

# Exponential inequalities for martingales with applications

Xiequan Fan, Ion Grama, Quansheng Liu

► **To cite this version:**

Xiequan Fan, Ion Grama, Quansheng Liu. Exponential inequalities for martingales with applications. Electronic Journal of Probability, Institute of Mathematical Statistics (IMS), 2015, 20, pp.1 - 22. 10.1214/EJP.v20-3496 . hal-01108032

**HAL Id: hal-01108032**

**<https://hal.inria.fr/hal-01108032>**

Submitted on 22 Jan 2015

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

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

## Exponential inequalities for martingales with applications

Xiequan Fan\*      Ion Grama†      Quansheng Liu†

### Abstract

The paper is devoted to establishing some general exponential inequalities for supermartingales. The inequalities improve or generalize many exponential inequalities of Bennett, Freedman, de la Peña, Pinelis and van de Geer. Moreover, our concentration inequalities also improve some known inequalities for sums of independent random variables. Applications associated with linear regressions, autoregressive processes and branching processes are provided. In particular, an interesting application of de la Peña’s inequality to self-normalized deviations is also provided.

**Keywords:** exponential inequality; martingales; changes of probability measure; Freedman’s inequality; de la Peña’s inequality; Pinelis’ inequality; Bernstein’s inequality; linear regressions; autoregressive processes.

**AMS 2010 Subject Classification:** 60E15; 60G42.

Submitted to EJP on May 6, 2014, final version accepted on December 27, 2014.

## 1 Introduction

Assume that we are given a sequence of real-valued supermartingale differences  $(\xi_i, \mathcal{F}_i)_{i=0, \dots, n}$  defined on some probability space  $(\Omega, \mathcal{F}, \mathbf{P})$ , where  $\xi_0 = 0$  and  $\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq \dots \subseteq \mathcal{F}_n \subseteq \mathcal{F}$  are increasing  $\sigma$ -fields. So we have  $\mathbf{E}(\xi_i | \mathcal{F}_{i-1}) \leq 0$ ,  $i = 1, \dots, n$ , by definition. Set

$$S_k = \sum_{i=1}^k \xi_i, \quad k = 1, \dots, n. \quad (1.1)$$

Then  $S = (S_k, \mathcal{F}_k)_{k=1, \dots, n}$  is a supermartingale. Let  $\langle S \rangle$  and  $[S]$  be respectively the quadratic characteristic and the squared variation of the supermartingale  $S$ :

$$\langle S \rangle_k = \sum_{i=1}^k \mathbf{E}(\xi_i^2 | \mathcal{F}_{i-1}) \quad \text{and} \quad [S]_k = \sum_{i=1}^k \xi_i^2. \quad (1.2)$$

\*Regularity Team, Inria and MAS Laboratory, Ecole Centrale Paris - Grande Voie des Vignes, 92295 Châtenay-Malabry, France. Corresponding author. E-mail: fanxiequan@hotmail.com

†Université de Bretagne-Sud, UMR 6205, LMBA, 56000 Vannes, France. E-mail: ion.grama@univ-ubs.fr, quansheng.liu@univ-ubs.fr

The following exponential inequality for supermartingales can be found in Freedman [16].

**Theorem A.** Suppose  $\xi_i \leq \epsilon$  for a positive constant  $\epsilon$ . Then, for all  $x, v > 0$ ,

$$\begin{aligned} \mathbf{P}(S_k \geq x \text{ and } \langle S \rangle_k \leq v^2 \text{ for some } k) &\leq B_2(x, \epsilon, v) \\ &:= \exp \left\{ -\frac{x^2}{2(v^2 + x\epsilon)} \right\}. \end{aligned} \quad (1.3)$$

After Freedman's seminal work, many interesting exponential inequalities for martingales have been established. For continuous-time martingales with bounded jumps, Freedman's inequality (1.3) has been established by Shorack and Wellner [34]. By imposing certain moment conditions, van de Geer [35] relaxed the condition of Shorack and Wellner and generalized inequality (1.3) for martingales with non-bounded jumps. Under the following conditional Bernstein condition: for a positive constant  $\epsilon$ ,

$$\mathbf{E}(|\xi_i|^l | \mathcal{F}_{i-1}) \leq \frac{1}{2} l! \epsilon^{l-2} \mathbf{E}(\xi_i^2 | \mathcal{F}_{i-1}), \quad \text{for all } l \geq 2, \quad (1.4)$$

de la Peña [8] have obtained the following Bernstein type inequality for martingales, for all  $x, v > 0$ ,

$$\begin{aligned} \mathbf{P}(S_k \geq x \text{ and } \langle S \rangle_k \leq v^2 \text{ for some } k) &\leq B_1(x, \epsilon, v) \\ &:= \exp \left\{ -\frac{x^2}{v^2(1 + \sqrt{1 + 2x\epsilon/v^2}) + x\epsilon} \right\} \\ &\leq B_2(x, \epsilon, v). \end{aligned} \quad (1.5) \quad (1.6)$$

Inequality (1.6) has also been obtained by van de Geer [35]. In particular, when  $(\xi_i)_{i=1, \dots, n}$  are independent, the inequalities (1.5) and (1.6) reduce, respectively, to the inequalities of Bennett [2] and Bernstein [6]. Many other generalizations of Freedman's inequality can be found in Haeusler [18], Pinelis [28], Dzharaparidze and van Zanten [13], Delyon [12] and Khan [22].

Following the work of Freedman [16], Shorack and Wellner [34], van de Geer [35] and de la Peña [8], we develop some new methods, based on changes of probability measure, for establishing some general exponential inequalities for supermartingales. The methods are user-friendly and efficient.

In Theorem 2.1, we obtain two exponential inequalities for supermartingales under a very general condition. Assume that

$$\mathbf{E}(\exp \{ \lambda \xi_i - g(\lambda) \xi_i^2 \} | \mathcal{F}_{i-1}) \leq 1 + f(\lambda) V_{i-1}$$

for some  $\lambda \in (0, \infty)$ , for two non-negative functions  $f(\lambda)$  and  $g(\lambda)$ , and for some non-negative and  $\mathcal{F}_{i-1}$ -measurable random variables  $V_{i-1}$ . Then, for all  $x, v, \omega > 0$ ,

$$\begin{aligned} &\mathbf{P} \left( S_k \geq x, [S]_k \leq v^2 \text{ and } \sum_{i=1}^k V_{i-1} \leq w \text{ for some } k \in [1, n] \right) \\ &\leq \exp \left\{ -\lambda x + g(\lambda) v^2 + n \log \left( 1 + \frac{f(\lambda)}{n} w \right) \right\} \end{aligned} \quad (1.7)$$

$$\leq \exp \left\{ -\lambda x + g(\lambda) v^2 + f(\lambda) w \right\}. \quad (1.8)$$

If  $\xi_i \geq -\epsilon$  for a positive constant  $\epsilon$ , then our result (1.8) implies that, for all  $x, v > 0$ ,

$$\mathbf{P}(S_k \geq x \text{ and } [S]_k \leq v^2 \text{ for some } k) \leq B_2(x, \epsilon, v). \quad (1.9)$$

This inequality is similar to the one of Freedman (1.3). To highlight the differences between (1.3) and (1.9), notice that the conditions  $\xi_i \leq \epsilon$  and conditional variance  $\langle S \rangle_k$  in Freedman’s inequality (1.3) are respectively replaced by the condition  $\xi_i \geq -\epsilon$  and squared variation  $[S]_k$  in our inequality (1.9). Moreover, inequality (1.9) completes Freedman’s inequality (1.3) by giving an estimation of deviation probabilities on the left side: if the martingale differences  $(\xi_i, \mathcal{F}_i)_{i=1, \dots, n}$  satisfy  $\xi_i \leq \epsilon$  for all  $i$ , then, for all  $x, v > 0$ ,

$$\mathbf{P}(S_k \leq -x \text{ and } [S]_k \leq v^2 \text{ for some } k) \leq B_2(x, \epsilon, v). \tag{1.10}$$

If the martingale differences verifies canonical assumption (which means  $g(\lambda) = \lambda^2/2$  and  $f(\lambda) = 0$ ), then (1.8) implies the following de la Peña inequality [8], for all  $x, v > 0$ ,

$$\mathbf{P}(S_k \geq x \text{ and } [S]_k \leq v^2 \text{ for some } k) \leq \exp\left\{-\frac{x^2}{2v^2}\right\}. \tag{1.11}$$

Moreover, we find that (1.11) implies the following self-normalized deviation result associated with independent and symmetric random variables, for all  $x > 0$ ,

$$\mathbf{P}\left(\max_{1 \leq k \leq n} \frac{S_k}{\sqrt{[S]_n}} \geq x\right) \leq \exp\left\{-\frac{x^2}{2}\right\}. \tag{1.12}$$

If  $\mathbf{E}|\xi_i|^3 < \infty$ , then (1.8) implies the following Bernstein type inequality, for all  $x, v, w > 0$ ,

$$\mathbf{P}(S_k \geq x, [S]_k \leq v^2 \text{ and } \Upsilon(S_k) \leq w \text{ for some } k) \leq B_1\left(x, \frac{w}{3v^2}, v\right) \tag{1.13}$$

$$\leq B_2\left(x, \frac{w}{3v^2}, v\right), \tag{1.14}$$

where

$$\Upsilon(S_k) = \sum_{i=1}^k \mathbf{E}(|\xi_i|^3 | \mathcal{F}_{i-1});$$

see Corollary 2.2. Compared to the inequalities (1.5) and (1.6), the advantage of the last two inequalities (1.13) and (1.14) is that we do not assume the existence of moments of all orders.

