Deviation inequalities for martingales with applications

Xiequan Fan[*,a,b], Ion Grama[c], Quansheng Liu[*,c]

[a] *Center for Applied Mathematics, Tianjin University, Tianjin 300072, China*
[b] *Regularity Team, Inria, France*
[c] *Université de Bretagne-Sud, LMBA, UMR CNRS 6205, Campus de Tohannic, 56017 Vannes, France*

Abstract

Using changes of probability measure developed by Grama and Haeusler (Stochastic Process. Appl., 2000), we extend the deviation inequalities of Lanzinger and Stadtmüller (Stochastic Process. Appl., 2000) and Fuk and Nagaev (Theory Probab. Appl., 1971) to the case of martingales. Our inequalities recover the best possible decaying rate in the independent case. In particular, these inequalities improve the results of Lesigne and Volný (Stochastic Process. Appl., 2001) under a stronger condition that the martingale differences have bounded conditional moments. Applications to linear regressions with martingale difference innovations, weak invariance principles for martingales and self-normalized deviations are provided. In particular, we establish a type of self-normalized deviation bounds for parameter estimation of linear regressions. Such type bounds have the advantage that they do not depend on the distribution of the regression random variables.

Keywords: Large deviations; Martingales; Deviation inequalities; Linear regressions; Weak invariance principles; Self-normalized deviations

2000 MSC: Primary 60G42, 60E15, 60F10; Secondary 62J05, 60F17.

1. Introduction

Assume that $(\xi_i)_{i\geq 1}$ is a sequence of centered random variables. Denote by $S_n = \sum_{i=1}^n \xi_i$ the partial sums of $(\xi_i)_{i\geq 1}$. If $(\xi_i)_{i\geq 1}$ are independent and identically distributed (i.i.d.) and satisfy the following subexponential condition: for a constant $\alpha \in (0,1)$,

$$K_\alpha := \mathbf{E}[\xi_1^2 \exp\{(\xi_1^+)^\alpha\}] < \infty, \tag{1.1}$$

where $x^+ = \max\{x, 0\}$, Lanzinger and Stadtmüller [26] have obtained the following subexponential inequality: for any $x, y > 0$,

$$\mathbf{P}\Big(S_n \geq x\Big) \quad \leq \quad \exp\left\{-\frac{x}{y^{1-\alpha}}\Big(1 - \frac{nK_\alpha}{2xy^{1-\alpha}}\Big)\right\} + \frac{n}{e^{y^\alpha}}\mathbf{E}[\exp\{(\xi_1^+)^\alpha\}]. \tag{1.2}$$

*Corresponding authors.
E-mail: fanxiequan@hotmail.com (X. Fan), ion.grama@univ-ubs.fr (I. Grama), quansheng.liu@univ-ubs.fr (Q. Liu).

In particular, by taking $y = x$, inequality (1.2) implies that for any $x > 0$,

$$\limsup_{n\to\infty} \frac{1}{n^\alpha} \log \mathbf{P}\Big(S_n \geq nx\Big) \quad \leq \quad -x^\alpha \tag{1.3}$$

and

$$\mathbf{P}\Big(S_n \geq n\Big) \quad = \quad O\Big(\exp\Big\{-c\,n^\alpha\Big\}\Big), \quad n \to \infty, \tag{1.4}$$

where $c > 0$ does not depend on n and can be any value smaller than $c_0 = \sup\{t > 0 : \mathbf{E}[\exp\{t(\xi_1^+)^\alpha\}] < \infty\}$ (see [26], Remark 3). The last two results (1.3) and (1.4) are the best possible under the present condition, since a large deviation principle (LDP) with good rate function x^α can be obtained in situations where some more information on the tail behavior of ξ_1 is available; see Theorem 2.3. Under the subexponential condition (1.1), more precise estimations on tail probabilities, or large deviation expansions, can be found in Nagaev [32, 33], Saulis and Statulevičius [36] and Borovkov [3, 4].

Recently, the generalizations of (1.4) have attracted certain interest. Doukhan and Neumann [10] have established a generalization of (1.4) under a new concept of weak dependence which extends usual mixing assumptions. This concept is particularly well suited for deriving estimates for the cumulants of sums of random variables. For a LDP for weighted sum of i.i.d. random variables satisfying conditions similar to (1.1) we refer to Kiesel and Stadtmüller [25] and Gantert, Ramanan and Rembart [17].

Our first aim is to give a generalization of (1.4) for martingales. Let $(\xi_i, \mathcal{F}_i)_{i\geq 1}$ be a sequence of martingale differences with respect to the filtration $(\mathcal{F}_i)_{i\geq 1}$. Under the Cramér condition $\sup_i \mathbf{E}[\exp\{|\xi_i|\}] < \infty$, Lesigne and Volný [27] first proved that (1.4) holds with $\alpha = 1/3$, and that the power $1/3$ is optimal even for the class of stationary and ergodic sequences of martingale differences. Later, Fan, Grama and Liu [12] generalized the result of Lesigne and Volný by proving that (1.4) holds under the more general exponential moment condition $\sup_i \mathbf{E}[\exp\{|\xi_i|^{\frac{2\alpha}{1-\alpha}}\}] < \infty$, $\alpha \in (0, 1)$, and that the power α in (1.4) is optimal for the class of stationary sequences of martingale differences. It is obvious that the condition $\sup_i \mathbf{E}[\exp\{|\xi_i|^{\frac{2\alpha}{1-\alpha}}\}] < \infty$ is much stronger than condition (1.1). Thus, the result in [12] does not imply (1.4) in the i.i.d. case.

To fill this gap, we consider the case of the martingale differences having bounded conditional subexponential moments. Under this assumption, we can recover the inequalities (1.2), (1.3) and (1.4); see Theorem 2.1. Denote by $||\xi||_\infty$ the essential supremum of a random variable ξ. Our first result implies that if

$$u_n := \max\left\{\Big|\Big|\sum_{i=1}^n \mathbf{E}[\xi_i^2 \exp\{(\xi_i^+)^\alpha\}|\mathcal{F}_{i-1}]\Big|\Big|_\infty, \quad 1\right\} < \infty,$$

then, for any $x > 0$,

$$\mathbf{P}\Big(\max_{1\leq k\leq n} S_k \geq x\Big) \quad \leq \quad 2\exp\left\{-\frac{x^2}{2(u_n + x^{2-\alpha})}\right\}. \tag{1.5}$$

To illustrate the bound (1.5), consider the simple case where $(\xi_i)_{i\geq 1}$ is a sequence of i.i.d. random variables: in this case $u_n = O(n)$ as $n \to \infty$. It is interesting to see that when $0 \leq x = o(n^{1/(2-p)})$, our bound (1.5) is sub-Gaussian. When $x = ny$ with $y > 0$ fixed, the bound (1.5) is subexponential $\exp\{-c_y n^\alpha\}$, where $c_y > 0$ does not depend on n, thus recovering (1.4).

For martingale differences $(\xi_i, \mathcal{F}_i)_{i\geq 1}$ satisfying $u_n = o(n^{2-\alpha})$, we find an improved version of the inequality (1.3); see Corollary 2.2. To show the best value of the constant in (1.3), for i.i.d. random variables we establish large deviation principles for $\frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^n \xi_i$ and $\frac{1}{n}\max_{1\leq k\leq n} S_k$ in Theorem 2.3.

For the methods, an approach for obtaining subexponential bound is to combine the method of the tower property of conditional expectation and uniform norm. This approach has been applied in Fuk [16], Chung and Lu [5], Liu and Watbled [30] and Dedecker and Fan [6]. With this approach, one can obtain inequality (1.2) with

$$nK_\alpha = \sum_{i=1}^n \Big|\Big| \mathbf{E}[\xi_i^2 \exp\{(\xi_i^+)^\alpha\}|\mathcal{F}_{i-1}] \Big|\Big|_\infty . \tag{1.6}$$

However, this result is not the best possible in some cases. It turns out that with the method based on the change of probability measure for martingales from Grama and Haeusler [18] one can obtain better bounds. Using this method, we show that the inequality (1.2) holds with

$$nK_\alpha = \Big|\Big| \sum_{i=1}^n \mathbf{E}[\xi_i^2 \exp\{(\xi_i^+)^\alpha\}|\mathcal{F}_{i-1}] \Big|\Big|_\infty . \tag{1.7}$$

Such a refinement has been first proved by Freedman [14] for improving Fuk's inequality [16] for martingales, and similar analogs were established by Haeusler [19], van de Geer [38], de la Peña [8] and [11]. Since the latter nK_α is less than the former one, i.e.,

$$\Big|\Big| \sum_{i=1}^n \mathbf{E}[\xi_i^2 \exp\{(\xi_i^+)^\alpha\}|\mathcal{F}_{i-1}] \Big|\Big|_\infty \ \leq\ \sum_{i=1}^n \Big|\Big| \mathbf{E}[\xi_i^2 \exp\{(\xi_i^+)^\alpha\}|\mathcal{F}_{i-1}] \Big|\Big|_\infty ,$$

our method has certain advantage. To illustrate it, consider the following example. Assume that $(\varepsilon_i)_{i\geq 1}$ is a sequence of independent and unbounded random variables, and that $(\varepsilon_i)_{i\geq 1}$ is independent of $(\xi_i, \mathcal{F}_i)_{i\geq 1}$. Assume that

$$\Big|\Big| \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] \Big|\Big|_\infty \geq 1 \quad \text{ and } \quad \Big|\Big| \mathbf{E}[\xi_i^2 \exp\{|\xi_i|^\alpha\}|\mathcal{F}_{i-1}] \Big|\Big|_\infty \leq D$$

for a constant D and all $i \geq 1$. Denote by $\xi_i' = \xi_i\varepsilon_i \Big/ \sqrt{\sum_{k=1}^n \varepsilon_k^2}$ and $\mathcal{F}_i' = \sigma\{\varepsilon_j, 1\leq j\leq i, \mathcal{F}_i, |\varepsilon_k|, 1 \leq k \leq n\}$. Then $(\xi_i', \mathcal{F}_i')_{i\geq 1}$ is also a sequence of martingale differences. It is easy to see that

$$\Big|\Big| \sum_{i=1}^n \mathbf{E}[(\xi_i')^2 \exp\{((\xi_i')^+)^\alpha\}|\mathcal{F}_{i-1}'] \Big|\Big|_\infty \ \leq\ \Big|\Big| \sum_{i=1}^n \frac{\varepsilon_i^2}{\sum_{k=1}^n \varepsilon_k^2} \mathbf{E}[\xi_i^2 \exp\{|\xi_i|^\alpha\}|\mathcal{F}_{i-1}] \Big|\Big|_\infty \ \leq\ D,$$

and that, by the fact that $(\varepsilon_i)_{i\geq 1}$ are unbounded,

$$\begin{aligned} \sum_{i=1}^n \Big|\Big| \mathbf{E}[(\xi_i')^2 \exp\{((\xi_i')^+)^\alpha\}|\mathcal{F}_{i-1}'] \Big|\Big|_\infty &\geq \sum_{i=1}^n \Big|\Big| \frac{\varepsilon_i^2}{\sum_{k=1}^n \varepsilon_k^2} \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] \Big|\Big|_\infty \\ &\geq \sum_{i=1}^n \Big|\Big| \frac{\varepsilon_i^2}{\sum_{k=1}^n \varepsilon_k^2} \Big|\Big|_\infty \ = \ n. \end{aligned}$$

Hence

$$\sup_n \Big|\Big| \sum_{i=1}^n \mathbf{E}[(\xi_i')^2 \exp\{((\xi_i')^+)^\alpha\}|\mathcal{F}_{i-1}'] \Big|\Big|_\infty \ \leq \ D,$$

and

$$\sup_n \sum_{i=1}^n \Big|\Big|\mathbf{E}[(\xi_i')^2 \exp\{((\xi_i')^+)^\alpha\}|\mathcal{F}_{i-1}']\Big|\Big|_\infty \ = \ \infty.$$

Thus our extension (1.7) in the right hand-side of (1.2) is nontrivial, while with (1.6) it is infinite.

Using the same method, we also generalize the following Fuk inequality for martingales (cf. Corollary $3'$ of Fuk [16]; see also Nagaev [33] for independent case): assume that $||\mathbf{E}[|\xi_i|^p|\mathcal{F}_{i-1}]||_\infty < \infty$ for some $p \geq 2$ and all $i \in [1, n]$, then for any $x > 0$,

$$\mathbf{P}\Big(\max_{1\leq k\leq n} S_k \geq x \Big) \ \leq \ \exp\Bigg\{ -\frac{x^2}{2\widetilde{V}^2} \Bigg\} + \frac{\widetilde{C}_p}{x^p}\,, \tag{1.8}$$

where

$$\widetilde{V}^2 := \frac{1}{4}(p+2)^2 e^p \sum_{i=1}^n \Big|\Big|\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}]\Big|\Big|_\infty \quad \text{and} \quad \widetilde{C}_p := \Big(1+\frac{2}{p}\Big)^p \sum_{i=1}^n \Big|\Big|\mathbf{E}[|\xi_i|^p|\mathcal{F}_{i-1}]\Big|\Big|_\infty.$$

In Corollary 2.5, we prove that (1.8) holds true when $\widetilde{V}^2$ and $\widetilde{C}_p$ are replaced by the following two smaller values V^2 and C_p respectively, where

$$V^2 := \frac{1}{4}(p+2)^2 e^p \Big|\Big|\sum_{i=1}^n \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}]\Big|\Big|_\infty \quad \text{and} \quad C_p := \Big(1+\frac{2}{p}\Big)^p \Big|\Big|\sum_{i=1}^n \mathbf{E}[|\xi_i|^p|\mathcal{F}_{i-1}]\Big|\Big|_\infty.$$

To illustrate this improvement on Fuk's inequality (1.8), consider the following comparison between C_p and $\widetilde{C}_p$ in the case of self-normalized deviations. Assume that $(\varepsilon_i)_{i=1,\ldots,n}$ is a sequence of independent, unbounded and symmetric random variables. Denote by $\xi_i = \varepsilon_i/(\sum_{k=1}^n |\varepsilon_k|^p)^{1/p}$, and $\mathcal{F}_i = \sigma\{\varepsilon_j, j \leq i, |\varepsilon_k|, 1 \leq k \leq n\}$. Then $(\xi_i, \mathcal{F}_i)_{i=1,\ldots,n}$ is a sequence of martingale differences. It is easy to see that

$$\Big|\Big|\sum_{i=1}^n \mathbf{E}[|\xi_i|^p|\mathcal{F}_{i-1}]\Big|\Big|_\infty \ = \ \sum_{i=1}^n \frac{|\varepsilon_i|^p}{\sum_{k=1}^n |\varepsilon_k|^p} \ = \ 1,$$

and that, by the fact that $(\varepsilon_i)_{i=1,\ldots,n}$ are unbounded,

$$\sum_{i=1}^n \Big|\Big|\mathbf{E}[|\xi_i|^p|\mathcal{F}_{i-1}]\Big|\Big|_\infty \ = \ \sum_{i=1}^n \Big|\Big|\frac{|\varepsilon_i|^p}{\sum_{k=1}^n |\varepsilon_k|^p}\Big|\Big|_\infty \ = \ n.$$

Hence C_p is of order $O(1)$ while $\widetilde{C}_p$ is of order $O(n)$, which implies a significant improvement on Fuk's inequality (1.8).

