}(X_{n,j}^2 I(|X_{n,j}| > \epsilon \sqrt{a_n}) = 0.$$

Then for all real t, $a_n \sum_{j=1}^{k_n} \log \mathbb{E} \exp(t X_{n,j}) \to t^2 \sigma^2 / 2$ and consequently the MDP holds for $(\sum_{j=1}^{k_n} X_{j,n})$ with speed a_n and good rate function $I(t) = t^2 / (2\sigma^2)$.

We first recall the following lemma, which is a well-known corollary of Ibragimov's covariance inequality for nonnegative and bounded random variables.

Lemma 15 (Ibragimov [14]). *Let $Z_1, \dots, Z_p$ be real-valued nonnegative random variables each a.s. bounded, and let*

$$\alpha = \sup_{k \in [1,p]} \alpha(\sigma(Z_i : i \le k), \sigma(Z_i : i > k)).$$

Then

$$\mathbb{E}(Z_1 \cdots Z_p) \le \mathbb{E}(Z_1) \cdots \mathbb{E}(Z_p) + (p-1)\alpha \|Z_1\|_\infty \cdots \|Z_p\|_\infty,$$

and

$$\mathbb{E}(Z_1) \cdots \mathbb{E}(Z_p) \le \mathbb{E}(Z_1 \cdots Z_p) + (p-1)\alpha \|Z_1\|_\infty \cdots \|Z_p\|_\infty.$$

Acknowledgments

The authors are indebted to the referee for carefully reading of the manuscript and for helpful comments.

References

[1] Adamczak, R. (2008). A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.* **13** 1000–1034.

[2] Arcones, M. A. (2003). Moderate deviations of empirical processes. In *Stochastic Inequalities and Applications. Progr. Probab.* **56** 189-212. Birkhäuser, Basel.

[3] Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction*, 2nd ed. *Lecture Notes in Statistics*. Springer-Verlag.

[4] Bradley, R. C. (1997). On quantiles and the central limit question for strongly mixing sequences. *J. Theor. Probab.* **10** 507–555.

[5] Bradley, R. C. (2007). *Introduction to Strong Mixing Conditions*, Vol. 1, 2, 3. Kendrick Press.

[6] Bryc, W. and Dembo, A. (1996). Large deviations and strong mixing. *Ann. Inst. Henri Poincaré* **32** 549–569.

[7] Castellana, J. V. and Leadbetter, M. R. (1986). On smoothed probability density estimation for stationary process. *Stochastic Process. Appl.* **21** 179–193.

[8] Chen, X. and de Acosta, A. (1998). Moderate deviations for empirical measures of Markov chains: Upper bounds. *J. Theoret. Probab.* **11**(4) 1075–1110.

[9] de Acosta, A. (1997). Moderate deviations for empirical measures of Markov chains: Lower bounds. *Ann. Probab.* **25**(1) 259–284.

[10] Dedecker, J. and Prieur, C. (2004). Coupling for tau-dependent sequences and applications. *J. Theoret. Probab.* **17** 861–885.

[11] Dedecker, J., Merlevède, F., Peligrad, M. and Utev, S. (2009). Moderate deviations for stationary sequences of bounded random variables. *Ann. Inst. H. Poincaré Probab. Statist.* **45** 453–476.

[12] Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. Springer, New York.

[13] Doukhan, P. and Neumann, M. (2007). Probability and moment inequalities for sums of weakly dependent random variables, with applications. *Stochastic Process. Appl.* **117**(7) 878–903.

[14] Ibragimov, I. A. (1962). Some limit theorems for stationary processes. *Theor. Probab. Appl.* **7** 349–382.

[15] Leblanc, F. (1997). Density estimation for a class of continuous time processes. *Math. Methods of Stat.* **6** 171–199.

[16] Merlevède, F. and Peligrad, M. (2009). Functional moderate deviations for triangular arrays and applications. *ALEA.* **5** 3–20.

[17] Merlevède, F., Peligrad, M. and Rio, E. (2009). A Bernstein type inequality and moderate deviations for weakly dependent sequences. Preprint. Available at arXiv: 0902.0582.

[18] Rio, E. (2000). Théorie asymptotique des processus aléatoires faiblement dépendants. *Mathématiques et Applications* **31**. Springer, Berlin.

[19] Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U. S. A.* **42** 43–47.

[20] Veretennikov, A. Yu. (1990). On hypoellipticity conditions and estimates of the mixing rate for stochastic differential equations. *Soviet Math. Dokl.* **40** 94–97.

[21] Tsirelson, B. (2008). Moderate deviations for random fields and random complex zeroes. Available at arXiv:0801.1050.