Assume that  $\mathbf{E}(e^{\lambda \xi_i} | \mathcal{F}_{i-1}) \leq 1 + f(\lambda)\mathbf{E}(\xi_i^2 | \mathcal{F}_{i-1})$  for some  $\lambda \in (0, \infty)$  and a positive function  $f(\lambda)$ . Then Theorem 2.1 implies that, for all  $x, v > 0$ ,

$$\begin{aligned} & \mathbf{P}(S_k \geq x \text{ and } \langle S \rangle_k \leq v^2 \text{ for some } k \in [1, n]) \\ & \leq \exp\left\{-\lambda x + n \log\left(1 + \frac{f(\lambda)}{n} v^2\right)\right\} \end{aligned} \tag{1.15}$$

$$\leq \exp\left\{-\lambda x + f(\lambda)v^2\right\}. \tag{1.16}$$

In particular, if  $(\xi_i, \mathcal{F}_i)_{i=1, \dots, n}$  satisfies condition (1.4), then it holds

$$\mathbf{E}(e^{\bar{\lambda} \xi_i} | \mathcal{F}_{i-1}) \leq 1 + f(\bar{\lambda})\mathbf{E}(\xi_i^2 | \mathcal{F}_{i-1}),$$

where

$$\bar{\lambda} = \frac{2x/v^2}{2x\epsilon/v^2 + 1 + \sqrt{1 + 2x\epsilon/v^2}} \quad \text{and} \quad f(\lambda) = \frac{\lambda^2}{2(1 - \lambda\epsilon)}.$$

Inequality (1.16) reduces to de la Peña’s inequality (1.5) with  $\lambda = \bar{\lambda}$ . Hence, our bound (1.15) with  $\lambda = \bar{\lambda}$  improves de la Peña’s inequality (1.5). In the i.i.d. case, bound (1.15)

significantly improves the large deviation bound (1.5) on large deviation tail probabilities  $\mathbf{P}(S_n \geq nx)$  by adding a factor with exponentially decay rate  $\exp\{-nc_x\}$ , where  $c_x > 0$  does not depend on  $n$ . In the applications for linear regression models, we find that such type refinements are useful; see Theorem 3.1.

In Theorem 2.6, we consider the case that supermartingale has sub-Gaussian differences. Assume that  $\mathbf{E}(e^{\lambda\xi_i}|\mathcal{F}_{i-1}) \leq \exp\{f(\lambda)V_{i-1}\}$  for some  $\lambda \in (0, \infty)$ , for a positive function  $f(\lambda)$  and for some  $\mathcal{F}_{i-1}$ -measurable random variables  $V_{i-1}$ . Then, for all  $x, v > 0$ ,

$$\mathbf{P}\left(S_k \geq x \text{ and } \sum_{i=1}^k V_{i-1} \leq v^2 \text{ for some } k\right) \leq \exp\left\{-\lambda x + f(\lambda)v^2\right\}. \quad (1.17)$$

In particular, when the function  $f(\lambda) = \lambda^2/2$  for all  $\lambda > 0$  and  $(V_i)_{i=1,\dots,n}$  are constants, inequality (1.17) reduces to Fuk's inequality with  $\lambda = \lambda(x) := x/\sum_{i=1}^n V_i^2$  (cf. Theorem 4 of [17]). Thus (1.17) is a generalization of Fuk's inequality [17] for supermartingales. If  $V_{i-1} = \mathbf{E}(\xi_i^2|\mathcal{F}_{i-1})$  is the conditional variance, inequality (1.17) reduces to Theorem 4.2 of Khan [22]. Inequality (1.17) implies the following result, where  $V_{i-1}$  is not the conditional variance. If  $\xi_i \leq U_{i-1}$  for some  $\mathcal{F}_{i-1}$ -measurable random variables  $U_{i-1}$ , then, for all  $x, v > 0$ ,

$$\mathbf{P}\left(S_k \geq x \text{ and } \sum_{i=1}^k C_{i-1}^2 \leq v^2 \text{ for some } k\right) \leq \exp\left\{-\frac{x^2}{2v^2}\right\}, \quad (1.18)$$

where

$$C_{i-1}^2 = \begin{cases} \mathbf{E}(\xi_i^2|\mathcal{F}_{i-1}), & \text{if } \mathbf{E}(\xi_i^2|\mathcal{F}_{i-1}) \geq U_{i-1}^2, \\ \frac{1}{4}\left(U_{i-1} + \frac{\mathbf{E}(\xi_i^2|\mathcal{F}_{i-1})}{U_{i-1}}\right)^2, & \text{otherwise.} \end{cases} \quad (1.19)$$

Then we show that (1.18) implies a generalization of Azuma-Hoeffding's inequality for martingales due to van de Geer [36]. Moreover, we also show that (1.18) significantly improves some recent inequalities of Bentkus [3] and Pinelis [29, 30] by adding an exponential decay factor in the case of  $\|\sum_{i=1}^n C_{i-1}^2\|_\infty < \sum_{i=1}^n \|C_{i-1}^2\|_\infty$ ; see (2.23) and *Example 1* for details. We find that such improvements are important in the applications for linear regression models and autoregressive processes; see Remarks 3.3 and 3.7.

The paper is organized as follows. We present our theoretical results in Section 2, give the applications of our results in Section 3 and devote to the proofs of our results in Sections 4 - 6. The proofs of the theorems and their corollaries are in the same sections.

## 2 Main results

Our first result is given under a very general condition.

**Theorem 2.1.** *Assume that  $V_{i-1}, i \in [1, n]$ , are non-negative and  $\mathcal{F}_{i-1}$ -measurable random variables. Suppose that*

$$\mathbf{E}(\exp\{\lambda\xi_i - g(\lambda)\xi_i^2\}|\mathcal{F}_{i-1}) \leq 1 + f(\lambda)V_{i-1} \quad (2.1)$$

*for some  $\lambda \in (0, \infty)$ , for two non-negative functions  $f(\lambda)$  and  $g(\lambda)$ , and for all  $i \in [1, n]$ . Then, for all  $x, v, \omega > 0$ ,*

$$\begin{aligned} & \mathbf{P}\left(S_k \geq x, [S]_k \leq v^2 \text{ and } \sum_{i=1}^k V_{i-1} \leq \omega \text{ for some } k \in [1, n]\right) \\ & \leq \exp\left\{-\lambda x + g(\lambda)v^2 + n \log\left(1 + \frac{f(\lambda)}{n}\omega\right)\right\} \end{aligned} \quad (2.2)$$

$$\leq \exp\left\{-\lambda x + g(\lambda)v^2 + f(\lambda)\omega\right\}. \quad (2.3)$$

Notice that when  $g(\lambda) = \lambda^2/2$  and  $f(\lambda) \equiv 0$ , condition (2.1) is called canonical assumption considered by de la Peña et al. [9, 10]. In particular, when  $V_{i-1}$  is a constant and  $g(\lambda) \equiv 0$ , condition (2.1) reduces to the condition considered by Rio [32].

Next we show that Theorem 2.1 is very useful for obtaining the concentration inequalities for supermartingales. Introducing the third moments of the supermartingale differences, we have the following Bernstein type inequalities.

**Corollary 2.2.** Assume  $\mathbf{E}(\xi_i^-)^3 < \infty$  for all  $i \in [1, n]$ . Denote by  $\langle\langle S \rangle\rangle_k = \sum_{i=1}^k \mathbf{E}((\xi_i^-)^3 | \mathcal{F}_{i-1})$  for all  $k \in [1, n]$ . Then, for all  $x, v, w > 0$ ,

$$\begin{aligned} & \mathbf{P}\left(S_k \geq x, [S]_k \leq v^2 \text{ and } \langle\langle S \rangle\rangle_k \leq w \text{ for some } k \in [1, n]\right) \\ & \leq \exp\left\{-\bar{\lambda}x + \frac{1}{2}\bar{\lambda}^2 v^2 + \frac{1}{3}\bar{\lambda}^3 w\right\} \end{aligned} \tag{2.4}$$

$$\leq B_1\left(x, \frac{w}{3v^2}, v\right) \tag{2.5}$$

$$\leq B_2\left(x, \frac{w}{3v^2}, v\right), \tag{2.6}$$

where  $\bar{\lambda} = 2x/(v^2 + \sqrt{v^4 + 4wx})$ .

Since  $\langle\langle S \rangle\rangle_k \leq \Upsilon(S_k)$ , the inequalities (2.4), (2.5) and (2.6) hold true when  $\langle\langle S \rangle\rangle_k$  is replaced by  $\Upsilon(S_k)$ . To the best of our knowledge, such inequalities have not been established for the sums of independent random variables.

Notice that (2.5) and (2.6) are respectively the bounds of Bennett and Bernstein. Compared to the conditional Bernstein condition (1.4), the condition of Corollary 2.2 does not assume the existence of the moments of all orders.

For supermartingales with differences bounded from below, we still have the following Bernstein type inequality.

**Corollary 2.3.** Assume  $\xi_i \geq -1$  for all  $i \in [1, n]$ . Then, for all  $x, v > 0$ ,

$$\begin{aligned} \mathbf{P}\left(S_k \geq x \text{ and } [S]_k \leq v^2 \text{ for some } k \in [1, n]\right) & \leq \left(1 + \frac{x}{v^2}\right)^{v^2} e^{-x} \\ & \leq B_1(x, 1, v) \\ & \leq B_2(x, 1, v). \end{aligned} \tag{2.7}$$

Inequality (2.7) is similar to Freedman’s inequality (1.3). However, there are two differences between (2.7) and (1.3). First, we assume  $\xi_i$  bounded from below instead of  $\xi_i$  bounded from above. Second, the quadratic characteristic  $\langle S \rangle_k$  in Freedman’s inequality is replaced by the squared variation  $[S]_k$  in our inequality (2.7). Such inequality could be useful for estimating the tail probabilities when the variances of  $(\xi_i)$  do not exist.