Under the condition that $(\xi_i, \mathcal{F}_i)_{i\geq 1}$ satisfy $\sup_i \mathbf{E}[|\xi_i|^p] < \infty$ for a $p \geq 2$, Lesigne and Volný [27] proved that

$$\mathbf{P}\Big(S_n \geq n\Big) \ = \ O\Big(\frac{1}{n^{p/2}}\Big), \qquad\qquad n \to \infty. \tag{1.9}$$

Under a stronger condition on the conditional moments, we obtain an improvement on the inequality of Lesigne and Volný (1.9). Assume either

$$\sup_i \mathbf{E}[|\xi_i|^{p+\delta}] < \infty \qquad \text{and} \qquad \sup_i ||\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}]|| < \infty$$

or

$$\sup_i ||\mathbf{E}[|\xi_i|^p|\mathcal{F}_{i-1}]|| < \infty$$

for two positive constants δ and $p \geq 2$ (do not depend on n). Then we have for any $\alpha \in (\frac{1}{2}, \infty)$,

$$\mathbf{P}\Big(\max_{1\leq k\leq n} S_k \geq n^\alpha \Big) \ = \ O\Big(\frac{1}{n^{\alpha p-1}}\Big), \qquad n \to \infty. \tag{1.10}$$

Since $p-1 \geq p/2$ for $p \geq 2$, it follows that (1.10) is an improvement of (1.9). We refer to our Theorem 2.6 and Corollary 2.5 where we give more precise bounds. From Theorem 4.1 of Hao and Liu [22] it follows that (1.10) is close to the optimal. For necessary and sufficient conditions to have (1.10) we refer to [22].

The paper is organized as follows. We present our main results in Section 2, and discuss the applications to linear regressions with martingale difference innovations, weak invariance principles for martingales and self-normalized deviations in Section 3. The proofs of theorems are given in Sections 4 - 9.

2. Main results

Assume that we are given a sequence of real-valued martingale differences $(\xi_i, \mathcal{F}_i)_{i=0,...,n}$ defined on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$, where $\xi_0 = 0$ and $\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq ... \subseteq \mathcal{F}_n \subseteq \mathcal{F}$ are increasing σ-fields. So, by definition, $\mathbf{E}[\xi_i|\mathcal{F}_{i-1}] = 0,\ i = 1, ..., n.$ Define the martingale $S := (S_k, \mathcal{F}_k)_{k=0,...,n}$ by setting

$$S_0 = 0 \qquad \text{and} \qquad S_k = \sum_{i=1}^k \xi_i, \qquad k = 1, ..., n. \tag{2.1}$$

Denote by $\langle S \rangle$ the quadratic characteristic of the martingale S :

$$\langle S\rangle_0 = 0 \qquad \text{and} \qquad \langle S\rangle_k = \sum_{i=1}^k \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}], \quad k = 1, ..., n. \tag{2.2}$$

For any $\alpha \in (0, 1)$, set

$$\Upsilon(S)_k = \sum_{i=1}^k \mathbf{E}[\xi_i^2 \exp\{(\xi_i^+)^\alpha\}|\mathcal{F}_{i-1}], \qquad k \in [1, n]. \tag{2.3}$$

Our first result is a subexponential inequality on tail probabilities for martingales. A similar inequality for separately Lipschitz functionals has been obtained recently by Dedecker and Fan [6].

Theorem 2.1. *Assume*

$$C_n := \sum_{i=1}^{n} \mathbf{E}[\xi_i^2 \exp\{(\xi_i^+)^\alpha\}] < \infty \tag{2.4}$$

for some $\alpha \in (0,1)$. *Then for all* $x, u > 0$,

$$\mathbf{P}\Big(S_k \geq x \text{ and } \Upsilon(S)_k \leq u \text{ for some } k \in [1,n]\Big)$$
$$\leq \begin{cases} \exp\Big\{-\dfrac{x^2}{2u}\Big\} + C_n\Big(\dfrac{x}{u}\Big)^{2/(1-\alpha)} \exp\Big\{-\Big(\dfrac{u}{x}\Big)^{\alpha/(1-\alpha)}\Big\} & \text{if } 0 \leq x < u^{1/(2-\alpha)} \\ \exp\Big\{-x^\alpha\Big(1-\dfrac{u}{2\,x^{2-\alpha}}\Big)\Big\} + C_n\dfrac{1}{x^2}\exp\Big\{-x^\alpha\Big\} & \text{if } x \geq u^{1/(2-\alpha)}. \end{cases} \tag{2.5}$$

It is obvious that

$$C_n = \mathbf{E}[\Upsilon(S)_n] \leq ||\Upsilon(S)_n||_\infty.$$

Hence, when $u \geq \max\{||\Upsilon(S)_n||_\infty, 1\}$, from (2.5) we get the following rough bounds

$$\mathbf{P}\Big(\max_{1\leq k\leq n} S_k \geq x\Big) \leq \begin{cases} 2\exp\Big\{-\dfrac{x^2}{2\,u}\Big\} & \text{if } 0 \leq x < u^{1/(2-\alpha)} \\ 2\exp\Big\{-\dfrac{1}{2}x^\alpha\Big\} & \text{if } x \geq u^{1/(2-\alpha)} \end{cases} \tag{2.6}$$

$$\leq 2\exp\Big\{-\frac{x^2}{2(u+x^{2-\alpha})}\Big\}. \tag{2.7}$$

Thus for moderate $x \in (0, u^{1/(2-\alpha)})$, the bound (2.5) is sub-Gaussian. For $x \geq u^{1/(2-\alpha)}$, the bound (2.5) is subexponential of the order $\exp\Big\{-\frac{1}{2}x^\alpha\Big\}$. Moreover, when $\frac{x}{u^{1/(2-\alpha)}} \to \infty$, by (2.5), this order can be improved to $\exp\Big\{-(1-\varepsilon)x^\alpha\Big\}$, for any given small $\varepsilon > 0$.

For any $\alpha \in (0,1)$ and $k \in [1,n]$, set

$$\widehat{\Upsilon}(S)_k = \sum_{i=1}^{k} \mathbf{E}[\xi_i^2 \exp\{|\xi_i|^\alpha\}|\mathcal{F}_{i-1}].$$

Since $\Upsilon(S)_k \leq \widehat{\Upsilon}(S)_k$, it is obvious that (2.5) is also an upper bound on the tail probabilities

$$\mathbf{P}\Big(\pm S_k \geq x \text{ and } \widehat{\Upsilon}(S)_k \leq u \text{ for some } k \in [1,n]\Big).$$

Moreover, if $||\widehat{\Upsilon}(S)_n||_\infty \leq u$, then (2.5) is an upper bound on tail probabilities of the maxima and minima of the partial sums $\mathbf{P}(\max_{1\leq k\leq n} S_k \geq x)$ and $\mathbf{P}(-\min_{1\leq k\leq n} S_k \geq x)$.

When $(\xi_i)_{i\geq 1}$ are i.i.d. random variables, we have $C_n = \Upsilon(S)_n = c_0 n$ with $c_0 = \mathbf{E}[\xi_1^2 \exp\{(\xi_1^+)^\alpha\}]$. In this case, inequality (2.6) implies the following large deviation bound: for any $x > 0$ and n sufficiently large,

$$\mathbf{P}\Big(\max_{1\leq k\leq n} S_k \geq nx\Big) \leq 2\exp\Big\{-c_x n^\alpha\Big\} \tag{2.8}$$

with $c_x = \frac{x^\alpha}{2}$. In fact, the constant c_x in (2.8) is close to x^α, as shown by the following large deviation bound for martingales, which is a consequence of Theorem 2.1.

Corollary 2.2. *Assume condition (2.4), for some $\alpha \in (0,1)$. If*

$$||\Upsilon(S)_n||_\infty \;=\; o(n^{2-\alpha}), \qquad n \to \infty, \tag{2.9}$$

then for any $x \geq 0$,

$$\limsup_{n\to\infty} \frac{1}{n^\alpha} \log \mathbf{P}\left(\frac{1}{n}\max_{1\leq k\leq n} S_k \geq x\right) \;\leq\; -x^\alpha. \tag{2.10}$$

The bound (2.10) cannot be improved under condition (2.9). This is shown by the following large deviation principles on $\frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^n \xi_i$ and $\frac{1}{n}\max_{1\leq k\leq n} S_k$ which hold true for i.i.d. random variables $(\xi_i)_{i\geq 1}$ with subexponential tails.

Theorem 2.3. *Consider an i.i.d. sequence $(\xi_i)_{i\geq 1}$ with $\mathbf{E}[\xi_1] = 0$.*

(i) If $\mathbf{E}[\xi_1^2] < \infty$ and for some constants $\alpha \in (0,1)$ and $c > 0$,

$$\lim_{x\to\infty} \frac{1}{x^\alpha} \log \mathbf{P}\Big(\xi_1 \geq x\Big) = -c, \tag{2.11}$$

then for all $x > 0$,

$$\lim_{n\to\infty} \frac{1}{n^\alpha} \log \mathbf{P}\left(\frac{1}{n}\max_{1\leq k\leq n} S_k \geq x\right) \;=\; \lim_{n\to\infty} \frac{1}{n^\alpha} \log \mathbf{P}\left(\frac{S_n}{n} \geq x\right) \;=\; -c\,x^\alpha. \tag{2.12}$$

(ii) If for some constant $\alpha \in (0,1)$ and $c > 0$,

$$\lim_{x\to\infty} \frac{1}{x^\alpha} \log \mathbf{P}\Big(|\xi_1| \geq x\Big) = -c, \tag{2.13}$$

then for all $x > 0$,

$$\lim_{n\to\infty} \frac{1}{n^\alpha} \log \mathbf{P}\left(\frac{1}{n}\max_{1\leq k\leq n} S_k \geq x\right) \;=\; \lim_{n\to\infty} \frac{1}{n^\alpha} \log \mathbf{P}\left(\frac{S_n}{n} \geq x\right) \;=\; -c\,x^\alpha \tag{2.14}$$

and

$$\lim_{n\to\infty} \frac{1}{n^\alpha} \log \mathbf{P}\left(\frac{1}{n}\min_{1\leq k\leq n} S_k \leq -x\right) \;=\; \lim_{n\to\infty} \frac{1}{n^\alpha} \log \mathbf{P}\left(\frac{S_n}{n} \leq -x\right) \;=\; -c\,x^\alpha. \tag{2.15}$$

In the special case where ξ_1 has a density $p(x)$ satisfying $p(x) \sim \exp\{-|x|^\alpha\}$ as $x \to \infty$, the asymptotic (2.12) for $\mathbf{P}\left(\frac{S_n}{n} \geq x\right)$ follows from Theorem 3 in Nagaev [32] (we also refer to Gantert, Ramanan and Rembart [17] for related results). Notice that condition (2.13) is weaker than that used in [32].

Since the rate function $x \mapsto c\,x^\alpha$ is continuous and strictly increasing on $\mathbf{R}_+ = [0,\infty)$, by Lemma 4.4 in Huang and Liu [23] we can reformulate the results of the previous theorem as large deviation principles (the proof therein was given for two sided tails but it can be easily adapted to one sided tails). We give the details below.

From part (i) of the previous theorem it follows that, if $\mathbf{E}[\xi_1^2] < \infty$ and (2.11) holds, then $\frac{S_n}{n}$ and $\frac{1}{n}\max_{1\le k\le n} S_k$ verify a LDP with norming factor $1/n^\alpha$ and rate function $x \mapsto x^\alpha$ on $\mathbf{R}_+$: for any Borel set $B \subset \mathbf{R}_+$, we have

$$-\inf_{x\in B^0} c\, x^\alpha = \liminf_{n\to\infty} \frac{1}{n^\alpha} \log \mathbf{P}\left(\frac{S_n}{n} \in B\right) \le \limsup_{n\to\infty} \frac{1}{n^\alpha} \log \mathbf{P}\left(\frac{S_n}{n} \in B\right) = -\inf_{x\in \overline{B}} c\, x^\alpha, \quad (2.16)$$

where B^0 and $\overline{B}$ denote the interior and the closure of B respectively; moreover the result (2.16) remains true when S_n is replaced by $\max_{1\le k\le n} S_k$.

From part (ii) it follows that, if (2.13) holds, then $\frac{S_n}{n}$ and $\frac{1}{n}\max_{1\le k\le n} S_k$ verify a LDP with norming factor $1/n^\alpha$ and rate function $x \mapsto x^\alpha$ on $\mathbf{R} = (-\infty, \infty)$, that is (2.16) holds for any Borel set $B \subset \mathbf{R}$.