Under the conditional Bernstein condition, we have

**Corollary 2.4.** Assume, for a constant  $\epsilon \in (0, \infty)$ ,

$$\mathbf{E}(\xi_i^l | \mathcal{F}_{i-1}) \leq \frac{1}{2} l! \epsilon^{l-2} \mathbf{E}(\xi_i^2 | \mathcal{F}_{i-1}) \text{ a.s. for all } l \geq 2 \text{ and all } i \in [1, n]. \tag{2.8}$$

Then, for all  $x, v > 0$ ,

$$\begin{aligned} & \mathbf{P}\left(S_k \geq x \text{ and } \langle S \rangle_k \leq v^2 \text{ for some } k \in [1, n]\right) \\ & \leq B_{1,n}(x, \epsilon, v) := \exp\left\{-\bar{\lambda}x + n \log\left(1 + \frac{\bar{\lambda}^2 v^2}{2n(1 - \bar{\lambda}\epsilon)}\right)\right\} \end{aligned} \tag{2.9}$$

$$\leq B_1(x, \epsilon, v), \tag{2.10}$$

where

$$\bar{\lambda} = \frac{2x/v^2}{2x\epsilon/v^2 + 1 + \sqrt{1 + 2x\epsilon/v^2}} \in (0, \epsilon^{-1}).$$

Notice that  $B_1(x, \epsilon, v) = \exp\left\{-\bar{\lambda}x + \frac{\bar{\lambda}^2 v^2}{2(1-\bar{\lambda}\epsilon)}\right\}$ . In the independent case, inequality (2.10) is known as Bennett's inequality [2]. To highlight how the bound  $B_{1,n}(x, \epsilon, v)$  improves Bennett's bound  $B_1(x, \epsilon, v)$ , we rewrite

$$B_{1,n}(x, \epsilon, v) = B_1(x, \epsilon, v) \exp\left\{-n\psi\left(\frac{\bar{\lambda}^2 v^2}{2n(1-\bar{\lambda}\epsilon)}\right)\right\},$$

where  $\psi(t) = t - \log(1+t)$  is a nonnegative convex function in  $t \geq 0$ . It is easy to see that, in the i.i.d. case with  $v^2 = n\sigma_1^2$  (or more generally when  $\frac{\epsilon}{v} = \frac{\sigma_1}{\sqrt{n}}$  for a constant  $\sigma_1 > 0$ ), we have

$$B_{1,n}(nx, \epsilon, \sqrt{n}\sigma_1) = B_1(nx, \epsilon, \sqrt{n}\sigma_1) \exp\{-n c_{x, \sigma_1, \epsilon}\}, \quad (2.11)$$

where  $c_{x, \sigma_1, \epsilon} = \psi\left(\frac{\bar{\lambda}^2 v^2}{2(1-\bar{\lambda}\epsilon)}\right) > 0$  does not depend on  $n$ . Thus Bennett's bound  $B_1(nx, \epsilon, \sqrt{n}\sigma_1)$  on tail probabilities  $\mathbf{P}(S_n \geq nx)$  is strengthened by adding a factor with exponential decay rate  $\exp\{-n c_{x, \sigma_1, \epsilon}\}$  as  $n \rightarrow \infty$ . Since the conditional Bernstein condition (1.4) implies condition (2.8), inequality (2.9) strengthens de la Peña's inequality (1.5).

One calls  $(\xi_i, \mathcal{F}_i)_{i=1, \dots, n}$  *conditionally symmetric*, if  $\mathbf{E}(\xi_i > y | \mathcal{F}_{i-1}) = \mathbf{E}(\xi_i < -y | \mathcal{F}_{i-1})$  for all  $i \in [1, n]$  and for any  $y \geq 0$ ; see Hitczenko [19], de la Peña [8] and Bercu and Touati [4]. It is obvious that if  $(\xi_i, \mathcal{F}_i)_{i=1, \dots, n}$  are conditionally symmetric, then, for any  $y > 0$ ,  $(\xi_i \mathbf{1}_{\{|\xi_i| > y\}}, \mathcal{F}_i)_{i=1, \dots, n}$  are also conditionally symmetric. In particular, the conditionally symmetric martingale differences satisfy the canonical assumption  $\mathbf{E}(\exp\{\lambda\xi_i - \lambda^2\xi_i^2/2\} | \mathcal{F}_{i-1}) \leq 1$  for all  $\lambda \geq 0$ ; see [8, 9, 10]. Thus, by Theorem 2.1 and optimizing on  $\lambda$ , inequality (2.3) implies de la Peña's inequality (1.11).

The following result is a Fuk-Nagaev type inequality [17, 27] for martingales with conditionally symmetric differences. Its proof is based on a truncation argument on martingale differences.

**Corollary 2.5.** *Assume that  $(\xi_i, \mathcal{F}_i)_{i=1, \dots, n}$  are conditionally symmetric. Let*

$$V_k^2(y) = \sum_{i=1}^k \mathbf{E}(\xi_i^2 \mathbf{1}_{\{|\xi_i| \leq y\}} | \mathcal{F}_{i-1}), \quad k \in [1, n].$$

Then, for all  $x, y, v > 0$  and  $v^2 \leq ny^2$ ,

$$\begin{aligned} & \mathbf{P}\left(S_k \geq x \text{ and } V_k^2(y) \leq v^2 \text{ for some } k \in [1, n]\right) \\ & \leq \exp\left\{-\underline{\lambda}x + n \log\left(1 + \frac{v^2}{ny^2} (\cosh(\underline{\lambda}y) - 1)\right)\right\} + \mathbf{P}\left(\max_{1 \leq i \leq n} \xi_i > y\right) \end{aligned} \quad (2.12)$$

$$\leq \exp\left\{-\bar{\lambda}x + \frac{v^2}{y^2} (\cosh(\bar{\lambda}y) - 1)\right\} + \mathbf{P}\left(\max_{1 \leq i \leq n} \xi_i > y\right), \quad (2.13)$$

where

$$\underline{\lambda} = \frac{1}{y} \log\left(\frac{\frac{xy}{v^2} - \frac{x}{ny} + \sqrt{1 + \frac{(xy)^2}{v^4} - 2\frac{x^2}{nv^2}}}{1 - \frac{x}{ny}}\right) \quad \text{and} \quad \bar{\lambda} = \frac{1}{y} \log\left(\frac{xy}{v^2} + \sqrt{1 + \frac{(xy)^2}{v^4}}\right).$$

Inequality (2.12) is the best possible that can be obtained from the exponential Markov inequality  $\mathbf{P}(S_n \geq x) \leq \inf_{\lambda > 0} \mathbf{E}e^{\lambda(S_n - x)}$  under the present assumption. Indeed, if  $(\xi_i)_{i=1, \dots, n}$  are i.i.d. and satisfy the following distribution

$$\mathbf{P}(\xi_i = y) = \mathbf{P}(\xi_i = -y) = \frac{v^2}{2ny^2} \quad \text{and} \quad \mathbf{P}(\xi_i = 0) = 1 - \frac{v^2}{ny^2}, \tag{2.14}$$

then the bound (2.12) equals to  $\inf_{\lambda > 0} \mathbf{E}e^{\lambda(S_n - x)}$ . In this sense, inequality (2.12) is a version of Hoeffding’s inequality (cf. (2.8) of [20]) for martingales with conditionally symmetric differences.

For martingales with bounded conditionally symmetric differences, Sason [33] has obtained (2.12) under the conditions  $|\xi_i| \leq y$  and  $\mathbf{E}(\xi_i^2 | \mathcal{F}_{i-1}) \leq v^2/n$ . He has also obtained (2.13) under the assumption  $|\xi_i| \leq y$ . Thus (2.12) improves and generalizes the Sason’s inequalities under a more general condition.

For the martingales with square integrable differences, several Nagaev type inequalities based on the truncation arguments on martingale differences can be found in Haeusler [18] and Courbot [7]. For optimal exponential convergence speed of such type bounds, we refer to Lesigne and Volný [23] and Fan et al. [14, 15].

Consider the case that the differences  $(\xi_i, \mathcal{F}_i)_{i=1, \dots, n}$  are sub-Gaussian. We have the following very general result.

**Theorem 2.6.** *Assume that  $V_{i-1}, i \in [1, n]$ , are positive and  $\mathcal{F}_{i-1}$ -measurable random variables. Suppose  $\mathbf{E}(e^{\lambda \xi_i} | \mathcal{F}_{i-1}) \leq \exp\{f(\lambda)V_{i-1}\}$  for all  $i \in [1, n]$  and for a positive function  $f(\lambda)$  for some  $\lambda \in (0, \infty)$ . Then, for all  $x, v > 0$ ,*

$$\mathbf{P}\left(S_k \geq x \text{ and } \sum_{i=1}^k V_{i-1} \leq v^2 \text{ for some } k \in [1, n]\right) \leq \exp\left\{-\lambda x + f(\lambda)v^2\right\}. \tag{2.15}$$

In the particular case where  $v^2 = \sum_{i=1}^n \|V_{i-1}\|_\infty$  and  $f(\lambda) = \lambda^2/2$ , Theorem 2.6 reduces to Theorem 4 of Fuk [17] after optimizing on  $\lambda$ . If  $V_{i-1} = \mathbf{E}(\xi_i^2 | \mathcal{F}_{i-1})$ , Theorem 2.6 reduces to Theorem 4.2 of Khan [22]. Thus (2.15) can be regarded as a generalization of the inequalities of Fuk [17] and Khan [22].

Using Theorem 2.6, we extend Azuma-Hoeffding’s inequality (cf. [1, 20]) to the case that the differences are only bounded from above.

**Corollary 2.7.** *Assume that  $U_{i-1}, i \in [1, n]$ , are nonnegative and  $\mathcal{F}_{i-1}$ -measurable random variables. Denote by*

$$C_{i-1}^2 = \begin{cases} \mathbf{E}(\xi_i^2 | \mathcal{F}_{i-1}), & \text{if } \mathbf{E}(\xi_i^2 | \mathcal{F}_{i-1}) \geq U_{i-1}^2, \\ \frac{1}{4} \left( U_{i-1} + \frac{\mathbf{E}(\xi_i^2 | \mathcal{F}_{i-1})}{U_{i-1}} \right)^2, & \text{otherwise.} \end{cases} \tag{2.16}$$

If  $\xi_i \leq U_{i-1}$  for all  $i \in [1, n]$ , then, for all  $\lambda > 0$ ,

$$\mathbf{E}(e^{\lambda \xi_i} | \mathcal{F}_{i-1}) \leq \exp\left\{\frac{\lambda^2}{2} C_{i-1}^2\right\}, \tag{2.17}$$

and, for all  $x, v > 0$ ,

$$\mathbf{P}\left(S_k \geq x \text{ and } \sum_{i=1}^k C_{i-1}^2 \leq v^2 \text{ for some } k \in [1, n]\right) \leq \exp\left\{-\frac{x^2}{2v^2}\right\}. \tag{2.18}$$

In particular, if  $\mathbf{E}(\xi_i^2 | \mathcal{F}_{i-1}) \geq U_{i-1}^2$  for all  $i \in [1, n]$ , then, for all  $x, v \geq 0$ ,

$$\mathbf{P}\left(\max_{1 \leq k \leq n} S_k \geq x \text{ and } \langle S \rangle_n \leq v^2\right) \leq \exp\left\{-\frac{x^2}{2v^2}\right\}. \tag{2.19}$$



Notice that if  $(\xi_i)_{i=1,\dots,n}$  are independent and satisfy the conditions  $\xi_i \leq c_i$  and  $\mathbf{E}\xi_i^2 \geq c_i^2$  for some constants  $(c_i)_{i=1,\dots,n}$ , then (2.19) is a gaussian bound with  $v^2 = \sum_{i=1}^n \mathbf{E}\xi_i^2$ . It is obvious that the Rademacher random variables satisfy this assumption.

For martingale differences  $(\xi_i, \mathcal{F}_i)_{i=1,\dots,n}$ , inequality (2.18) generalizes the following inequality due to van de Geer (cf. Theorem 2.5 of [36]): if  $L_{i-1} \leq \xi_i \leq U_{i-1}$  for some  $\mathcal{F}_{i-1}$ -measurable random variables  $L_{i-1}$  and  $U_{i-1}$ , then, for all  $x, v > 0$ ,

$$\mathbf{P}\left(S_k \geq x \text{ and } \frac{1}{4} \sum_{i=1}^k (U_{i-1} - L_{i-1})^2 \leq v^2 \text{ for some } k \in [1, n]\right) \leq \exp\left\{-\frac{x^2}{2v^2}\right\}. \quad (2.20)$$

Indeed, since

$$\mathbf{E}(\xi_i^2 | \mathcal{F}_{i-1}) = \mathbf{E}((\xi_i - L_{i-1})\xi_i | \mathcal{F}_{i-1}) \leq \mathbf{E}((\xi_i - L_{i-1})U_{i-1} | \mathcal{F}_{i-1}) \leq -L_{i-1}U_{i-1},$$

we have

$$\sum_{i=1}^k C_{i-1}^2 \leq \sum_{i=1}^k \frac{1}{4} \left( U_{i-1} + \frac{\mathbf{E}(\xi_i^2 | \mathcal{F}_{i-1})}{U_{i-1}} \right)^2 \leq \frac{1}{4} \sum_{i=1}^k (U_{i-1} - L_{i-1})^2 \quad (2.21)$$

and

$$\left\{ \frac{1}{4} \sum_{i=1}^k (U_{i-1} - L_{i-1})^2 \leq v^2 \right\} \subseteq \left\{ \sum_{i=1}^k C_{i-1}^2 \leq v^2 \right\},$$

which together with (2.18) implies (2.20).

Under the assumption of Corollary 2.7, Pinelis [29, 30] (see also Bentkus [3]) proved the following inequality, for all  $x > 0$ ,

$$\begin{aligned} \mathbf{P}\left(\max_{1 \leq k \leq n} S_k \geq x\right) &\leq c \left(1 - \Phi\left(\frac{x}{\hat{v}}\right)\right) \\ &= O\left(\frac{1}{1+x/\hat{v}} \exp\left\{-\frac{x^2}{2\hat{v}^2}\right\}\right), \quad x \rightarrow \infty, \end{aligned} \quad (2.22)$$

where  $c$  is an absolute constant and

$$\hat{v}^2 = \sum_{i=1}^n \left\| \frac{1}{4} \left( U_{i-1} + \frac{\mathbf{E}(\xi_i^2 | \mathcal{F}_{i-1})}{U_{i-1}} \right)^2 \right\|_{\infty}.$$

Notice that  $\hat{v}^2 \geq \|\sum_{i=1}^n C_{i-1}^2\|_{\infty}$ . If  $\hat{v}^2 = \|\sum_{i=1}^n C_{i-1}^2\|_{\infty}$ , Pinelis' inequality (2.22) is better than ours (2.18) by adding a factor  $\frac{O(1)}{1+x/\hat{v}}$ . Otherwise  $\hat{v}^2 > \|\sum_{i=1}^n C_{i-1}^2\|_{\infty}$ , our inequality (2.18) improves Pinelis' inequality (2.22) by adding an exponential decay factor of order

$$\left(1 + \frac{x}{\hat{v}}\right) \exp\left\{-\frac{x^2}{2}\delta\right\}, \quad x \rightarrow \infty, \quad (2.23)$$

where

$$\delta = \frac{\hat{v}^2 - \|\sum_{i=1}^n C_{i-1}^2\|_{\infty}}{\hat{v}^2 \|\sum_{i=1}^n C_{i-1}^2\|_{\infty}} > 0.$$

To illustrate this factor, consider the following example. For a much more significant improvement, we refer to Remark 3.3.

*Example 1:* Assume that  $(\varepsilon_i)_{i=1,\dots,n}$  is a sequence of Rademacher random variables, and that  $\mathcal{N}$  is a random variable independent of  $(\varepsilon_i)_{i=1,\dots,n}$ . Set

$$\xi_i = \left(\frac{\varepsilon_i}{\sqrt{n}} \sin \mathcal{N}\right) \mathbf{1}_{\{i \text{ is odd}\}} + \left(\frac{\varepsilon_i}{\sqrt{n}} \cos \mathcal{N}\right) \mathbf{1}_{\{i \text{ is even}\}},$$

$\mathcal{F}_0 = \sigma\{\mathcal{N}\}$  and  $\mathcal{F}_i = \sigma\{\mathcal{N}, \varepsilon_j, 1 \leq j \leq i\}$ . So we have

$$\xi_i \leq U_i := \left(\frac{1}{\sqrt{n}}|\sin \mathcal{N}|\right) \mathbf{1}_{\{i \text{ is odd}\}} + \left(\frac{1}{\sqrt{n}}|\cos \mathcal{N}|\right) \mathbf{1}_{\{i \text{ is even}\}}$$

and

$$\sum_{i=1}^n C_{i-1}^2 = \langle S \rangle_n = \sum_{i=1}^n \left( \frac{\sin^2 \mathcal{N}}{n} \mathbf{1}_{\{i \text{ is odd}\}} + \frac{\cos^2 \mathcal{N}}{n} \mathbf{1}_{\{i \text{ is even}\}} \right).$$

Hence, for any even number  $n$ , it is easy to see that  $v^2 = 1 > \frac{1}{2} = \sum_{i=1}^n C_{i-1}^2 = \langle S \rangle_n$ . Then Pinelis' inequality (2.22) shows that:

$$\mathbf{P} \left( \max_{1 \leq k \leq n} S_k \geq x \right) = O \left( \frac{1}{1+x} \exp \left\{ -\frac{x^2}{2} \right\} \right), \quad x \rightarrow \infty,$$

while our inequality (2.18) implies that:

$$\mathbf{P} \left( \max_{1 \leq k \leq n} S_k \geq x \right) \leq \exp \left\{ -x^2 \right\}.$$

Thus our inequality (2.18) improves Pinelis' inequality (2.22) by adding a factor with the exponential decay rate  $(1+x) \exp \left\{ -\frac{x^2}{2} \right\}$ .

**Remark 2.8.** Corollary 2.7 implies a simple proof of the following self-normalized deviation inequality. Assume that  $(\xi_i)_{i=1, \dots, n}$  are independent and symmetric. Then, for all  $x > 0$ ,

$$\mathbf{P} \left( \max_{1 \leq k \leq n} \frac{S_k}{\sqrt{[S]_n}} \geq x \right) \leq \exp \left\{ -\frac{x^2}{2} \right\}, \tag{2.24}$$

where by convention  $\frac{0}{0} = 0$ . A similar result can be found in Hitczenko [19]. Hitczenko has obtained the same upper bound on tail probabilities  $\mathbf{P} \left( S_n \geq x \mid \sqrt{[S]_n} \mid \infty \right)$ . For more precise results, we refer to Wang and Jing [37]. In particular, the Cramér type large deviations have been established by Jing, Shao and Wang [21] without assuming that  $(\xi_i)_{i=1, \dots, n}$  are symmetric (or  $(\xi_i)_{i=1, \dots, n}$  have exponential moments).