Now we consider the case when the martingale differences $(\xi_i, \mathcal{F}_i)_{i=0,\dots,n}$ have absolute moments of order $p \ge 2$. For any $p > 2$ denote

$$\Xi(S)_k = \sum_{i=1}^k \mathbf{E}[(\xi_i^+)^p|\mathcal{F}_{i-1}], \qquad k \in [1, n].$$

We prove the following inequality, which is similar to the results of Haeusler [19] and [11].

Theorem 2.4. *Let $p \ge 2$. Assume $\mathbf{E}[|\xi_i|^p] < \infty$ for all $i \in [1, n]$. Then for all $x, y, v, w > 0$,*

$$\begin{aligned}&\mathbf{P}\Big(S_k \ge x,\ \langle S\rangle_k \le v \text{ and } \Xi(S)_k \le w \text{ for some } k \in [1, n]\Big)\\ &\quad\le \exp\left\{-\frac{\alpha^2 x^2}{2e^p v}\right\} + \exp\left\{-\frac{\beta x}{y}\log\left(1 + \frac{\beta x y^{p-1}}{w}\right)\right\} + \mathbf{P}\left(\max_{1\le i\le n} \xi_i > y\right), \end{aligned} \quad (2.17)$$

where

$$\alpha = \frac{2}{p+2} \qquad \text{and} \qquad \beta = 1 - \alpha. \quad (2.18)$$

Setting $y = \beta x$, we obtain the following generalization of the Fuk-Nagaev inequality (1.8).

Corollary 2.5. *Let $p \ge 2$. Assume $||\mathbf{E}[|\xi_i|^p|\mathcal{F}_{i-1}]||_\infty < \infty$ for all $i \in [1, n]$. Then for all $x > 0$,*

$$\mathbf{P}\Big(\max_{1\le k\le n} S_k \ge x\Big) \le \exp\left\{-\frac{x^2}{2\,V^2}\right\} + \frac{C_p}{x^p}, \quad (2.19)$$

where

$$V^2 = \frac{1}{4}(p+2)^2 e^p \Big|\Big|\langle S\rangle_n\Big|\Big|_\infty \qquad \text{and} \qquad C_p = \Big(1 + \frac{2}{p}\Big)^p \Big|\Big|\sum_{i=1}^n \mathbf{E}[|\xi_i|^p|\mathcal{F}_{i-1}]\Big|\Big|_\infty. \quad (2.20)$$

It is worth noting that if $\sup_i ||\mathbf{E}[|\xi_i|^p|\mathcal{F}_{i-1}]||_\infty \leq C$ for a constant C, then, by Jensen's inequality, it holds $\sup_i ||\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}]||_\infty \leq C^{2/p}$. Therefore, inequality (2.19) implies the following sub-Gaussian bound: for any $x = O(\sqrt{n}\,(\log n)^\beta)$, $n \to \infty$, with β satisfying $\beta > 0$ if $p = 2$ and $\beta \in (0, 1/2]$ if $p > 2$,

$$\mathbf{P}\Big(\max_{1\leq k\leq n} S_k \geq x\Big) \;=\; O\bigg(\exp\bigg\{-C\,\frac{x^2}{n}\bigg\}\bigg), \tag{2.21}$$

where $C > 0$ does not depend on x and n. Inequality (2.19) also implies that for any $\alpha \in (\frac{1}{2}, \infty)$ and any $x > 0$,

$$\mathbf{P}\Big(\max_{1\leq k\leq n} S_k > n^\alpha x\Big) \;=\; O\Big(\frac{c_x}{n^{\alpha p-1}}\Big), \qquad n \to \infty, \tag{2.22}$$

where $c_x > 0$ does not depend on n. The asymptotic (2.22) was first obtained by Fuk [16] and it is the best possible under the stated condition even for the sums of independent random variables (cf. Fuk and Nagaev [15]).

When the martingale differences $(\xi_i, \mathcal{F}_i)_{i\geq 1}$ satisfy $\sup_i \mathbf{E}[|\xi_i|^p] < \infty$, Lesigne and Volný [27] proved that for any $x > 0$,

$$\mathbf{P}\Big(S_n > nx\Big) \;=\; O\Big(\frac{c_x}{n^{p/2}}\Big), \qquad n \to \infty, \tag{2.23}$$

where $c_x > 0$ does not depend on n, and that the order $n^{-p/2}$ is optimal for the class of stationary and ergodic sequences of martingale differences. When $\alpha = 1$, equality (2.22) implies the following large deviation convergence rate: for any $x > 0$,

$$\mathbf{P}\Big(\max_{1\leq k\leq n} S_k > nx\Big) \;=\; O\Big(\frac{c_x}{n^{p-1}}\Big), \qquad n \to \infty, \tag{2.24}$$

where $c_x > 0$ does not depend on n. When $p \geq 2$, it holds $p - 1 \geq p/2$. Thus (2.24) refines the bound (2.23) under the stronger assumption that the p-th conditional moments are uniformly bounded $\sup_i ||\mathbf{E}[|\xi_i|^p|\mathcal{F}_{i-1}]||_\infty < \infty$. Moreover, the following proposition of Lesigne and Volný [27] shows that the estimate of (2.24) cannot be essentially improved even in the i.i.d. case.

Proposition A. *Let $p \geq 1$ and $(c_n)_{n\geq 1}$ be a real positive sequence approaching zero. There exists a sequence of i.i.d. random variables $(\xi_i)_{i\geq 1}$ such that $\mathbf{E}[|\xi_i|^p] < \infty$, $\mathbf{E}[\xi_i] = 0$ and*

$$\limsup_{n\to\infty} \frac{n^{p-1}}{c_n}\mathbf{P}(|S_n| \geq n) = \infty.$$

When $\mathbf{E}[|\xi_i|^2|\mathcal{F}_{i-1}]$ and $\mathbf{E}[|\xi_i|^p]$, $i = 1, ..., n$, are uniformly bounded for some $p > 2$, (but for the same p the condition $\mathbf{E}[|\xi_i|^p|\mathcal{F}_{i-1}] \leq C$ may be violated for some $i \in [1, n]$), we have the following result.

Theorem 2.6. *Let $p \geq 2$. Assume $\mathbf{E}[|\xi_i|^{p+\delta}] < \infty$ for a small $\delta > 0$ and all $i \in [1, n]$. Then for all $x, v > 0$,*

$$\begin{aligned}\mathbf{P}\Big(S_k \geq x \textit{ and } \langle S\rangle_k \leq v^2 \textit{ for some } k \in [1, n]\Big)&\\ \leq \exp\left\{-\frac{x^2}{2\Big(v^2 + \frac{1}{3}x^{(2p+\delta)/(p+\delta)}\Big)}\right\} + \frac{1}{x^p}\sum_{i=1}^n \mathbf{E}\Big[|\xi_i|^{p+\delta}\mathbf{1}_{\{\xi_i > x^{p/(p+\delta)}\}}\Big].&\end{aligned} \tag{2.25}$$

In particular, as a consequence of (2.25), if $\sup_i \mathbf{E}[|\xi_i|^{p+\delta}] < \infty$ and $\sup_i ||\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}]||_\infty < \infty$, we obtain that for any $\alpha \in (\frac{1}{2}, \infty)$ and any $x > 0$,

$$\begin{aligned}\mathbf{P}\Big(\max_{1\leq k\leq n} S_k \geq n^\alpha x\Big) &\leq \exp\Big\{ - c_x \min\Big\{n^{\frac{\alpha\delta}{p+\delta}}, n^{2\alpha-1}\Big\}\Big\} + \frac{C/x^p}{n^{\alpha p-1}} \\ &= O\Big(\frac{c_x}{n^{\alpha p-1}}\Big), \qquad n\to\infty, \qquad (2.26)\end{aligned}$$

where $c_x > 0$ does not depend on n.

Note that (2.22) and (2.26) have the same convergence rate. To highlight the difference between the conditions under which we obtain (2.22) and (2.26), notice that the assumption $\sup_i ||\mathbf{E}[|\xi_i|^p|\mathcal{F}_{i-1}]||_\infty < \infty$ has been replaced by the two assumptions $\sup_i \mathbf{E}[|\xi_i|^{p+\delta}] < \infty$ and $\sup_i ||\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}]||_\infty < \infty$. It is obvious that the two assumptions $\sup_i ||\mathbf{E}[|\xi_i|^p|\mathcal{F}_{i-1}]||_\infty < \infty$ and $\sup_i \mathbf{E}[|\xi_i|^{p+\delta}] < \infty$ do not imply each other. Therefore Corollary 2.5 and Theorem 2.6 do not imply each other. Actually, Hao and Liu [22, Theorem 4.1] gave necessary and sufficient conditions for (2.26) to hold, in particular from their result it follows that (2.26) holds under the weaker conditions $\sup_i \mathbf{E}[|\xi_i|^p] < \infty$ and $\sup_i ||\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}]||_\infty < \infty$ and the exponent in (2.26) is optimal.

The following corollary of Theorem 2.6 is obvious.

Corollary 2.7. *Assume the condition of Theorem 2.6. Then for all $x, v > 0$,*

$$\begin{aligned}\mathbf{P}\Big(\max_{1\leq k\leq n} S_k \geq x\Big) &\leq \exp\Big\{-\frac{x^2}{2(nv^2 + \frac{1}{3}x^{(2p+\delta)/(p+\delta)})}\Big\} \\ &\quad + \frac{1}{x^p}\sum_{i=1}^n \mathbf{E}\Big[|\xi_i|^{p+\delta}\mathbf{1}_{\{\xi_i > x^{p/(p+\delta)}\}}\Big] + \frac{1}{v^{p+\delta}}\mathbf{E}\Big[\Big|\frac{\langle S\rangle_n}{n}\Big|^{(p+\delta)/2}\Big]. \qquad (2.27)\end{aligned}$$

Moreover,

$$\mathbf{E}\Big[\Big|\frac{\langle S\rangle_n}{n}\Big|^{(p+\delta)/2}\Big] \leq \frac{1}{n}\sum_{i=1}^n \mathbf{E}[|\xi_i|^{p+\delta}]. \qquad (2.28)$$

Inequality (2.28) implies that if $\sup_i \mathbf{E}[|\xi_i|^p] < \infty$ for a $p \geq 2$, then $\mathbf{E}[|\langle S\rangle_n/n|^{p/2}]$ are uniformly bounded for all n.

Compared to Theorem 2.6, Corollary 2.7 is simpler, since it only requires the moment of $\langle S\rangle_n$ instead of bounding $\langle S\rangle_n$ uniformly in n.

Assume $\sup_i \mathbf{E}[|\xi_i|^{p+\delta}] < \infty$ for same $p \geq 2$ (without any condition on $\langle S\rangle_n$). Applying (2.28) to (2.27) with $nv^2 = \frac{2}{3}x^{(2p+\delta)/(p+\delta)}$, we obtain for all $x, v > 0$,

$$\mathbf{P}\Big(\max_{1\leq k\leq n} S_k \geq x\Big) \leq \exp\Big\{-\frac{1}{2}x^{\delta/(p+\delta)}\Big\} + \frac{nC}{x^p} + \Big(\frac{3n}{2}\Big)^{\frac{p+\delta}{2}}\frac{C}{x^{p+\delta/2}}. \qquad (2.29)$$

The last inequality shows that for any $x > 0$,

$$\mathbf{P}\Big(\max_{1\leq k\leq n} S_k \geq n\Big) = O\Big(\frac{1}{n^{p/2}}\Big), \qquad n\to\infty. \qquad (2.30)$$

Since $\delta > 0$ is arbitrary small, the asymptotic (2.30) is close to the best possible large deviation convergence rate $n^{-(p+\delta)/2}$ given by Lesigne and Volný [27] (cf. (2.23)).

3. Applications

The exponential concentration inequalities for martingales have many applications. McDiarmid [31], Rio [35] and Dedecker and Fan [6] applied such type inequalities to estimate the concentration of separately Lipschhitz functions. Liu and Watbled [30] adopted these inequalities to deduce asymptotic properties of the free energy of directed polymers in a random environment. We refer to Bercu and Touati [1] and [13] for more interesting applications of the concentration inequalities for martingales. In the sequel, we provide some applications of our results to linear regressions with martingale difference innovations, weak invariance principles for martingales and self-normalized deviations.

(a). Linear regressions

Linear regressions can be used to investigate the impact of one variable on the other, or to predict the value of one variable based on the other. For instance, if one wants to see impact of footprint size on height, or predict height according to a certain given value of footprint size. The stochastic linear regression model is given by, for all $k \in [1, n]$,

$$X_k = \theta \phi_k + \varepsilon_k \, , \tag{3.1}$$

where $(X_k)_{k=1,...,n}$, $(\phi_k)_{k=1,...,n}$ and $(\varepsilon_k)_{k=1,...,n}$ are the observations, the regression random variables and the driven noises, respectively. We assume that $(\phi_k)_{k=1,...,n}$ is a sequence of independent random variables, and that $(\varepsilon_k)_{k=1,...,n}$ is a sequence of martingale differences with respect to the natural filtration. Moreover, we suppose that $(\phi_k)_{k=1,...,n}$ and $(\varepsilon_k)_{k=1,...,n}$ are independent. Our interest is to estimate the unknown parameter θ. The well-known least-squares estimator θ_n is given below

$$\theta_n = \frac{\sum_{k=1}^n \phi_k X_k}{\sum_{k=1}^n \phi_k^2}. \tag{3.2}$$

Recently, Bercu and Touati [1] have obtained some very precise exponential bounds on the tail probabilities $\mathbf{P}\left(|\theta_n - \theta| \geq x\right)$. However, their precise bounds depend on the distribution of the regression random variables $(\phi_k)_{k=1,...,n}$, which restricts the applications of these bounds when the distributions of the regression variables are unknown. When $(\varepsilon_k)_{k=1,...,n}$ are independent normal random variables with a common variation $\sigma^2 > 0$, Liptser and Spokoiny [29] have established the following estimation: for all $x \geq 1$,

$$\mathbf{P}\left(\pm (\theta_n - \theta)\sqrt{\Sigma_{k=1}^n \phi_k^2} \geq x \right) \ \leq \ \sqrt{\frac{2}{\pi}} \, \frac{\sigma}{x} \, \exp\left\{ -\frac{x^2}{2\,\sigma^2} \right\}. \tag{3.3}$$

When $(\varepsilon_k)_{k=1,...,n}$ are conditionally sub-Gaussian, similar estimation is allowed to be obtained in Liptser and Spokoiny [29]. An interesting feature of bound (3.3) is a type of self-normalized deviations and the bound does not depend on the distribution of the regression random variables. Thus the self-normalized approximation $(\theta_n - \theta)\sqrt{\Sigma_{k=1}^n \phi_k^2}$ has certain advantage on the usual approximation of $\theta_n - \theta$.

Next, we would like to consider the case that $(\varepsilon_k)_{k=1,...,n}$ are martingale differences. When the regression random variables are constants, such case has been considered in Section 5 of Dedecker and Merlevède [7], where the authors gave rate of convergence for the normal approximation of $\theta_n - \theta$ in terms of minimal distance. In the following theorems, we will assume that the regression variables are random. More importantly, we will consider the self-normalized approximation $(\theta_n-\theta)\sqrt{\Sigma_{k=1}^n\phi_k^2}$ instead of the usual approximation of $\theta_n-\theta$ considered in Bercu and Touati [1] and Dedecker and Merlevède [7].

Theorem 3.1. *Assume for two constants $\alpha\in(0,1)$ and D,*

$$\mathbf{E}\Big[\varepsilon_i^2 e^{|\varepsilon_i|^\alpha}\Big|\sigma\{\varepsilon_j, j\leq i-1\}\Big]\leq D$$

for all $i\in[1,n]$. Then for any $u\geq\max\{D,1\}$ and all $x>0$,

$$\mathbf{P}\left(\pm(\theta_n-\theta)\sqrt{\Sigma_{k=1}^n\phi_k^2}\geq x\right) \leq \begin{cases} 2\exp\Big\{-\dfrac{x^2}{2\,u}\Big\} & \text{if } 0\leq x<u^{1/(2-\alpha)} \\ 2\exp\Big\{-\dfrac{1}{2}x^\alpha\Big\} & \text{if } x\geq u^{1/(2-\alpha)} \end{cases} \tag{3.4}$$

$$\leq 2\exp\left\{-\frac{x^2}{2(u+x^{2-\alpha})}\right\}. \tag{3.5}$$

In particular, it holds for any $x>0$,

$$\mathbf{P}\bigg(\pm(\theta_n-\theta)\sqrt{\Sigma_{k=1}^n\phi_k^2}\geq\sqrt{n}\,x\bigg) = O\bigg(\exp\Big\{-c_x n^{\alpha/2}\Big\}\bigg), \qquad n\to\infty, \tag{3.6}$$

where $c_x>0$ does not depend on n.

If $(\varepsilon_k)_{k=1,\ldots,n}$ have the Weibull distributions and the conditional variances are uniformly bounded, then we have the following inequality which has the same exponentially decaying rate of (3.6).

Theorem 3.2. *Assume for three constants $\alpha\in(0,1)$, E and F,*

$$\mathbf{E}\Big[\varepsilon_i^2\Big|\sigma\{\varepsilon_j, j\leq i-1\}\Big]\leq E \qquad \text{and} \qquad \mathbf{E}\Big[\exp\{|\varepsilon_i|^{\frac{\alpha}{1-\alpha}}\}\Big]\leq F$$

for all $i\in[1,n]$. Then for all $x>0$,

$$\mathbf{P}\left(\pm(\theta_n-\theta)\sqrt{\Sigma_{k=1}^n\phi_k^2}\geq x\right) \leq \exp\left\{-\frac{x^2}{2(E+\frac{1}{3}x^{2-\alpha})}\right\}+nF\exp\Big\{-x^\alpha\Big\}. \tag{3.7}$$

In particular, equality (3.6) holds.

If $(\varepsilon_k)_{k=1,\ldots,n}$ have finite conditional moments, by Corollary 2.5, then we have the following result.

Theorem 3.3. *Let $p\geq2$. Assume for a constant A,*

$$\mathbf{E}\Big[|\varepsilon_i|^p\Big|\sigma\{\varepsilon_j, j\leq i-1\}\Big]\leq A$$

for all $i \in [1, n]$. Then for all $x > 0$,

$$\mathbf{P}\Big(\pm (\theta_n - \theta)\sqrt{\Sigma_{k=1}^n \phi_k^2} \geq x \Big) \ \leq \ \exp\left\{ -\frac{x^2}{2\,V^2} \right\} + \frac{C_p}{x^p}\,, \tag{3.8}$$

where

$$V^2 = \frac{1}{4}(p+2)^2 e^p\, A^{2/p} \quad \text{and} \quad C_p = \Big(1 + \frac{2}{p}\Big)^p A. \tag{3.9}$$

In particular, it holds for any $x > 0$,

$$\mathbf{P}\Big(\pm (\theta_n - \theta)\sqrt{\Sigma_{k=1}^n \phi_k^2} \geq \sqrt{n}\, x \Big) \ = \ O\Big(\frac{c_x}{n^{p/2}} \Big), \qquad n \to \infty, \tag{3.10}$$

where $c_x > 0$ does not depend on n.

A similar inequality can be obtained by applying the Fuk inequality (1.8) to the martingale difference sequence (cf. (7.1) for the definition of $(\xi_i, \mathcal{F}_i)_{i=1,...,n}$). The Fuk inequality implies that for all $x > 0$,

$$\mathbf{P}\Big(\pm (\theta_n - \theta)\sqrt{\Sigma_{k=1}^n \phi_k^2} \geq x \Big) \ \leq \ \exp\left\{ -\frac{x^2}{2nV^2} \right\} + \frac{nC_p}{x^p}, \tag{3.11}$$

where V^2 and C_p are defined by (3.9). In particular, it implies that for any $x > 0$,

$$\mathbf{P}\Big(\pm (\theta_n - \theta)\sqrt{\Sigma_{k=1}^n \phi_k^2} \geq \sqrt{n}\, x \Big) \ = \ O\Big(\frac{c_x}{n^{p/2-1}} \Big), \qquad n \to \infty, \tag{3.12}$$

where $c_x > 0$ does not depend on n. The order of (3.10) is much better than that of (3.12). Thus the refinement of (3.8) on (3.11) is significant.

If $(\varepsilon_k)_{k=1,...,n}$ have finite moments and uniformly bounded conditional variances, by Theorem 2.6, we obtain the following result which has the same polynomially decaying rate of Theorem 3.3.

Theorem 3.4. *Let $p \geq 2$. Assume for two constants A and B,*

$$\mathbf{E}\Big[\varepsilon_i^2 \,\Big|\, \sigma\{\varepsilon_j, j \leq i-1\}\Big] \leq A \qquad \text{and} \qquad \mathbf{E}\Big[|\varepsilon_i|^{p+\delta}\Big] \leq B$$

for a small $\delta > 0$ and all $i \in [1, n]$. Then for all $x > 0$,

$$\mathbf{P}\Big(\pm (\theta_n - \theta)\sqrt{\Sigma_{k=1}^n \phi_k^2} \geq x \Big) \ \leq \ \exp\left\{ -\frac{x^2}{2\Big(A + \frac{1}{3}x^{(2p+\delta)/(p+\delta)}\Big)} \right\} + \frac{B}{x^p}. \tag{3.13}$$

In particular, equality (3.10) holds.

In the following theorem, we assume that $(\varepsilon_i)_{i=1,...,n}$ have only a moment of order $p \in [1, 2]$.

Theorem 3.5. *Let $p \in [1,2]$. Assume for a constant A,*

$$\mathbf{E}[|\varepsilon_i|^p] \leq A$$

for all $i \in [1,n]$. Then for all $x > 0$,

$$\mathbf{P}\bigg(\pm (\theta_n - \theta)\sqrt{\Sigma_{k=1}^n \phi_k^2} \geq x \bigg) \ \leq \ \frac{2A}{x^p}. \tag{3.14}$$

In particular, equality (3.10) holds.

Theorems 3.1 and 3.5 focus on obtaining the large deviation inequalities. These inequalities do not depend on the distribution of input random variables $(\phi_k)_{k=1,...,n}$. Similar bounds are also expected to be obtained via the decoupling techniques of de la Peña [8] and de la Peña and Giné [9]. In particular, if $(\varepsilon_k)_{k=1,...,n}$ are independent random variables (instead of martingale differences), with the method of conditionally independent in de la Peña and Giné [9], more precise bounds, but depending on the distribution of input random variables, may be established.

Haeusler and Joos [20] proved that if the martingale differences satisfy $\mathbf{E}[|\xi_i|^{2+\delta}] < \infty$ for a constant $\delta > 0$ and all $i \in [1,n]$, then there exists a constant C_δ, depending only on δ, such that for all $x \in \mathbf{R}$,

$$\bigg|\mathbf{P}(S_n \leq x) - \Phi(x)\bigg| \leq C_\delta \left(\sum_{i=1}^n \mathbf{E}\Big[|\xi_i|^{2+\delta}\Big] + \mathbf{E}\Big[|\langle S\rangle_n - 1|^{1+\delta/2}\Big] \right)^{1/(3+\delta)} \frac{1}{1+|x|^{2+\delta}}, \tag{3.15}$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\{-t^2/2\}dt$ is the standard normal distribution; see also Hall and Heyde [21] with the larger factor $\frac{1}{1+|x|^{4(1+\delta/2)^2/(3+\delta)}}$ replacing $\frac{1}{1+|x|^{2+\delta}}$. Using (3.15), we obtain the following nonuniform Berry-Esseen bound, which depends on the distribution of input random variables.

Theorem 3.6. *Let $p > 2$. Assume that $(\varepsilon_i)_{i=1,...,n}$ satisfy $\mathbf{E}[\varepsilon_i^2 \ |\sigma\{\varepsilon_j, j \leq i-1\}] = \sigma^2$ a.s. for a positive constant σ and all $i \in [1,n]$. Assume $\mathbf{E}[|\varepsilon_i|^p] \leq A$ for a constant A and all $i \in [1,n]$. Then for all $x \in \mathbf{R}$,*

$$\bigg|\mathbf{P}\bigg((\theta_n - \theta)\sqrt{\Sigma_{k=1}^n \phi_k^2} \leq x\sigma\bigg) - \Phi(x)\bigg| \leq C_p \left(\sum_{i=1}^n \mathbf{E}\left[\left|\frac{\phi_i}{\sqrt{\Sigma_{k=1}^n \phi_k^2}}\right|^p\right]\right)^{1/(1+p)} \frac{1}{1+|x|^p}, \tag{3.16}$$

where C_p is a constant depending only on A, σ and p.

Notice that

$$\sum_{i=1}^n \mathbf{E}\left[\left|\frac{\phi_i}{\sqrt{\Sigma_{k=1}^n \phi_k^2}}\right|^p\right] \leq \sum_{i=1}^n \mathbf{E}\left[\left|\frac{\phi_i}{\sqrt{\Sigma_{k=1}^n \phi_k^2}}\right|^2\right] = 1.$$

Thus (3.16) implies that the tail probability $\mathbf{P}\Big((\theta_n - \theta)\sqrt{\Sigma_{k=1}^n \phi_k^2} \geq x\Big)$ has the decaying rate x^{-p} as $x \to \infty$, which is coincident with the inequalities (3.8) and (3.13).

(b). Weak invariance principles

In this subsection, let $(\xi_i, \mathcal{F}_i)_{i\geq 1}$ be a sequence of martingale differences. The following rate of convergence in the central limit theorem (CLT) for martingale difference sequences is due to Ouchti (cf. Corollary 1 of [34]). Assume that there exists a constant $M > 0$ such that $\mathbf{E}[|\xi_i|^3|\mathcal{F}_{i-1}] \leq M\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}]$ a.s. for all $i \in \mathbf{N}$. If the series $\sum_{i=1}^{\infty} \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}]$ diverges a.s. then there is a constant $C_M > 0$, depending on M, such that

$$\sup_{x\in\mathbf{R}} \left| \mathbf{P}\Big(S_{v(n)} \leq x\sqrt{n} \Big) - \Phi(x) \right| \ \leq \ \frac{C_M}{n^{1/4}}\, , \tag{3.17}$$

where

$$v(n) = \inf \Big\{ k \in \mathbf{N}, \quad \langle S\rangle_k \geq n \Big\}.$$

Define the process $\{H_n(t), 0 \leq t \leq 1\}$ by

$$H_n(t) = \frac{1}{\sqrt{n}} S_{v(k)} \qquad \text{for } 0 \leq t \leq 1,$$

for $t_k = k/n$ and $k = 0, \ldots, n$ and by interpolation on the interval $(t_{k-1}, t_k]$ for $k = 1, \ldots, n$. Then the following invariance principle holds.

Theorem 3.7. *Assume that there exists a constant $M > 0$ such that $\mathbf{E}[|\xi_i|^3|\mathcal{F}_{i-1}] \leq M\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}]$ a.s. for all $i \in \mathbf{N}$. If the series $\sum_{i=1}^{\infty} \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}]$ diverges a.s., then the sequence of processes $\{H_n(t), 0 \leq t \leq 1\}$ converges in distribution to the standard Wiener process in the space $\mathcal{D}_{[0,1]}$ endowed with the Srorokhod metric.*

The result can be obtained from Theorem 2 of Chapter 7 (see also Theorem 2 of Chapter 5) of Liptser and Shiryaev [28]. As an illustration of our Theorem 2.4, we will give a short proof of the tightness of the processes $\{H_n(t), 0 \leq t \leq 1\}$ which is much simpler than the proof in [28].