### 3 Applications to statistical estimation

The exponential concentration inequalities for martingales certainly have many applications. McDiarmid [26] and Rio [31] applied such type inequalities to estimate the concentration of separately Lipschitz functions. Van de Geer [35] found that such inequalities can be used for maximum likelihood estimation for counting processes. Liu and Watbled [25] considered the free energy of directed polymers in a random environment via martingale inequalities. Dedecker and Fan [11] gave an application of these inequalities to the Wasserstein distance between the empirical measure and the invariant distribution. We refer to Bercu [5] for more interesting applications of the concentration inequalities for martingales.

In the sequel, we discuss how to apply our results to linear regression models, autoregressive processes and branching processes. We find these models in Liptser and Spokoiny [24] and Bercu and Touati [4].

**1. Linear regression models.** Consider the stochastic linear regression models given, for all  $k \in [1, n]$ , by

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

where  $X_k, \phi_k$  and  $\varepsilon_k$  are the observations, the regression variables and the driven noises, respectively. We assume that  $(\phi_k)$  is a sequence of independent random variables. We also assume that  $(\varepsilon_k)$  is a sequence of independent and identically distributed (i.i.d.) random variables, with mean zero and variation  $\sigma^2 > 0$ . Moreover, we suppose that  $(\phi_k)$  and  $(\varepsilon_k)$  are independent. Our interest is to estimate the unknown parameter  $\theta$ . The well-known least-squares estimator  $\theta_n$  is given below

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

When  $(\phi_k)$  and  $(\varepsilon_k)$  are sub-Gaussian, exponential inequalities on the convergence of  $\theta_n - \theta$  have been established by Bercu and Touati [4]. When  $(\varepsilon_k)$  are the normal random variables, Liptser and Spokoiny [24] have established the following estimation: for all  $x \geq 1$ ,

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

Here, we would like to give a generalization of this inequality. Consider the case that the random variables  $(\varepsilon_k)$  satisfy the Bernstein condition.

**Theorem 3.1.** Assume  $|\phi_k|/\sqrt{\sum_{k=1}^n \phi_k^2} \leq \epsilon_1$  and

$$|\mathbf{E}\varepsilon_i^k| \leq \frac{1}{2} k! \epsilon_2^{k-2} \mathbf{E}\varepsilon_i^2, \quad \text{for all } k \geq 2 \text{ and all } i \in [1, n],$$

for two positive numbers  $\epsilon_1$  and  $\epsilon_2$ . Let  $\epsilon = \epsilon_1 \epsilon_2 / \sigma$ . Then, for all  $x \geq 0$ ,

$$\mathbf{P} \left( \pm (\theta_n - \theta) \sqrt{\sum_{k=1}^n \phi_k^2} \geq x\sigma \right) \leq B_{1,n}(x, \epsilon, 1) \leq \exp \left\{ -\frac{x^2}{2(1+x\epsilon)} \right\}. \quad (3.4)$$

Since  $|\phi_k|/\sqrt{\sum_{k=1}^n \phi_k^2} \leq 1$ , the condition imposed on  $(\phi_k)$  of Theorem 3.1 can be dropped by taking  $\epsilon_1 = 1$ . It is interesting to see that by taking  $\epsilon_1 = 1$ , bound (3.4) does not depend on the distribution of the regression variables  $(\phi_k)$ . This is a big advantage in practice.

If  $a \leq |\phi_k| \leq b$  for two positive constants  $a$  and  $b$ , then the condition of Theorem 2.1 is satisfied with  $\epsilon_1 = \frac{b}{a\sqrt{n}}$ . Indeed, it is easy to see that

$$\frac{|\phi_k|}{\sqrt{\sum_{i=1}^n \phi_i^2}} \leq \frac{b}{\sqrt{na^2}} = \epsilon_1.$$

In this case, bound (3.4) behaviors like  $\exp\{-x^2/2\}$  when  $x = o(\sqrt{n})$  as  $n \rightarrow \infty$ . When  $x$  is large, bound (3.4) behaviors like  $\exp\{-x\}$ .

If  $(\varepsilon_k)$  are bounded from above, we have the following sub-Gaussian tail bound from Corollary 2.7.

**Theorem 3.2.** If  $\varepsilon_k \leq \epsilon$  for all  $k \in [1, n]$ , then, for all  $x \geq 0$ ,

$$\mathbf{P} \left( (\theta_n - \theta) \sqrt{\sum_{k=1}^n \phi_k^2} \geq x\sigma \right) \leq \exp \left\{ -\frac{x^2}{2C_n} \right\}, \quad (3.5)$$

where

$$C_n = \frac{1}{4} \left( \frac{\epsilon}{\sigma} + \frac{\sigma}{\epsilon} \right)^2.$$

In particular, if  $|\varepsilon_k| \leq \epsilon$ , bound (3.5) holds true on the tail probabilities

$$\mathbf{P} \left( \pm (\theta_n - \theta) \sqrt{\sum_{k=1}^n \phi_k^2} \geq x\sigma \right).$$

**Remark 3.3.** If  $|\varepsilon_k| \leq \epsilon$ , we can obtain some similar bounds by using van de Geer’s inequality (2.20) or Pinelis’ inequality (2.22). However, those bounds are less tight than (3.5). Indeed, by van de Geer’s inequality, we can obtain the bound (3.5) with a larger  $C_n = (\epsilon/\sigma)^2$ . If we make use of Pinelis’ inequality (or Bentkus’ inequality [3]), the bound will be as large as

$$\frac{O(1)}{x} \exp \left\{ -\frac{x^2}{2nC_n} \right\}.$$

Next, consider the tail probabilities of  $(\theta_n - \theta) \sum_{k=1}^n \phi_k^2$ . It seems that our inequalities fit well to such type estimations.

**Theorem 3.4.** Assume that there exist  $\alpha \in (1, 2]$  and  $c > 0$  such that

$$\mathbf{E}e^{\lambda\varepsilon_i} \leq e^{c|\lambda|^\alpha} \quad \text{for all } i \in [1, n] \text{ and all } \lambda \in \mathbf{R}.$$

Then, for all  $x, v \geq 0$ ,

$$\mathbf{P} \left( \pm (\theta_n - \theta) \sum_{k=1}^n \phi_k^2 \geq x \text{ and } \sum_{k=1}^n |\phi_k|^\alpha \leq v^\alpha \right) \leq \exp \left\{ -C(\alpha) \left( \frac{x}{v} \right)^{\frac{\alpha}{\alpha-1}} \right\}, \quad (3.6)$$

where

$$C(\alpha) = (c\alpha)^{\frac{1}{1-\alpha}} (1 - \alpha^{-1}).$$

When the condition of Theorem 3.4 is verified with  $\alpha = 2$ , then  $(\varepsilon_i)$  are known as sub-Gaussian random variables. It is known that the bounded random variables and the normal random variables are all sub-Gaussian random variables. In particular, if  $(\varepsilon_i)$  are the standard normal random variables, then bound (3.6) is valid with  $\alpha = 2$  and  $c = C(2) = 1/2$ .

**2. Autoregressive processes.** The model of autoregressive can be stated as follows: for all  $k \in [1, n]$ ,

$$X_k = \theta X_{k-1} + \varepsilon_k, \quad (3.7)$$

where  $(X_k)$  and  $(\varepsilon_k)$  are the observations and driven noises, respectively. We assume that  $(\varepsilon_k)$  is a sequence of i.i.d. centered random variables with variation  $\sigma^2 > 0$ . The process is said to be stable if  $|\theta| \leq 1$ , unstable if  $|\theta| = 1$  and explosive if  $|\theta| > 1$ . We can estimate the unknown parameter  $\theta$  by the least-squares estimator given by, for all  $n \geq 1$ ,

$$\theta'_n = \frac{\sum_{k=1}^n X_k X_{k-1}}{\sum_{k=1}^n X_{k-1}^2}. \quad (3.8)$$

When  $X_0$  and  $(\varepsilon_k)$  are the normal random variables, the convergence rate of  $\theta'_n - \theta$  has been established by Bercu and Touati [4]. Here, we would like to give an almost sure convergence rate of  $(\theta'_n - \theta) \sum_{k=1}^n X_{k-1}^2$ .

By an argument similar to that of Theorem 3.4, we have the following result.

**Theorem 3.5.** Assume the condition of Theorem 3.4. Then bound (3.6) holds true on the tail probabilities

$$\mathbf{P} \left( \pm (\theta'_n - \theta) \sum_{k=1}^n X_{k-1}^2 \geq x \text{ and } \sum_{k=1}^n |X_{k-1}|^\alpha \leq v^\alpha \right). \quad (3.9)$$

If  $(\varepsilon_i)$  are bounded, then we have

**Theorem 3.6.** Assume  $|\varepsilon_i| \leq \epsilon$  for all  $i \in [1, n]$ . Then, for all  $x, v > 0$ ,

$$\mathbf{P} \left( \pm (\theta'_n - \theta) \sum_{k=1}^n X_{k-1}^2 \geq x \text{ and } L_n \leq v^2 \right) \leq \exp \left\{ -\frac{x^2}{2v^2} \right\}, \quad (3.10)$$

where

$$L_n = \frac{1}{4} \left( \epsilon + \frac{\sigma^2}{\epsilon} \right)^2 \sum_{k=1}^n X_{k-1}^2.$$

**Remark 3.7.** We can obtain some similar bounds by using Corollary 2.3 or van de Geer's inequality. However, those bounds are less tight than (3.10). For instance, by van de Geer's inequality, we can obtain the bound (3.10) with a larger  $L_n = \epsilon^2 \sum_{k=1}^n X_{k-1}^2$ .