(c). Self-normalized deviation inequalities

Consider the self-normalized deviations for independent random variables. Assume that $(\xi_i)_{i=1,\ldots,n}$ is a sequence of independent and symmetric random variables. Denote by

$$V_n(\beta) = \Big(\sum_{i=1}^{n} |\xi_i|^\beta \Big)^{1/\beta}$$

for a constant $\beta \in (1, 2]$. We have the following self-normalized deviation inequality.

Theorem 3.8. *Assume that $(\xi_i)_{i=1,\ldots,n}$ is a sequence of independent and symmetric random variables. Let t_0 be a positive number such that*

$$\frac{1}{2}\Big(e^{x - t_0|x|^\beta} + e^{-x - t_0|x|^\beta} \Big) \leq 1 \tag{3.18}$$

for a constant $\beta \in (1, 2]$ and all $x \in \mathbf{R}$. Then for all $x > 0$,

$$\mathbf{P}\left(\max_{1\leq k\leq n} \frac{S_k}{V_n(\beta)} \geq x \right) \ \leq \ \exp\Big\{ - C(\beta, t_0)\ x^{\frac{\beta}{\beta-1}} \Big\}, \tag{3.19}$$

where

$$C(\beta, t_0) = \left(\frac{1}{t_0}\right)^{\frac{1}{\beta-1}} \left(\frac{1}{\beta}\right)^{\frac{\beta}{\beta-1}} (\beta - 1).$$

Remark 3.9. *Let us comment on Theorem 3.8.*

1. *Ideally we want to make bound (3.19) as small as possible. Since $C(\beta, t_0)$ is decreasing in t_0, it is enough to choose t_0 satisfying (3.18) as small as possible for a given β, i.e.,*

$$t_0(\beta) = \inf\Big\{t > 0,\ \frac{1}{2}\Big(e^{x-t|x|^\beta} + e^{-x-t|x|^\beta}\Big) \le 1 \text{ for all } x \in \mathbf{R}\Big\}.$$

It is obvious that $t_0(\beta) = \infty$ for $\beta > 2$ since for $\beta > 2$,

$$e^{x-t_0|x|^\beta} + e^{-x-t_0|x|^\beta} = 2 + x^2 + o(x^2), \qquad x \to 0.$$

Moreover, $t_0(\beta) \to 1$ as β is decreasing to 1, $t_0(\beta) \to 1/2$ as β is increasing to 2, and

$$t_0(\beta) \le 1 \quad \text{for all } x \in \mathbf{R} \text{ and } \beta \in (1, 2].$$

For practical purposes one can use the following table:

β	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
t_0	0.742	0.627	0.554	0.504	0.470	0.448	0.436	0.436	0.450	0.5

2. *For certain special cases, Jing, Liang and Zhou [24] have obtained the following self-normalized moderate deviation principle (MDP) result (see also Shao [37] for self-normalized LDP result). Assume that*

$$\mathbf{P}(\xi_i \ge x) = \mathbf{P}(\xi_i \le -x) \sim \frac{c}{x^\alpha} h_i(x), \qquad x \to \infty,$$

where $\alpha \in (0, 2)$, $c > 0$ and $h_i(x)$'s are slowly varying at ∞. Under certain regularity conditions on the tail probabilities of ξ_i (cf. Theorem 2.3 of [24] for details), the following limit holds for $x_n \to \infty$ and $x_n = o(n^{(\beta-1)/\beta})$ and $\beta > \max\{1, \alpha\}$,

$$\lim_{n\to\infty} x_n^{-\frac{\beta}{\beta-1}} \log \mathbf{P}\left(S_n/V_n(\beta) \ge x_n\right) = -(\beta - 1) C_\alpha(\beta), \tag{3.20}$$

where $C_\alpha(\beta)$ is a positive constant depending on α and β. For self-normalized sums of symmetric random variables, the constant $C_\alpha(\beta)$ is the solution of

$$\int_0^\infty \frac{2 - \exp\{\beta x - x^\beta/C^{\beta-1}\} - \exp\{-\beta x - x^\beta/C^{\beta-1}\}}{x^{1+\alpha}} dx = 0. \tag{3.21}$$

According to the MDP result of Jing, Liang and Zhou [24], the power $\frac{\beta}{\beta-1}$ in the right hand sides of (3.19) is the best possible for moderate x's.

3. *For $\beta \in (1, 2]$, inequality (3.19) implies the following upper bounds of LDP*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbf{P}\left(\max_{1\le k\le n} \frac{S_k}{V_n(\beta)} \ge x\, n^{\frac{\beta-1}{\beta}}\right) \le -C(\beta, t_0)\ x^{\frac{\beta}{\beta-1}}, \quad x \in (0, 1]. \tag{3.22}$$

According to the LDP result of Shao [37], the power $\frac{\beta}{\beta-1}$ in the right hand side of (3.19) is also the best possible for large x's.

4. Proof of Theorem 2.1

To prove Theorem 2.1, we need the following technical lemma based on a truncation argument.

Lemma 4.1. *Assume* $\mathbf{E}[\xi_i^2 \exp\{|\xi_i|^\alpha\}] < \infty$ *for a constant* $\alpha \in (0,1)$. *Set* $\eta_i = \xi_i \mathbf{1}_{\{\xi_i \leq y\}}$ *for* $y > 0$. *Then for all* $\lambda > 0$,

$$\mathbf{E}[e^{\lambda \eta_i} | \mathcal{F}_{i-1}] \leq 1 + \frac{\lambda^2}{2} \mathbf{E}[\eta_i^2 \exp\{\lambda y^{1-\alpha} (\eta_i^+)^\alpha\} | \mathcal{F}_{i-1}].$$

The proof of Lemma 4.1 can be found in the proof of Proposition 3.5 in Dedecker and Fan [6]. However, instead of using the tower property of conditional expectation as in Dedecker and Fan [6], we use changes of probability measure in the proof of this theorem. Set $\eta_i = \xi_i \mathbf{1}_{\{\xi_i \leq y\}}$ for some $y > 0$. The exact value of y is given later. Then $(\eta_i, \mathcal{F}_i)_{i=1,...,n}$ is a sequence of supermartingale differences, and it holds $\mathbf{E}[\exp\{\lambda \eta_i\}] < \infty$ for all $\lambda \in (0, \infty)$ and all i. Define the exponential multiplicative martingale $Z(\lambda) = (Z_k(\lambda), \mathcal{F}_k)_{k=0,...,n}$, where

$$Z_k(\lambda) = \prod_{i=1}^{k} \frac{\exp\{\lambda \eta_i\}}{\mathbf{E}[\exp\{\lambda \eta_i\} | \mathcal{F}_{i-1}]}, \qquad Z_0(\lambda) = 1.$$

If T is a stopping time, then $Z_{T \wedge k}(\lambda)$ is also a martingale, where

$$Z_{T \wedge k}(\lambda) = \prod_{i=1}^{T \wedge k} \frac{\exp\{\lambda \eta_i\}}{\mathbf{E}[\exp\{\lambda \eta_i\} | \mathcal{F}_{i-1}]}, \qquad Z_0(\lambda) = 1.$$

Thus, the random variable $Z_{T \wedge k}(\lambda)$ is a probability density on $(\Omega, \mathcal{F}, \mathbf{P})$, i.e.

$$\int Z_{T \wedge k}(\lambda) d\mathbf{P} = \mathbf{E}[Z_{T \wedge k}(\lambda)] = 1.$$

Define the *conjugate probability measure*

$$d\mathbf{P}_\lambda = Z_{T \wedge n}(\lambda) d\mathbf{P}, \tag{4.1}$$

and denote by $\mathbf{E}_\lambda$ the expectation with respect to $\mathbf{P}_\lambda$. Since $\xi_i = \eta_i + \xi_i \mathbf{1}_{\{\xi_i > y\}}$, it follows that for any $x, y, u > 0$,

$$\begin{aligned}
& \mathbf{P}\Big(S_k \geq x \text{ and } \Upsilon(S)_k \leq u \text{ for some } k \in [1, n]\Big) \\
\leq \ & \mathbf{P}\left(\sum_{i=1}^{k} \eta_i \geq x \text{ and } \Upsilon(S)_k \leq u \text{ for some } k \in [1, n]\right) \\
& + \mathbf{P}\left(\sum_{i=1}^{k} \xi_i \mathbf{1}_{\{\xi_i > y\}} > 0 \text{ for some } k \in [1, n]\right) \\
=: \ & P_1 + \mathbf{P}\left(\max_{1 \leq i \leq n} \xi_i > y\right).
\end{aligned} \tag{4.2}$$

For any $x, u > 0$, define the stopping time

$$T(x,u) = \min\left\{k \in [1,n] : \sum_{i=1}^{k} \eta_i \geq x \text{ and } \Upsilon(S)_k \leq u\right\},$$

with the convention that $\min \emptyset = 0$. Then,

$$\mathbf{1}_{\{S_k \geq x \text{ and } \Upsilon(S)_k \leq u \text{ for some } k \in [1,n]\}} = \sum_{k=1}^{n} \mathbf{1}_{\{T(x,u)=k\}}.$$

By the change of measure (9.2), we deduce that for any $x, \lambda, u > 0$,

$$\begin{aligned} P_1 &= \mathbf{E}_\lambda\Big[Z_{T\wedge n}(\lambda)^{-1}\mathbf{1}_{\{S_k \geq x \text{ and } \Upsilon(S)_k \leq u \text{ for some } k\in[1,n]\}}\Big] \\ &= \sum_{k=1}^{n} \mathbf{E}_\lambda\Big[\exp\Big\{-\lambda\Big(\sum_{i=1}^{k}\eta_i\Big) + \Psi_k(\lambda)\Big\}\mathbf{1}_{\{T(x,u)=k\}}\Big], \end{aligned} \tag{4.3}$$

where

$$\Psi_k(\lambda) = \sum_{i=1}^{k} \log \mathbf{E}\left[\exp\left\{\lambda \eta_i\right\}|\mathcal{F}_{i-1}\right]. \tag{4.4}$$

Set $\lambda = y^{\alpha-1}$. By Lemma 4.1 and the inequality $\log(1+t) \leq t$ for all $t \geq 0$, it is easy to see that for any $x > 0$,

$$\begin{aligned} \Psi_k(\lambda) &\leq \sum_{i=1}^{k} \log\left(1 + \frac{\lambda^2}{2}\mathbf{E}[\eta_i^2 \exp\{\lambda y^{1-\alpha}(\eta_i^+)^\alpha\}|\mathcal{F}_{i-1}]\right) \\ &\leq \sum_{i=1}^{k} \frac{\lambda^2}{2}\mathbf{E}[\eta_i^2 \exp\{\lambda y^{1-\alpha}(\eta_i^+)^\alpha\}|\mathcal{F}_{i-1}] \\ &\leq \frac{1}{2}y^{2\alpha-2}\Upsilon(S)_k. \end{aligned}$$

Using the fact that $\sum_{i=1}^{k}\eta_i \geq x$ and $\Psi_k(\lambda) \leq \frac{1}{2}y^{2\alpha-2}u$ on the set $\{T(x,u)=k\}$, we find that for any $x, u > 0$,

$$\begin{aligned} P_1 &\leq \exp\left\{-\lambda x + \frac{1}{2}y^{2\alpha-2}u\right\}\mathbf{E}_\lambda\Big[\sum_{k=1}^{n}\mathbf{1}_{\{T(x,u)=k\}}\Big] \\ &\leq \exp\left\{-y^{\alpha-1}x + \frac{1}{2}y^{2\alpha-2}u\right\}. \end{aligned}$$

From (4.2), it follows that

$$\begin{aligned} &\mathbf{P}\Big(S_k \geq x \text{ and } \Upsilon(S)_k \leq u \text{ for some } k \in [1,n]\Big) \\ &\qquad\qquad \leq \exp\left\{-y^{\alpha-1}x + \frac{1}{2}y^{2\alpha-2}u\right\} + \mathbf{P}\left(\max_{1\leq i\leq n}\xi_i > y\right). \end{aligned} \tag{4.5}$$

By the exponential Markov inequality, we have the following estimation: for any $y > 0$,

$$\begin{aligned}
\mathbf{P}\left(\max_{1\le i\le n}\xi_i > y\right) &\le \sum_{i=1}^{n}\mathbf{P}\left(\xi_i > y\right)\\
&\le \frac{1}{y^2}\exp\{-y^\alpha\}\sum_{i=1}^{n}\mathbf{E}[\xi_i^2\exp\{(\xi_i^+)^\alpha\}]\\
&\le \frac{C_n}{y^2}\exp\{-y^\alpha\}. \qquad (4.6)
\end{aligned}$$

Taking

$$y = \begin{cases} \left(\dfrac{u}{x}\right)^{1/(1-\alpha)} & \text{if } 0 \le x < u^{1/(2-\alpha)} \\ x & \text{if } x \ge u^{1/(2-\alpha)}, \end{cases}$$

from (4.5) and (4.6), we obtain the desired inequality. This completes the proof of Theorem 2.1. □

Proof of Corollary 2.2. Set $u_n = ||\Upsilon(S)_n||_\infty$. Then $u_n = o(n^{2-\alpha}), n \to \infty$, by the assumptions of Theorem 2.2. For any $x > 0$, by Theorem 2.1, we have

$$\begin{aligned}
\mathbf{P}\left(\max_{1\le k\le n} S_k \ge nx\right) &\le \exp\left\{-(nx)^\alpha\left(1-\frac{u_n}{2\,(nx)^{2-\alpha}}\right)\right\} + \frac{C_n}{(nx)^2}\exp\left\{-(nx)^\alpha\right\}\\
&\le \left(1+\frac{C_n}{(nx)^2}\right)\exp\left\{-(nx)^\alpha\left(1-\frac{u_n}{2\,(nx)^{2-\alpha}}\right)\right\}.
\end{aligned}$$