**3. Branching processes.** Consider the Galton-Watson process starting from  $X_0 = 1$  and given, for all  $n \geq 1$ , by

$$X_n = \sum_{k=1}^{X_{n-1}} Y_{n,k},$$

where  $(Y_{n,k})$  is a sequence i.i.d. and nonnegative integer-valued random variables. The distribution of  $(Y_{n,k})$ , with finite mean  $m$  and variance  $\sigma^2$ , is commonly called the offspring or reproduction distribution. We are interested in the estimation of the offspring mean  $m$ . The Lotka-Nagaev estimator is given by

$$m_n = \frac{X_n}{X_{n-1}}.$$

Assume  $X_n > 0$  a.s. such that the Lotka-Nagaev estimator  $m_n$  is always well defined. Our goal is to establish exponential inequalities for  $m_n$ . Denote by

$$\xi_{n,k} = Y_{n,k} - m.$$

Then

$$(m_n - m)X_{n-1} = X_n - mX_{n-1} = \sum_{k=1}^{X_{n-1}} \xi_{n,k}.$$

Thus  $(m_n - m)X_{n-1}$  is a sum of independent random variables by given  $X_{n-1}$ . By Corollary 2.4, we easily obtain the following exponential inequalities.

**Theorem 3.8.** Assume, for a constant  $\epsilon \in (0, \infty)$ ,

$$|\mathbf{E}\xi_{n,k}^l| \leq \frac{1}{2} l! \epsilon^{l-2} \mathbf{E}\xi_{n,k}^2 \quad \text{for all } l \geq 2 \text{ and all } k \in [1, X_{n-1}].$$

Then, for all  $x, v > 0$ , it holds

$$\begin{aligned} \mathbf{P}\left(|m_n - m|X_{n-1} \geq x \text{ and } X_{n-1}\sigma^2 \leq v^2 \mid X_{n-1}\right) &\leq 2B_{1,n}(x, \epsilon, v) \\ &\leq 2 \exp\left\{-\frac{x^2}{2(v^2 + x\epsilon)}\right\}. \end{aligned}$$

In particular, it implies that, for all  $x > 0$ ,

$$\mathbf{P}\left(|m_n - m| \geq x\right) \leq 2 \mathbf{E}\left(\exp\left\{-X_{n-1} \frac{x^2}{2(\sigma^2 + x\epsilon)}\right\}\right).$$

Since  $\xi_{n,k} \geq -m$ , we have the following one side sub-Gaussian bound by Corollary 2.7. This bound cannot be obtained from Azuma-Hoeffding's inequality.

**Theorem 3.9.** For all  $x, v > 0$ , it holds

$$\mathbf{P}\left((m_n - m)X_{n-1} \leq -x, MX_{n-1} \leq v^2 \mid X_{n-1}\right) \leq \exp\left\{-\frac{x^2}{2v^2}\right\},$$

where

$$M = \begin{cases} \sigma^2, & \text{if } \sigma \geq m, \\ \frac{1}{4} \left( m + \frac{\sigma^2}{m} \right)^2, & \text{if } \sigma < m. \end{cases}$$

In particular, it implies that, for all  $x > 0$ ,

$$\mathbf{P} \left( (m_n - m) \leq -x \right) \leq \mathbf{E} \left( \exp \left\{ -X_{n-1} \frac{x^2}{2M} \right\} \right).$$

More generale estimations on the tail probabilities  $\mathbf{P} (|m_n - m| \geq x)$ , we refer to Bercu and Touati [4]. In particular, Bercu and Touati have established the Bernstein bounds associated with the cumulant generating function of  $\xi_{n,k}$ .

#### 4 Proof of Theorem 2.1

Suppose  $\mathbf{E}(\exp \{ \lambda \xi_i - g(\lambda) \xi_i^2 \} | \mathcal{F}_{i-1}) \leq 1 + f(\lambda) V_{i-1}$  for a constant  $\lambda \in (0, \infty)$  and all  $i \in [1, n]$ . Define the exponential multiplicative martingale  $Z(\lambda) = (Z_k(\lambda), \mathcal{F}_k)_{k=0, \dots, n}$ , where

$$Z_k(\lambda) = \prod_{i=1}^k \frac{\exp \{ \lambda \xi_i - g(\lambda) \xi_i^2 \}}{\mathbf{E}(\exp \{ \lambda \xi_i - g(\lambda) \xi_i^2 \} | \mathcal{F}_{i-1})}, \quad Z_0(\lambda) = 1.$$

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

$$Z_{T \wedge k}(\lambda) = \prod_{i=1}^{T \wedge k} \frac{\exp \{ \lambda \xi_i - g(\lambda) \xi_i^2 \}}{\mathbf{E}(\exp \{ \lambda \xi_i - g(\lambda) \xi_i^2 \} | \mathcal{F}_{i-1})}, \quad Z_0(\lambda) = 1.$$

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

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

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

*Proof of Theorem 2.1.* For any  $x, v, w > 0$ , define the stopping time

$$T(x, v, w) = \min \left\{ k \in [1, n] : S_k \geq x, [S]_k \leq v^2 \text{ and } \sum_{i=1}^k V_{i-1} \leq w \right\},$$

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

$$\mathbf{1}_{\{S_k \geq x, [S]_k \leq v^2 \text{ and } \sum_{i=1}^k V_{i-1} \leq w \text{ for some } k \in [1, n]\}} = \sum_{k=1}^n \mathbf{1}_{\{T(x, v, w) = k\}}.$$

By the change of measure (4.1), we deduce that, for all  $x, \lambda, v, w > 0$ ,

$$\begin{aligned} & \mathbf{P} \left( S_k \geq x, [S]_k \leq v^2 \text{ and } \sum_{i=1}^k V_{i-1} \leq w \text{ for some } k \in [1, n] \right) \\ &= \mathbf{E}_\lambda \left( Z_{T \wedge n}(\lambda)^{-1} \mathbf{1}_{\{S_k \geq x, [S]_k \leq v^2 \text{ and } \sum_{i=1}^k V_{i-1} \leq w \text{ for some } k \in [1, n]\}} \right) \\ &= \sum_{k=1}^n \mathbf{E}_\lambda \left( \exp \{ -\lambda S_k + g(\lambda) [S]_k + \Xi_k(\lambda) \} \mathbf{1}_{\{T(x, v, w) = k\}} \right), \end{aligned} \tag{4.2}$$

where

$$\Xi_k(\lambda) = \sum_{i=1}^k \log \mathbf{E} \left( \exp \{ \lambda \xi_i - g(\lambda) \xi_i^2 \} \mid \mathcal{F}_{i-1} \right).$$

Using Jensen's inequality and the condition  $\mathbf{E}(\exp \{ \lambda \xi_i - g(\lambda) \xi_i^2 \} \mid \mathcal{F}_{i-1}) \leq 1 + f(\lambda) V_{i-1}$ , we have

$$\begin{aligned} \Xi_k(\lambda) &\leq k \log \left( \frac{1}{k} \sum_{i=1}^k \mathbf{E} \left( \exp \{ \lambda \xi_i - g(\lambda) \xi_i^2 \} \mid \mathcal{F}_{i-1} \right) \right) \\ &\leq k \log \left( 1 + \frac{1}{k} f(\lambda) \sum_{i=1}^k V_{i-1} \right). \end{aligned} \quad (4.3)$$

Thus (4.2) implies that, for all  $x, \lambda, v, w > 0$ ,

$$\begin{aligned} &\mathbf{P} \left( S_k \geq x, [S]_k \leq v^2 \text{ and } \sum_{i=1}^k V_{i-1} \leq w \text{ for some } k \in [1, n] \right) \\ &\leq \sum_{k=1}^n \mathbf{E}_\lambda \left( \exp \left\{ -\lambda S_k + g(\lambda) [S]_k + k \log \left( 1 + \frac{1}{k} f(\lambda) \sum_{i=1}^k V_{i-1} \right) \right\} \mathbf{1}_{\{T(x, v, w) = k\}} \right) \end{aligned} \quad (4.4)$$

By the fact  $S_k \geq x, [S]_k \leq v^2$  and  $\sum_{i=1}^k V_{i-1} \leq w$  on the set  $\{T(x, v, w) = k\}$ , we find that, for all  $x, \lambda, v, w > 0$ ,

$$\begin{aligned} &\mathbf{P} \left( S_k \geq x, [S]_k \leq v^2 \text{ and } \sum_{i=1}^k V_{i-1} \leq w \text{ for some } k \in [1, n] \right) \\ &\leq \exp \left\{ -\lambda x + g(\lambda) v^2 + k \log \left( 1 + \frac{1}{k} f(\lambda) w \right) \right\} \mathbf{E}_\lambda \left( \sum_{k=1}^n \mathbf{1}_{\{T(x, v, w) = k\}} \right) \\ &\leq \exp \left\{ -\lambda x + g(\lambda) v^2 + n \log \left( 1 + \frac{1}{n} f(\lambda) w \right) \right\} \end{aligned} \quad (4.5)$$

$$\leq \exp \{ -\lambda x + g(\lambda) v^2 + f(\lambda) w \}. \quad (4.6)$$

This gives the desired inequalities (2.2) and (2.3), and completes the proof of Theorem 2.1.