Since $u_n \ge C_n$, we have $C_n = o(n^{2-\alpha}), n \to \infty$. Hence it holds

$$\limsup_{n\to\infty}\frac{1}{n^\alpha}\log\mathbf{P}\left(\max_{1\le k\le n} S_k \ge nx\right) \le -x^\alpha.$$

This completes the proof of Corollary 2.2. □

Proof of Theorem 2.3. (i) We suppose that $c = 1$ by considering $c^{1/\alpha}\xi_i$ instead of ξ_i. Let $\varepsilon \in (0,1)$ be fixed and consider $\xi_i' = \xi_i(1-\varepsilon)^\beta$ with $\beta > 1/\alpha$. By the assumption of the theorem $\mathbf{P}(\xi_1 \ge x) \le e^{-(1-\varepsilon)x^\alpha}$ for $x > 0$ large enough, so that $\mathbf{P}(\xi_1' \ge x) \le e^{-(1-\varepsilon)^{1-\beta\alpha}x^\alpha}$ for $x > 0$ large enough, and

$$\mathbf{E}[(\xi_1^{'+})^2\exp\{(\xi_1^{'+})^\alpha\}] = \int_0^\infty \mathbf{P}(\xi_1' \ge x)\Big(2x+\alpha x^{1+\alpha}\Big)e^{x^\alpha}dx < \infty.$$

Together with $\mathbf{E}[(\xi_1^{'})^2] < \infty$ this implies that $c_0 := \mathbf{E}[(\xi_1^{'})^2\exp\{(\xi_1^{'+})^\alpha\}] < \infty$. Thus $\Upsilon(S')_n := \sum_{i=1}^n \mathbf{E}[(\xi_1^{'})^2\exp\{(\xi_1^{'+})^\alpha\}] = nc_0 = o(n^{2-\alpha})$ as $n \to \infty$. Hence, from Corollary 2.2 applied to $(\xi_i')_{i\ge1}$ we have for any $x > 0$,

$$\limsup_{n\to\infty}\frac{1}{n^\alpha}\log\mathbf{P}\left((1-\varepsilon)^\beta\max_{1\le k\le n} S_k \ge nx\right) \le -x^\alpha,$$

which implies that for any $x > 0$,

$$\limsup_{n\to\infty} \frac{1}{n^\alpha} \log \mathbf{P}\left(\max_{1\le k\le n} S_k \ge nx \right) \;\le\; -(1-\varepsilon)^{\beta\alpha} x^\alpha \,.$$

Letting $\varepsilon \to 0$ we obtain for any $x > 0$,

$$\limsup_{n\to\infty} \frac{1}{n^\alpha} \log \mathbf{P}\left(\max_{1\le k\le n} S_k \ge nx \right) \;\le\; -x^\alpha \,. \tag{4.7}$$

We now consider the lower bound. By the independence of ξ_i's, it is easy to see that for any $x, \varepsilon > 0$, we have

$$\begin{aligned} \mathbf{P}\left(\max_{1\le k\le n} S_k \ge nx \right) &\ge \mathbf{P}\Big(S_n \ge nx \Big) \\ &\ge \mathbf{P}\Big(\sum_{i=2}^n \xi_i \ge -n\varepsilon, \xi_1 \ge n(x+\varepsilon) \Big) \\ &= \mathbf{P}\Big(\sum_{i=2}^n \xi_i \ge -n\varepsilon \Big) \, \mathbf{P}\Big(\xi_1 \ge n(x+\varepsilon) \Big). \end{aligned}$$

The first probability on the right-hand side converges to 1 as $n \to \infty$ due to the law of large numbers. By (2.11), the second term on the right-hand side has the following lower bound

$$\mathbf{P}\Big(\xi_1 \ge n(x+\varepsilon) \Big) \;\ge\; \exp\Big\{ - \Big(n(x+\varepsilon) \Big)^\alpha (1+\varepsilon) \Big\}$$

for all n large enough. Hence

$$\liminf_{n\to\infty} \frac{1}{n^\alpha} \log \mathbf{P}\left(\frac{S_n}{n} \ge x \right) \;\ge\; -(x+\varepsilon)^\alpha (1+\varepsilon).$$

Letting $\varepsilon \to 0$, we obtain

$$\liminf_{n\to\infty} \frac{1}{n^\alpha} \log \mathbf{P}\left(\frac{S_n}{n} \ge x \right) \;\ge\; -x^\alpha. \tag{4.8}$$

Combining the lower bound (4.8) with the upper bound (4.7), we get (2.12). This ends the proof of part (i).

(ii) The proof of (2.14) is similar to that of part (i). The assertion (2.15) follows from (2.14) applied to $(-\xi_i)_{i\ge1}$. □

5. Proof of Theorem 2.4

To prove Theorem 2.4, we need the following technical lemma.

Lemma 5.1. *Let $p \geq 2$. Assume $\mathbf{E}[|\xi_i|^p] < \infty$ for all $i \in [1, n]$. Set $\eta_i = \xi_i \mathbf{1}_{\{\xi_i \leq y\}}$ for $y > 0$. Then for all $\lambda > 0$,*

$$\mathbf{E}[e^{\lambda \eta_i}|\mathcal{F}_{i-1}] \leq 1 + \frac{1}{2} e^p \lambda^2 \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] + f(y)\mathbf{E}[(\xi_i^+)^p|\mathcal{F}_{i-1}]\,,$$

where the function

$$f(u) = \frac{e^{\lambda u} - 1 - \lambda u}{u^p}, \quad u > 0. \tag{5.1}$$

Proof. We argue as in Fuk and Nagaev [15] (see also Fuk [16]). Using a two term Taylor's expansion, we have for some $\theta \in [0, 1]$,

$$e^{\lambda \eta_i} \;\; \leq \;\; 1 + \lambda \eta_i + \frac{\lambda^2}{2} \eta_i^2 \mathbf{1}_{\{\lambda \eta_i \leq p\}} e^{\lambda \theta \eta_i} + f(\eta_i)(\eta_i^+)^p \mathbf{1}_{\{\lambda \eta_i > p\}}.$$

Remark that the function f is positive and increasing for $\lambda u \geq p$. Since $\mathbf{E}[\eta_i|\mathcal{F}_{i-1}] \leq \mathbf{E}[\xi_i|\mathcal{F}_{i-1}] = 0$ and $\eta_i \leq y$, it follows that

$$\begin{aligned}
\mathbf{E}[e^{\lambda \eta_i}|\mathcal{F}_{i-1}] &\leq 1 + \frac{1}{2} e^p \lambda^2 \mathbf{E}[\eta_i^2|\mathcal{F}_{i-1}] + f(y)\mathbf{E}[(\eta_i^+)^p|\mathcal{F}_{i-1}] \\
&\leq 1 + \frac{1}{2} e^p \lambda^2 \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] + f(y)\mathbf{E}[(\xi_i^+)^p|\mathcal{F}_{i-1}]\,,
\end{aligned}$$

which gives the desired inequality. □

We make use of Lemma 5.1 to prove Theorem 2.4. Set $\eta_i = \xi_i \mathbf{1}_{\{\xi_i \leq y\}}$ for $y > 0$. Define the conjugate probability measure $d\mathbf{P}_\lambda$ by (9.2) and denote by $\mathbf{E}_\lambda$ the expectation with respect to $\mathbf{P}_\lambda$. Since $\xi_i = \eta_i + \xi_i \mathbf{1}_{\{\xi_i > y\}}$, it follows that for any $x, y, u, w > 0$,

$$\begin{aligned}
& \mathbf{P}\left(S_k > x,\ \langle S\rangle_k \leq v \text{ and } \Xi(S)_k \leq w \text{ for some } k \in [1, n]\right) \\
\leq\;& \mathbf{P}\left(\sum_{i=1}^k \eta_i \geq x,\ \langle S\rangle_k \leq v \text{ and } \Xi(S)_k \leq w \text{ for some } k \in [1, n]\right) \\
& + \mathbf{P}\left(\sum_{i=1}^k \xi_i \mathbf{1}_{\{\xi_i > y\}} > 0 \text{ for some } k \in [1, n]\right) \\
=:\;& P_2 + \mathbf{P}\left(\max_{1 \leq i \leq n} \xi_i > y\right).
\end{aligned} \tag{5.2}$$

For any $x, v, w > 0$, define the stopping time $T:$

$$T(x, v, w) = \min\Big\{k \in [1, n] : S_k \geq x,\ \langle S\rangle_k \leq v \text{ and } \Xi(S)_k \leq w\Big\},$$

with the convention that $\min \emptyset = 0$. Then

$$\mathbf{1}_{\{S_k > x,\ \langle S\rangle_k \leq v \text{ and } \Xi(S)_k \leq w \text{ for some } k \in [1, n]\}} = \sum_{k=1}^n \mathbf{1}_{\{T = k\}}.$$

By the change of measure (9.2), we deduce that for any $x, y, \lambda, u, w > 0$,

$$\begin{aligned} P_2 &= \mathbf{E}_\lambda\Big[Z_{T\wedge n}(\lambda)^{-1}\mathbf{1}_{\{S_k>x,\ \langle S\rangle_k\leq v \text{ and } \Xi(S)_k\leq w \text{ for some } k\in[1,n]\}}\Big] \\ &= \sum_{k=1}^{n}\mathbf{E}_\lambda\Big[\exp\Big\{-\lambda\Big(\sum_{i=1}^{k}\eta_i\Big)+\Psi_k(\lambda)\Big\}\mathbf{1}_{\{T=k\}}\Big], \end{aligned}$$

where $\Psi_k(\lambda)$ is defined by (4.4). By Lemma 5.1 and the inequality $\log(1+t)\leq t$ for $t\geq 0$, it is easy to see that for any $x, y, \lambda, u, w > 0$,

$$\begin{aligned} \Psi_k(\lambda) &\leq \sum_{i=1}^{k}\log\left(1+\frac{1}{2}e^p\lambda^2\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}]+f(y)\mathbf{E}[(\xi_i^+)^p|\mathcal{F}_{i-1}]\right) \\ &\leq \sum_{i=1}^{k}\left(\frac{1}{2}e^p\lambda^2\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}]+f(y)\mathbf{E}[(\xi_i^+)^p|\mathcal{F}_{i-1}]\right), \end{aligned}$$

where $f(y)$ is defined by (5.1). By the fact that $\sum_{i=1}^{k}\eta_i\geq x$ and $\Psi_k(\lambda)\leq\frac{1}{2}e^p\lambda^2v+f(y)w$ on the set $\{T=k\}$. we find that for any $x, y, \lambda, u, w > 0$,

$$\begin{aligned} P_2 &\leq \exp\left\{-\lambda x+\frac{1}{2}e^p\lambda^2v+f(y)w\right\}\mathbf{E}_\lambda\Big[\sum_{k=1}^{n}\mathbf{1}_{\{T=k\}}\Big] \\ &\leq \exp\left\{-\lambda x+\frac{1}{2}e^p\lambda^2v+f(y)w\right\}. \end{aligned}$$

Next we carry out an argument as in Fuk and Nagaev [15]. Then

$$P_2 \leq \exp\left\{-\frac{\alpha^2x^2}{2e^pv}\right\}+\exp\left\{-\frac{\beta x}{y}\log\left(1+\frac{\beta xy^{p-1}}{w}\right)\right\}, \tag{5.3}$$

where α and β are defined by (2.18). Combining the inequalities (5.2) and (5.3) together, we obtain the desired inequality. This completes the proof of Theorem 2.4 □

Proof of Corollary 2.5. When $y=\beta x$, from (2.17), it is easy to see that for all $x>0$,

$$\mathbf{P}\left(\max_{1\leq i\leq n}\xi_i>y\right) \leq \sum_{i=1}^{n}\mathbf{P}\Big(\xi_i>\beta x\Big) \leq \frac{1}{\beta^px^p}\sum_{i=1}^{n}\mathbf{E}[|\xi_i|^p] \leq \frac{C_p}{x^p}$$

and

$$\exp\left\{-\frac{\beta x}{y}\log\left(1+\frac{\beta xy^{p-1}}{w}\right)\right\} \leq \frac{w}{\beta xy^{p-1}+w} \leq \frac{w}{\beta^px^p} \leq \frac{C_p}{x^p},$$

where C_p is defined by (2.20). Thus (2.17) implies (2.19). □

6. Proofs of Theorem 2.6 and Corollary 2.7

To prove Theorem 2.6, we need the following inequality whose proof can be found in Fan, Grama and Liu [11] (cf. Corollary 2.3 and Remark 2.1 therein).

Lemma 6.1. *Assume* $\mathbf{E}[\xi_i^2] < \infty$ *for all* $i \in [1, n]$. *Then for all* $x, y, v > 0$,

$$\mathbf{P}\Big(S_k \geq x \ \textit{and} \ \langle S\rangle_k \leq v^2 \ \textit{for some} \ k \in [1, n]\Big)$$
$$\leq \ \exp\left\{-\frac{x^2}{2(v^2 + \frac{1}{3}xy)}\right\} + \mathbf{P}\left(\max_{1\leq i\leq n} \xi_i > y\right).$$

Proof of Theorem 2.6. By Lemma 6.1 and the Markov inequality, it follows that for all $x, y, v > 0$,

$$\begin{aligned}
&\mathbf{P}\Big(S_k \geq x \text{ and } \langle S\rangle_k \leq v^2 \text{ for some } k \in [1, n]\Big) \\
&\quad \leq \ \exp\left\{-\frac{x^2}{2(v^2 + \frac{1}{3}xy)}\right\} + \sum_{i=1}^n \mathbf{P}\Big(\xi_i > y\Big) \\
&\quad \leq \ \exp\left\{-\frac{x^2}{2(v^2 + \frac{1}{3}xy)}\right\} + \frac{1}{y^{p+\delta}} \sum_{i=1}^n \mathbf{E}\Big[|\xi_i|^{p+\delta} \mathbf{1}_{\{\xi_i > y\}}\Big].
\end{aligned}$$

Taking $y = x^{p/(p+\delta)}$ in the last inequality, we obtain the desired inequality. This completes the proof of Theorem 2.6. □

Proof of Corollary 2.7. Notice that $p + \delta > 2$. It is easy to see that for any $x, v > 0$,

$$\begin{aligned}
\mathbf{P}\Big(\max_{1\leq k\leq n} S_k \geq x\Big) \ &\leq \ \mathbf{P}\Big(\max_{1\leq k\leq n} S_k \geq x \ \text{ and } \langle S\rangle_n \leq nv^2\Big) + \mathbf{P}\Big(\langle S\rangle_n > nv^2\Big) \\
&\leq \ \mathbf{P}\Big(S_k \geq x \text{ and } \langle S\rangle_k \leq nv^2 \text{ for some } k \in [1, n]\Big) + \mathbf{P}\Big(\langle S\rangle_n > nv^2\Big) \\
&\leq \ \mathbf{P}\Big(S_k \geq x \text{ and } \langle S\rangle_k \leq nv^2 \text{ for some } k \in [1, n]\Big) \\
&\quad + \frac{\mathbf{E}[|\langle S\rangle_n|^{(p+\delta)/2}]}{n^{(p+\delta)/2} v^{p+\delta}},
\end{aligned}$$

which gives the first desired inequality. By the Hölder inequality, it follows that

$$\sum_{i=1}^n a_i \ \leq \ n^{1-2/(p+\delta)} \Big(\sum_{i=1}^n a_i^{(p+\delta)/2}\Big)^{2/(p+\delta)}, \qquad a_i \geq 0, \ i = 1, ..., n.$$

Hence

$$\Big(\sum_{i=1}^n a_i\Big)^{(p+\delta)/2} \ \leq \ n^{(p-2+\delta)/2} \sum_{i=1}^n a_i^{(p+\delta)/2}, \qquad a_i \geq 0, \ i = 1, ..., n.$$

Then we have

$$\mathbf{E}[|\langle S\rangle_n|^{(p+\delta)/2}] \ \leq \ n^{(p-2+\delta)/2} \sum_{i=1}^n \mathbf{E}\Big[\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}]^{(p+\delta)/2}\Big]$$

$$
\begin{aligned}
&\le n^{(p-2+\delta)/2}\sum_{i=1}^{n}\mathbf{E}\Big[\mathbf{E}[|\xi_i|^{p+\delta}|\mathcal{F}_{i-1}]\Big] \\
&= n^{(p-2+\delta)/2}\sum_{i=1}^{n}\mathbf{E}[|\xi_i|^{p+\delta}].
\end{aligned}
$$

This completes the proof of corollary. □

7. Proofs of Theorems 3.1 - 3.6

From (3.1) and (3.2), it is easy to see that

$$\theta_n - \theta = \sum_{k=1}^{n}\frac{\phi_k\varepsilon_k}{\Sigma_{k=1}^n\phi_k^2}.$$

For any $i = 1, ..., n$, set

$$\xi_i = \frac{\phi_i\varepsilon_i}{\sqrt{\Sigma_{k=1}^n\phi_k^2}} \quad \text{and} \quad \mathcal{F}_i = \sigma\Big\{\phi_k, \varepsilon_k, 1\le k\le i,\ \phi_k^2, 1\le k\le n\Big\}. \tag{7.1}$$

Then $(\xi_i, \mathcal{F}_i)_{i=1,...,n}$ is a sequence of martingale differences, and satisfies

$$S_n = \sum_{i=1}^{n}\xi_i = (\theta_n-\theta)\sqrt{\Sigma_{k=1}^n\phi_k^2}\ .$$

Proof of Theorem 3.1. Notice that

$$\Upsilon(S)_n \ \le\ \sum_{i=1}^{n}\frac{\phi_i^2}{\Sigma_{k=1}^n\phi_k^2}\mathbf{E}[\varepsilon_i^2\exp\{|\varepsilon_i|^\alpha\}|\mathcal{F}_{i-1}] \ \le\ \sum_{i=1}^{n}\frac{\phi_i^2 D}{\Sigma_{k=1}^n\phi_k^2} \ =\ D.$$

Applying Theorem 2.1 to $(\xi_i, \mathcal{F}_i)_{i=1,...,n}$, we find that (2.6), with $u \ge \max\{D, 1\}$, is an upper bound on the tail probabilities $\mathbf{P}\left((\theta_n-\theta)\sqrt{\Sigma_{k=1}^n\phi_k^2}\ge x\right)$.

Similarly, applying Theorem 2.1 to $(-\xi_i, \mathcal{F}_i)_{i=1,...,n}$, we find that (2.6), with $u \ge \max\{D, 1\}$, is also an upper bound on the tail probabilities $\mathbf{P}\left(-(\theta_n-\theta)\sqrt{\Sigma_{k=1}^n\phi_k^2}\ge x\right)$. This completes the proof of Theorem 3.1. □

Proof of Theorem 3.2. By the fact

$$\mathbf{E}[\varepsilon_i^2|\mathcal{F}_{i-1}] = \mathbf{E}[\varepsilon_i^2|\sigma\{\varepsilon_k, 1\le k\le i-1\}] \le E,$$

it follows that

$$\langle S\rangle_n \ \le\ \sum_{i=1}^{n}\frac{\phi_i^2}{(\Sigma_{k=1}^n\phi_k^2)}\mathbf{E}[\varepsilon_i^2|\mathcal{F}_{i-1}] \ \le\ E.$$

Similarly, by the fact $\mathbf{E}[\exp\{|\varepsilon_i|^{\frac{\alpha}{1-\alpha}}\}] \leq F$, it is easy to see that

$$\mathbf{E}[\exp\{|\xi_i|^{\frac{\alpha}{1-\alpha}}\}] \ \leq \ \mathbf{E}[\exp\{|\varepsilon_i|^{\frac{\alpha}{1-\alpha}}\}] \ \leq \ F.$$

Applying Theorem 2.2 of Fan, Grama and Liu [12] to $(\pm\xi_i, \mathcal{F}_i)_{i=1,...,n}$, we obtain the desired inequality. □

Proof of Theorem 3.3. By the fact

$$\mathbf{E}[|\varepsilon_i|^p|\mathcal{F}_{i-1}] = \mathbf{E}[|\varepsilon_i|^p|\sigma\{\varepsilon_k, 1 \leq k \leq i-1\}] \leq A,$$

it follows that

$$\langle S\rangle_n \ \leq \ \sum_{i=1}^n \frac{\phi_i^2}{(\Sigma_{k=1}^n \phi_k^2)} \mathbf{E}[\varepsilon_i^2|\mathcal{F}_{i-1}] \ \leq \ \sum_{i=1}^n \frac{\phi_i^2}{(\Sigma_{k=1}^n \phi_k^2)} \Big(\mathbf{E}[|\varepsilon_i|^p|\mathcal{F}_{i-1}]\Big)^{2/p} = A^{2/p}$$

and

$$\sum_{i=1}^n \mathbf{E}[|\xi_i|^p|\mathcal{F}_{i-1}] \ \leq \ \sum_{i=1}^n \frac{\phi_i^2}{(\Sigma_{k=1}^n \phi_k^2)} \mathbf{E}[|\varepsilon_i|^p|\mathcal{F}_{i-1}] \ \leq \ A.$$

Applying Corollary 2.5 to $(\pm\xi_i, \mathcal{F}_i)_{i=1,...,n}$, we obtain the desired inequality. □

Proof of Theorem 3.4. By the fact

$$\mathbf{E}[\varepsilon_i^2|\mathcal{F}_{i-1}] = \mathbf{E}[\varepsilon_i^2|\sigma\{\varepsilon_k, 1 \leq k \leq i-1\}] \leq A,$$

it follows that

$$\langle S\rangle_n \ \leq \ \sum_{i=1}^n \frac{\phi_i^2}{(\Sigma_{k=1}^n \phi_k^2)} \mathbf{E}[\varepsilon_i^2|\mathcal{F}_{i-1}] \ \leq \ A.$$

Similarly, by the fact $\mathbf{E}[|\varepsilon_i|^{p+\delta}] \leq B$, it follows that

$$\sum_{i=1}^n \mathbf{E}[|\xi_i|^{p+\delta}] \ = \ \sum_{i=1}^n \mathbf{E}\Bigg[\Big|\frac{\phi_i^2}{\Sigma_{k=1}^n \phi_k^2}\Big|^{\frac{p+\delta}{2}}\Bigg] \mathbf{E}[|\varepsilon_i|^{p+\delta}] \ \leq \ \mathbf{E}\Bigg[\sum_{i=1}^n \frac{\phi_i^2}{\Sigma_{k=1}^n \phi_k^2}\Bigg] B \ = \ B.$$

Applying Theorem 2.6 to $(\pm\xi_i, \mathcal{F}_i)_{i=1,...,n}$, we obtain the desired inequality. □

Proof of Theorem 3.5. Let $p \in [1, 2]$. By the inequality

$$\Big(\sum_{i=1}^n a_i\Big)^\alpha \leq \sum_{i=1}^n a_i^\alpha, \quad a_i \geq 0 \ \text{ and } \ \alpha \in (0, 1],$$

we have

$$\sum_{i=1}^n \mathbf{E}[|\xi_i|^p] \ = \ \sum_{i=1}^n \mathbf{E}\Bigg[\frac{(\phi_i^2)^{p/2}}{(\Sigma_{k=1}^n \phi_k^2)^{p/2}}\Bigg] \mathbf{E}[|\varepsilon_i|^p] \ \leq \ \mathbf{E}\Bigg[\frac{(\sum_{i=1}^n \phi_i^2)^{p/2}}{(\Sigma_{k=1}^n \phi_k^2)^{p/2}}\Bigg] A \ = \ A.$$

By the inequality of von Bahr and Esseen (cf. Theorem 2 of [39]), we get

$$\mathbf{E}[|S_n|^p] \ \leq \ 2\sum_{i=1}^n \mathbf{E}[|\xi_i|^p] \ \leq \ 2A.$$

Then for all $x > 0$,

$$\mathbf{P}\Big(\pm (\theta_n - \theta)\sqrt{\Sigma_{k=1}^n \phi_k^2} \geq x \Big) \ = \ \mathbf{P}\Big(\pm S_n \geq x \Big) \ \leq \ \frac{\mathbf{E}[|S_n|^p]}{x^p} \ \leq \ \frac{2A}{x^p}.$$

This completes the proof of theorem. □

Proof of Theorem 3.6. It is obvious that

$$\frac{\theta_n - \theta}{\sigma}\sqrt{\Sigma_{k=1}^n \phi_k^2} = \sum_{i=1}^n \eta_i,$$

where $\eta_i = \xi_i/\sigma$. Notice that $\mathbf{E}[\varepsilon_i^2|\mathcal{F}_{i-1}] = \mathbf{E}[\varepsilon_i^2 \ |\sigma\{\varepsilon_j, j \leq i-1\}] = \sigma^2$ a.s.. Then we have

$$\sum_{i=1}^n \mathbf{E}[\eta_i^2|\mathcal{F}_{i-1}] \ = \ \frac{\langle S\rangle_n}{\sigma^2} \ = \ \sum_{i=1}^n \frac{\phi_i^2}{(\Sigma_{k=1}^n \phi_k^2)} \frac{\mathbf{E}[\varepsilon_i^2|\mathcal{F}_{i-1}]}{\sigma^2} \ = \ \sum_{i=1}^n \frac{\phi_i^2}{\Sigma_{k=1}^n \phi_k^2} \ = \ 1$$

and

$$\sum_{i=1}^n \mathbf{E}[|\eta_i|^p|\mathcal{F}_{i-1}] \ \leq \ \sum_{i=1}^n \mathbf{E}\Bigg[\Bigg|\frac{\phi_i}{\sqrt{\Sigma_{k=1}^n \phi_k^2}}\Bigg|^p\Bigg] \frac{\mathbf{E}[|\varepsilon_i|^p|\mathcal{F}_{i-1}]}{\sigma^p} \ \leq \ \frac{A}{\sigma^p} \sum_{i=1}^n \mathbf{E}\Bigg[\Bigg|\frac{\phi_i}{\sqrt{\Sigma_{k=1}^n \phi_k^2}}\Bigg|^p\Bigg].$$

Applying inequality (3.15) to the martingale difference sequence $(\eta_i, \mathcal{F}_i)_{i=1,...,n}$ with $\delta = p-2$, we obtain the desired inequality. □

8. Proof of tightness in Theorem 3.7

By Theorem 8.4, Chapter 3, Section 8 of Billingsley [2] (see also Chapter 3, Section 16 for the convergence in the space), we only need to show that for any $\varepsilon > 0$, there exist a $\lambda > 1$ and an integer n_0 such that for every $n \geq n_0$,

$$\mathbf{P}\Big(\max_{k \leq i \leq k+n} |S_i - S_k| \geq \lambda\sqrt{n} \Big) \ \leq \ \frac{\varepsilon}{\lambda^2}. \tag{8.1}$$

Since

$$\mathbf{E}[|\xi_i|^3|\mathcal{F}_{i-1}] \leq M\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}],$$

we deduce that

$$(\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}])^{3/2} \leq \mathbf{E}[|\xi_i|^3|\mathcal{F}_{i-1}] \leq M\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}]. \tag{8.2}$$

This, in turn, implies that

$$\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] \leq M^2, \qquad \langle S\rangle_{k+n} - \langle S\rangle_k \leq nM^2 \quad \text{and} \quad \sum_{i=k}^{k+n} \mathbf{E}[|\xi_i|^3|\mathcal{F}_{i-1}] \leq nM^3.$$