*Proof of Corollary 2.2.* To prove Corollary 2.2, we should use the following basic inequality:

$$\exp \left\{ x - \frac{1}{2} x^2 \right\} \leq 1 + x + \frac{1}{3} (x^-)^3, \quad x \in \mathbf{R}.$$

By the last inequality, it follows that, for all  $\lambda > 0$ ,

$$\mathbf{E} \left( \exp \left\{ \lambda \xi_i - \frac{1}{2} (\lambda \xi_i)^2 \right\} \mid \mathcal{F}_{i-1} \right) \leq 1 + \frac{1}{3} \lambda^3 \mathbf{E} \left( (\xi_i^-)^3 \mid \mathcal{F}_{i-1} \right). \quad (4.7)$$

Applying the inequalities (2.2) and (2.3) with  $g(\lambda) = \frac{\lambda^2}{2}$ ,  $f(\lambda) = \frac{\lambda^3}{3}$  and  $V_{i-1} = \mathbf{E} \left( (\xi_i^-)^3 \mid \mathcal{F}_{i-1} \right)$ , we get (2.4) by noting the fact that

$$\inf_{\lambda > 0} \exp \left\{ -\lambda x + \frac{1}{2} \lambda^2 v^2 + \frac{1}{3} \lambda^3 w \right\} = \exp \left\{ -\bar{\lambda} x + \frac{1}{2} \bar{\lambda}^2 v^2 + \frac{1}{3} \bar{\lambda}^3 w \right\}, \quad (4.8)$$

where  $\bar{\lambda} = 2x / (v^2 + \sqrt{v^4 + 4wx})$ . By a simple calculation, we find that, for all  $v, w > 0$  and all  $0 < \lambda < \frac{3v^2}{w}$ ,

$$\frac{1}{2} \lambda^2 v^2 + \frac{1}{3} \lambda^3 w \leq \frac{\lambda^2 v^2}{2(1 - \frac{\lambda w}{3v^2})}. \quad (4.9)$$

Thus, for all  $x, v, w > 0$ ,

$$\begin{aligned} \exp\left\{-\bar{\lambda}x + \frac{1-\bar{\lambda}^2}{2}v^2 + \frac{1-\bar{\lambda}^3}{3}w\right\} &\leq \inf_{0 < \lambda < \frac{3v^2}{w}} \exp\left\{-\lambda x + \frac{\lambda^2 v^2}{2(1-\frac{\lambda w}{3v^2})}\right\} \\ &= B_1\left(x, \frac{w}{3v^2}, v\right) \\ &\leq B_2\left(x, \frac{w}{3v^2}, v\right). \end{aligned}$$

Combining this inequality with (2.4), we obtain the desired inequalities (2.5) and (2.6) of the corollary.

To prove Corollary 2.3, we need the following lemma.

**Lemma 4.1.** *If  $\xi$  is a random variable such that  $\xi \geq -1$  and  $\mathbf{E}\xi \leq 0$ , then, for all  $\lambda \in [0, 1)$ ,*

$$\mathbf{E}\left(\exp\left\{\lambda\xi + (\lambda + \log(1 - \lambda))\xi^2\right\}\right) \leq 1.$$

*Proof.* Assume  $\xi \geq -1$  and  $\lambda \in [0, 1)$ . Then  $\lambda\xi \geq -\lambda > -1$ . Since the function

$$f(x) = \frac{\log(1+x) - x}{x^2/2}, \quad x > -1, \tag{4.10}$$

is increasing in  $x$ , we have

$$\begin{aligned} \log(1 + \lambda\xi) &\geq \lambda\xi + \frac{1}{2}(\lambda\xi)^2 f(-\lambda) \\ &= \lambda\xi + \xi^2(\lambda + \log(1 - \lambda)). \end{aligned} \tag{4.11}$$

Thus

$$\exp\left\{\lambda\xi + \xi^2(\lambda + \log(1 - \lambda))\right\} \leq 1 + \lambda\xi. \tag{4.12}$$

Since  $\mathbf{E}\xi \leq 0$ , it follows that

$$\mathbf{E}\left(\exp\left\{\lambda\xi + \xi^2(\lambda + \log(1 - \lambda))\right\}\right) \leq 1,$$

which gives the desired inequality.

*Proof of Corollary 2.3.* Let  $T = \min\{k \in [1, n] : S_k \geq x \text{ and } [S]_k \leq v^2\}$ . Applying inequality (2.3) with  $g(\lambda) = -(\lambda + \log(1 - \lambda))$  and  $f(\lambda) = 0$ , from Lemma 4.1, we obtain, for all  $x, v > 0$  and all  $\lambda \in [0, 1)$ ,

$$\begin{aligned} &\mathbf{P}(S_k \geq x \text{ and } [S]_k \leq v^2 \text{ for some } k \in [1, n]) \\ &\leq \exp\{-\lambda x - (\lambda + \log(1 - \lambda))v^2\}. \end{aligned} \tag{4.13}$$

It is easy to see that bound (4.13) attains its minimum at

$$\lambda = \lambda(x) = \frac{x}{v^2 + x}. \tag{4.14}$$

Substituting  $\lambda = \lambda(x)$  in (4.13), we get, for all  $x, v > 0$ ,

$$\begin{aligned} &\mathbf{P}(S_k \geq x \text{ and } [S]_k \leq v^2 \text{ for some } k \in [1, n]) \\ &\leq \inf_{\lambda \in [0, 1)} \exp\{-\lambda x - (\lambda + \log(1 - \lambda))v^2\} \end{aligned} \tag{4.15}$$

$$= \left(1 + \frac{x}{v^2}\right)^{v^2} e^{-x}. \tag{4.16}$$



Using Taylor's expansion, we deduce that, for all  $\lambda \in [0, 1)$ ,

$$\begin{aligned} \lambda + \log(1 - \lambda) &= -\frac{\lambda^2}{2} \left( 1 + \frac{2}{3}\lambda + \frac{2}{4}\lambda^2 + \dots \right) \\ &\geq -\frac{\lambda^2}{2} (1 + \lambda + \lambda^2 + \dots) \\ &= -\frac{\lambda^2}{2(1 - \lambda)}. \end{aligned} \quad (4.17)$$

Thus we have, for all  $x, v > 0$ ,

$$\inf_{\lambda \in [0, 1)} \exp\{-\lambda x - (\lambda + \log(1 - \lambda))v^2\} \leq \inf_{\lambda \in [0, 1)} \exp\left\{-\lambda x + \frac{\lambda^2 v^2}{2(1 - \lambda)}\right\} \quad (4.18)$$

$$= B_1(x, 1, v) \quad (4.19)$$

$$\leq B_2(x, 1, v).$$

Combining (4.16), (4.18) and (4.19) together, we obtain the desired inequalities of Corollary 2.3.

*Proof of Corollary 2.4.* Assume  $\mathbf{E}(\xi_i^l | \mathcal{F}_{i-1}) \leq \frac{1}{2} l! \epsilon^{l-2} \mathbf{E}(\xi_i^2 | \mathcal{F}_{i-1})$  for all  $l \geq 2$  and a constant  $\epsilon \in (0, \infty)$ . Then, for all  $0 \leq \lambda < \epsilon^{-1}$ ,

$$\begin{aligned} \mathbf{E}(e^{\lambda \xi_i} | \mathcal{F}_{i-1}) - 1 &= \sum_{k=2}^{+\infty} \frac{\lambda^k}{k!} \mathbf{E}(\xi_i^k | \mathcal{F}_{i-1}) \\ &\leq \frac{\lambda^2}{2} \mathbf{E}(\xi_i^2 | \mathcal{F}_{i-1}) \sum_{k=2}^{\infty} (\lambda \epsilon)^{k-2} \\ &= \frac{\lambda^2}{2(1 - \lambda \epsilon)} \mathbf{E}(\xi_i^2 | \mathcal{F}_{i-1}). \end{aligned}$$

Using Theorem 2.1, we obtain the desired inequality (2.9) with  $\lambda = \bar{\lambda}$ . Since  $n \log(1 + \frac{t}{n}) \leq t$  for all  $t \geq 0$ , it follows that, for all  $x, v > 0$ ,

$$B_{1,n}(x, \epsilon, v) \leq \exp\left\{-\bar{\lambda}x + \frac{\bar{\lambda}^2 v^2}{2(1 - \bar{\lambda} \epsilon)}\right\} = B_1(x, \epsilon, v). \quad (4.20)$$

This completes the proof of Corollary 2.4.

*Proof of Corollary 2.5.* Assume that  $(\xi_i, \mathcal{F}_i)_{i=1, \dots, n}$  are conditionally symmetric. For any  $y > 0$ , let  $\eta_i = \xi_i \mathbf{1}_{\{|\xi_i| \leq y\}}$ . Then  $(\eta_i, \mathcal{F}_i)_{i=1, \dots, n}$  is a sequence of bounded and conditionally symmetric martingale differences. Using Taylor's expansion, we obtain the following estimation of the moment generating function of  $\eta_i$ ,

$$\begin{aligned} \mathbf{E}(e^{\lambda \eta_i} | \mathcal{F}_{i-1}) &= \mathbf{E}\left(\frac{e^{\lambda \eta_i} + e^{-\lambda \eta_i}}{2} \mid \mathcal{F}_{i-1}\right) \\ &= 1 + \sum_{k=1}^{\infty} \frac{\lambda^{2k}}{(2k)!} \mathbf{E}(\eta_i^{2k} | \mathcal{F}_{i-1}). \end{aligned}$$

Since  $|\eta_i| \leq y$ , it follows that  $\mathbf{E}(\eta_i^{2k} | \mathcal{F}_{i-1}) \leq y^{2k-2} \mathbf{E}(\eta_i^2 | \mathcal{F}_{i-1})$  and that

$$\begin{aligned} \mathbf{E}(e^{\lambda \eta_i} | \mathcal{F}_{i-1}) &\leq 1 + \frac{\mathbf{E}(\eta_i^2 | \mathcal{F}_{i-1})}{y^2} \sum_{k=1}^{\infty} \frac{(\lambda y)^{2k}}{(2k)!} \\ &\leq 1 + \frac{\mathbf{E}(\eta_i^2 | \mathcal{F}_{i-1})}{y^2} (\cosh(\lambda y) - 1). \end{aligned} \quad (4.21)$$