Applying (2.17) with $p=3, x=2y=\lambda\sqrt{n}$, we obtain, by (8.2),

$$
\begin{aligned}
& \mathbf{P}\Big(\max_{k\leq i\leq k+n} |S_i - S_k| \geq \lambda\sqrt{n}\Big) \\
\leq\ & 2\exp\Big\{-\frac{2\,\lambda^2}{25e^3M^2}\Big\} + 2\exp\Big\{-\frac{6}{5}\log\Big(1+\frac{3\lambda^3\sqrt{n}}{20M^3}\Big)\Big\} \\
& \quad + \mathbf{P}\Big(\max_{k\leq i\leq k+n} \xi_i > \frac{1}{2}\lambda\sqrt{n}\Big) + \mathbf{P}\Big(\max_{k\leq i\leq k+n} (-\xi_i) > \frac{1}{2}\lambda\sqrt{n}\Big) \\
\leq\ & 2\exp\Big\{-\frac{4\,\lambda^2}{50e^3M^2}\Big\} + 2\Big(\frac{3\lambda^3\sqrt{n}}{20M^3}\Big)^{-6/5} + \frac{16}{\lambda^3 n^{3/2}}\sum_{i=k}^{k+n}\mathbf{E}\Big[|\xi_i|^3\Big] \\
\leq\ & \frac{\varepsilon}{\lambda^2},
\end{aligned}
$$

provided that λ is sufficiently large, which proves (8.1). □

9. Proof of Theorem 3.8

For any $i=1,...,n$, set

$$
\eta_i = \frac{\xi_i}{V_n(\beta)}, \quad \mathcal{F}_0 = \sigma\Big(|\xi_j|, 1\leq j\leq n\Big) \ \text{ and } \ \mathcal{F}_i = \sigma\Big(\xi_k, 1\leq k\leq i,\ |\xi_j|, 1\leq j\leq n\Big). \tag{9.1}
$$

Since $(\xi_i)_{i=1,...,n}$ are independent and symmetric, then

$$
\mathbf{E}[\xi_i > y\mid \mathcal{F}_{i-1}] = \mathbf{E}[\xi_i > y\mid |\xi_i|] = \mathbf{E}[-\xi_i > y\mid |-\xi_i|] = \mathbf{E}[-\xi_i > y\mid \mathcal{F}_{i-1}].
$$

Thus $(\eta_i, \mathcal{F}_i)_{i=1,...,n}$ is a sequence of conditionally symmetric martingale differences, i.e.

$$
\mathbf{E}[\eta_i > y\,|\,\mathcal{F}_{i-1}] = \mathbf{E}[-\eta_i > y\,|\,\mathcal{F}_{i-1}].
$$

It is easy to see that

$$
\frac{S_n}{\mathrm{V}_n(\beta)} = \sum_{i=1}^n \eta_i
$$

is a sum of martingale differences, and that $(\eta_i, \mathcal{F}_i)_{i=1,...,n}$ satisfies

$$
\sum_{i=1}^n \mathbf{E}[|\eta_i|^\beta|\,\mathcal{F}_{i-1}] = \sum_{i=1}^n |\eta_i|^\beta \ = \ \sum_{i=1}^n \frac{|\xi_i|^\beta}{V_n(\beta)^\beta} \ = \ 1.
$$

For any $x>0$, define the stopping time T:

$$
T(x) = \min\Big\{k\in[1,n] : \sum_{i=1}^k \eta_i \geq x\Big\},
$$

with the convention that $\min\emptyset = 0$. Then it follows that

$$\mathbf{1}_{\left\{\max_{1\leq k\leq n} S_k/V_n(\beta)\geq x\right\}} = \sum_{k=1}^{n} \mathbf{1}_{\{T(x)=k\}}.$$

For any nonnegative number λ, define the martingale $M(\lambda) = (M_k(\lambda), \mathcal{F}_k)_{k=0,...,n}$, where

$$M_k(\lambda) = \prod_{i=1}^{k} \frac{\exp\{\lambda\eta_i\}}{\mathbf{E}\left[\exp\{\lambda\eta_i\}\,|\mathcal{F}_{i-1}\right]}, \qquad M_0(\lambda) = 1.$$

Since T is a stopping time, then $M_{T\wedge k}(\lambda)$, $\lambda > 0$, is also a martingale. Define the conjugate probability measure $\mathbf{P}_\lambda$ on $(\Omega, \mathcal{F})$:

$$d\mathbf{P}_\lambda = M_{T\wedge n}(\lambda)d\mathbf{P}. \tag{9.2}$$

Denote $\mathbf{E}_\lambda$ the expectation with respect to $\mathbf{P}_\lambda$. Denote by

$$\Psi_k(\lambda) = \sum_{i=1}^{k} \log \mathbf{E}\left[\exp\{\lambda\eta_i\}\Big|\mathcal{F}_{i-1}\right], \qquad k \in [1, n].$$

Using the change of probability measure (9.2), we have for all $x > 0$,

$$\begin{aligned}
\mathbf{P}\left(\max_{1\leq k\leq n}\frac{S_k}{V_n(\beta)} \geq x\right) &= \mathbf{E}_\lambda\left[M_{T\wedge n}(\lambda)^{-1}\mathbf{1}_{\left\{\max_{1\leq k\leq n} S_k/V_n(\beta)\geq x\right\}}\right] \\
&= \sum_{k=1}^{n}\mathbf{E}_\lambda\left[\exp\left\{-\lambda\sum_{i=1}^{k}\eta_i + \Psi_k(\lambda)\right\}\mathbf{1}_{\{T(x)=k\}}\right] \\
&\leq \sum_{k=1}^{n}\mathbf{E}_\lambda\left[\exp\left\{-\lambda x + \Psi_k(\lambda)\right\}\mathbf{1}_{\{T(x)=k\}}\right],
\end{aligned} \tag{9.3}$$

where the last line follows by the fact that $\sum_{i=1}^{k}\eta_i \geq x$ on the set $\{T(x) = k\}$. Since $(\eta_i, \mathcal{F}_i)_{i=1,...,n}$ is conditionally symmetric, one has

$$\mathbf{E}\left[\exp\{\lambda\eta_i\}\Big|\mathcal{F}_{i-1}\right] = \mathbf{E}\left[\exp\{-\lambda\eta_i\}\Big|\mathcal{F}_{i-1}\right],$$

and thus it holds

$$\mathbf{E}\left[\exp\{\lambda\eta_i\}\Big|\mathcal{F}_{i-1}\right] = \mathbf{E}\left[\cosh(\lambda\eta_i)\Big|\mathcal{F}_{i-1}\right].$$

Notice that η_i^2 is $\mathcal{F}_{i-1}-$measurable, and that

$$\cosh(x) = \sum_{k=0}^{\infty}\frac{1}{(2k)!}x^{2k}.$$

Thus

$$\Psi_k(\lambda) = \sum_{i=1}^{k}\log\Big(\cosh(\lambda\eta_i)\Big).$$

Let t_0 be a number such that

$$\frac{1}{2}\Big(e^{x-t_0|x|^\beta}+e^{-x-t_0|x|^\beta}\Big)\leq 1 \qquad \text{for all } x\in \mathbf{R}. \tag{9.4}$$

Then we have

$$\Psi_k(\lambda)\leq \sum_{i=1}^k \log\Big(e^{t_0|\lambda\eta_i|^\beta}\Big)=t_0\lambda^\beta\sum_{i=1}^k|\eta_i|^\beta\leq t_0\lambda^\beta.$$

and

$$\begin{aligned}\mathbf{P}\bigg(\max_{1\leq k\leq n}\frac{S_k}{V_n(\beta)}\geq x\bigg) &\leq \inf_{\lambda>0}\exp\Big\{-\lambda x+t_0\lambda^\beta\Big\}\\ &= \exp\Big\{-C(\beta,t_0)x^{\frac{\beta}{\beta-1}}\Big\}.\end{aligned} \tag{9.5}$$

This completes the proof.

Acknowledgements

The authors would like to thank the anonymous referee for valuable comments and remarks. The work has been partially supported by the National Natural Science Foundation of China (Grants no. 11601375, no. 11501146, no. 11571052 and no. 11401590).

[1] Bercu, B., Touati, A., 2008. Exponential inequalities for self-normalized martingales with applications. *Ann. Appl. Probab.* **18**(5): 1848–1869.

[2] Billingsley, P. *Convergence of Probability Measures.* John Wiley: New York, 1968.

[3] Borovkov, A. A., 2000. Estimates for the distribution of sums and maxima of sums of random variables when the Cramer condition is not satisfied. *Siberian Math. J.* **41**(5): 811–848.

[4] Borovkov, A. A., 2000. Probabilities of large deviations for random walks with semi-exponential distributions. *Siberian Math. J.* **41**(6): 1061–1093.

[5] Chung, F., Lu, L., 2006. Concentration inequalities and martingale inequalities: a survey. *Internet Math.* **3**(1): 79–127.

[6] Dedecker, J., Fan, X., 2015. Deviation inequalities for separately Lipschitz functionals of iterated random functions, *Stochastic Process. Appl.* **125**(1): 60–90.

[7] Dedecker, J., Merlevède, F., 2011. Rates of convergence in the central limit theorem for linear statistics of martingale differences. *Stochastic Process. Appl.* **121**(5): 1013–1043.

[8] de la Peña, V. H., 1999. A general class of exponential inequalities for martingales and ratios. *Ann. Probab.* **27** (1): 537–564.

[9] de la Peña, V. H., Giné, E., 1999. *Decoupling: from dependence to independence.* Springer.

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

[11] Fan, X., Grama, I., Liu, Q., 2012. Hoeffding's inequality for supermartingales. *Stochastic Process. Appl.* **122**(10): 3545–3559.

[12] Fan, X., Grama, I., Liu, Q., 2012. Large deviation exponential inequalities for supermartingales. *Electron. Commun. Probab.* **17**(59): 1–8.

[13] Fan, X., Grama, I. and Liu, Q., 2015. Exponential inequalities for martingales with applications. *Electron. J. Probab.* **20**(1): 1–22.

[14] Freedman, D.A., 1975. On tail probabilities for martingales. *Ann. Probab.* **3**, No. 1, 100–118.

[15] Fuk, D. Kh., Nagaev, S. V., 1971. Probabilistic inequalities for partial sums of independent variables. *Theory Probab. Appl.* **16**(4): 660–675.

[16] Fuk, D. Kh., 1973. Some probabilistic inequalities for martingales. *Siberian. Math. J.* **14**(1): 185–193.

[17] Gantert, N., Ramanan, K., Rembart, F., 2014. Large deviations for weighted sums of stretched exponential random variables. *Electron. Commun. Probab.* **19**(41): 1–14.

[18] Grama, I., Haeusler, E., 2000. Large deviations for martingales via Cramér's method. *Stochastic Process. Appl.* **85**(2): 279–293.

[19] Haeusler, E., 1984. An exact rate of convergence in the functional central limit theorem for special martingale difference arrays. *Probab. Theory Relat. Fields* **65**(4): 523–534.

[20] Haeusler, E., Joos, K., 1988. A nonuniform bound on the rate of convergence in the martingale central limit theorem. *Ann. Probab.* **16**(4): 1699–1720.

[21] Hall, P., Heyde, C. C., 1980. *Martingale Limit Theory and Its Application.* Academic Press.

[22] Hao, S., Liu, Q., 2014. Convergence rates in the lax of large numbers for arrays of martingale differences. *J. Math. Anal. Appl.* **417**(2): 733–773.

[23] Huang, C., Liu, Q., 2012. Moments, moderate and large deviations for a branching process in a random environment. *Stoch. Proc. Appl.* **122**: 522–545.

[24] Jing, B. Y., Liang, H. Y., Zhou, W., 2012. Self-normalized moderate deviations for independent random variables. *Sci. China Math.* **55**(11): 2297–2315

[25] Kiesel, R., Stadtmüller, U., 2000. A large deviation principle for weighted sums of independent and identically distributed random variables. *J. Math. Anal. Appl.* **251**(2): 929–939.

[26] Lanzinger, H., Stadtmüller, U., 2000. Maxima of increments of partial sums for certain subexponential distributions. *Stochastic Process. Appl.* **86**(2): 307–322.

[27] Lesigne, E., Volný, D., 2001. Large deviations for martingales. *Stochastic Process. Appl.* **96**(1): 143–159.

[28] Liptser, R., Shiryaev, A., 1989. *Theory of Martingales.* Kluwer Academic Publishers.

[29] Liptser, R., Spokoiny, V., 2000. Deviation probability bound for martingales with applications to statistical estimation. *Statist. Probab. Lett.* **46**(4): 347–357.

[30] Liu, Q., Watbled, F., 2009. Exponential inequalities for martingales and asymptotic properties of the free energy of directed polymers in a random environment. *Stochastic Process. Appl.* **119**(10): 3101–3132.

[31] McDiarmid, C., 1989. On the method of bounded differences. *Surveys in combi.*

[32] Nagaev, A. V., 1969. Integral limit theorems taking into account large deviations when Cramér's condition does not hold. I. *Theory Probab. Appl.* **14**(1): 51–63.

[33] Nagaev, S. V., 1979. Large deviations of sums of independent random variables. *Ann. Probab.* **7**: 745–789.

[34] Ouchti, L., 2005. On the rate of convergence in the central limit theorem for martingale difference sequences. *Ann. Inst. Henri Poincaré Probab. Stat.* **41**(1): 35–43.

[35] Rio, E., 2013. On McDiarmid's concentration inequality. *Electron. Commun. Probab.* **18**(44): 1–11.

[36] Saulis, L., Statulevičius, V. A., 1978. *Limit theorems for large deviations.* Kluwer Academic Publishers.

[37] Shao, Q. M., 1997. Self-normalized large deviations. *Ann. Probab.* **25**(1): 285–328.

[38] van de Geer, S., 1995. Exponential inequalities for martingales, with application to maximum likelihood estimation for counting process. *Ann. Statist.* **23**, 1779–1801.

[39] von Bahr, B., Esseen, C. G., 1965. Inequalities for the rth absolute moment of a sum of random variables, $1 \leqq r \leqq 2$. *Ann. Math. Statist.* **36**(1): 299–303.