Set  $V_k^2(y) = \sum_{i=1}^k \mathbf{E}(\eta_i^2 | \mathcal{F}_{i-1})$  for all  $k \in [1, n]$ . Using Theorem 2.1, we obtain, for all  $x, v > 0$ ,

$$\begin{aligned}
 P_1 &:= \mathbf{P} \left( \sum_{i=1}^k \eta_i \geq x \text{ and } V_k^2(y) \leq v^2 \text{ for some } k \in [1, n] \right) \\
 &\leq \inf_{\lambda \geq 0} \exp \left\{ -\lambda x + n \log \left( 1 + \frac{v^2}{n y^2} (\cosh(\lambda y) - 1) \right) \right\} \tag{4.22}
 \end{aligned}$$

$$\leq \inf_{\lambda \geq 0} \exp \left\{ -\lambda x + \frac{v^2}{y^2} (\cosh(\lambda y) - 1) \right\}. \tag{4.23}$$

By some simple calculations, we find that (4.22) and (4.23) attain their minimums at  $\underline{\lambda}$  and  $\bar{\lambda}$  of Corollary 2.5, respectively. It is easy to see that

$$\begin{aligned}
 &\mathbf{P} (S_k \geq x \text{ and } V_k^2(y) \leq v^2 \text{ for some } k \in [1, n]) \\
 &\leq \mathbf{P} \left( \sum_{i=1}^k (\eta_i + \xi_i \mathbf{1}_{\{\xi_i < -y\}}) \geq x \text{ and } V_k^2(y) \leq v^2 \text{ for some } k \in [1, n] \right) \\
 &\quad + \mathbf{P} \left( \sum_{i=1}^k \xi_i \mathbf{1}_{\{\xi_i > y\}} > 0 \text{ and } V_k^2(y) \leq v^2 \text{ for some } k \in [1, n] \right) \\
 &\leq P_1 + \mathbf{P} \left( \max_{1 \leq i \leq n} \xi_i > y \right). \tag{4.24}
 \end{aligned}$$

Implementing (4.22) and (4.23) into (4.24), we get the desired inequalities (2.12) and (2.13).

### 5 Proof of Theorem 2.6 and its corollaries

The proof of Theorem 2.6 is similar to the argument of Theorem 2.1.

*Proof of Theorem 2.6.* Let  $T = \min\{k \in [1, n] : S_k \geq x \text{ and } \sum_{i=1}^k V_{i-1} \leq v^2\}$ . According to (4.2) with  $g(\lambda) \equiv 0$ , we have the following estimation, for all  $x, v > 0$ ,

$$\mathbf{P} \left( S_k \geq x \text{ and } \sum_{i=1}^k V_{i-1} \leq v^2 \text{ for some } k \in [1, n] \right) \leq \sum_{k=1}^n \mathbf{E}_\lambda \left( \exp \{-\lambda x + \Psi_k(\lambda)\} \mathbf{1}_{\{T=k\}} \right),$$

where

$$\Psi_k(\lambda) = \sum_{i=1}^k \log \mathbf{E}(e^{\lambda \xi_i} | \mathcal{F}_{i-1}).$$

Using the condition  $\mathbf{E}(e^{\lambda \xi_i} | \mathcal{F}_{i-1}) \leq \exp\{f(\lambda)V_{i-1}\}$  and the fact  $\sum_{i=1}^k V_{i-1} \leq v^2$  on the set  $\{T = k\}$ , we obtain

$$\begin{aligned}
 &\mathbf{P} \left( S_k \geq x \text{ and } \sum_{i=1}^k V_{i-1} \leq v^2 \text{ for some } k \in [1, n] \right) \\
 &\leq \sum_{k=1}^n \mathbf{E}_\lambda \left( \exp \left\{ -\lambda x + f(\lambda) \sum_{i=1}^k V_{i-1} \right\} \mathbf{1}_{\{T=k\}} \right) \\
 &\leq \exp \{-\lambda x + f(\lambda)v^2\},
 \end{aligned}$$

which gives (2.15) of Theorem 2.6.

In the proof of Corollary 2.7, we shall need the following two lemmas.

**Lemma 5.1.** *If  $\xi$  is a random variable satisfying  $\xi \leq 1$ ,  $\mathbf{E}\xi \leq 0$  and  $\mathbf{E}\xi^2 = \sigma^2$ , then, for all  $\lambda > 0$ ,*

$$\mathbf{E}e^{\lambda\xi} \leq \frac{1}{1+\sigma^2} \exp\{-\lambda\sigma^2\} + \frac{\sigma^2}{1+\sigma^2} \exp\{\lambda\}.$$

A proof can be found in Fan, Grama and Liu [14].

**Lemma 5.2.** *Assume that  $\xi$  is a random variable satisfying  $\mathbf{E}\xi \leq 0$ ,  $\xi \leq b$  for a constant  $b > 0$  and  $\mathbf{E}\xi^2 = \sigma^2$ . Set*

$$s^2 = \begin{cases} \sigma^2, & \text{if } \sigma \geq b, \\ \frac{1}{4} \left(b + \frac{\sigma^2}{b}\right)^2, & \text{if } \sigma < b. \end{cases} \quad (5.1)$$

Then, for all  $\lambda > 0$ ,

$$\mathbf{E}e^{\lambda\xi} \leq \exp\left\{\frac{\lambda^2 s^2}{2}\right\}. \quad (5.2)$$

*Proof.* If  $\sigma \geq b$ , by Lemma 5.1, then, for all  $t \geq 0$ ,

$$\mathbf{E}e^{t\xi/\sigma} \leq \frac{1}{2}(e^{-t} + e^t) \leq \exp\left\{\frac{t^2}{2}\right\}.$$

Taking  $t = \lambda\sigma \geq 0$ , we have

$$\mathbf{E}e^{\lambda\xi} \leq \exp\left\{\frac{\lambda^2 \sigma^2}{2}\right\} = \exp\left\{\frac{\lambda^2 s^2}{2}\right\}. \quad (5.3)$$

If  $\sigma < b$ , by Lemma 5.1, we get, for all  $t \geq 0$ ,

$$\begin{aligned} \mathbf{E}e^{t\xi/b} &\leq \frac{1}{1+\sigma^2/b^2} \exp\{-t\sigma^2/b^2\} + \frac{\sigma^2/b^2}{1+\sigma^2/b^2} \exp\{t\} \\ &= \exp\{f(z)\}, \end{aligned}$$

where  $z = t(1 + \sigma^2/b^2)$  and  $f(z) = -zp + \log(1 - p + pe^z)$  with  $p = \frac{\sigma^2/b^2}{1+\sigma^2/b^2}$ . Since  $f(0) = f'(0) = 0$ ,

$$f'(z) = -p + \frac{p}{p + (1-p)e^{-z}}$$

and

$$f''(z) = \frac{p(1-p)e^{-z}}{(p + (1-p)e^{-z})^2} \leq \frac{1}{4},$$

we have

$$f(z) \leq \frac{1}{8}z^2 = \frac{t^2}{8} \left(1 + \frac{\sigma^2}{b^2}\right)^2 \quad \text{and} \quad \mathbf{E}e^{t\xi/b} \leq \exp\left\{\frac{t^2}{8} \left(1 + \frac{\sigma^2}{b^2}\right)^2\right\}.$$

Taking  $t = \lambda b \geq 0$ , we obtain

$$\mathbf{E}e^{\lambda\xi} \leq \exp\left\{\frac{\lambda^2}{8} \left(b + \frac{\sigma^2}{b}\right)^2\right\} = \exp\left\{\frac{\lambda^2 s^2}{2}\right\}. \quad (5.4)$$

Combining (5.3) and (5.4) together, we obtain (5.2).

*Proof of Corollary 2.7.* Inequality (2.17) follows immediately from Lemma 5.2. Using Theorem 2.6, we obtain, for all  $x, \lambda, v > 0$ ,

$$\mathbf{P}\left(S_k \geq x \text{ and } \sum_{i=1}^k C_{i-1}^2 \leq v^2 \text{ for some } k \in [1, n]\right) \leq \exp\left\{-\lambda x + \frac{\lambda^2 v^2}{2}\right\}.$$

Minimizing the right hand side of the last inequality with respect to  $\lambda \geq 0$ , we easily obtain (2.18).

*Proof of Remark 2.8.* Assume that  $(\xi_i)_{i=1,\dots,n}$  are independent and symmetric. Set

$$\mathcal{F}_i = \sigma \{ \xi_k, k \leq i, \xi_j^2, 1 \leq j \leq n \}.$$

Since  $\xi_i$  is symmetric, we deduce that

$$\mathbf{E}(\xi_i > y | \mathcal{F}_{i-1}) = \mathbf{E}(\xi_i > y | \xi_i^2) = \mathbf{E}(-\xi_i > y | (-\xi_i)^2) = \mathbf{E}(-\xi_i > y | \mathcal{F}_{i-1}).$$

Thus  $(\frac{\xi_i}{\sqrt{[S]_n}}, \mathcal{F}_i)_{i=1,\dots,n}$  are conditionally symmetric martingale differences. For all  $1 \leq i \leq n$ , we have

$$\mathbf{E} \left( \exp \left\{ \lambda \frac{\xi_i}{\sqrt{[S]_n}} \right\} \middle| \mathcal{F}_{i-1} \right) = \frac{1}{2} \mathbf{E} \left( \exp \left\{ \lambda \frac{\xi_i}{\sqrt{[S]_n}} \right\} + \exp \left\{ -\lambda \frac{\xi_i}{\sqrt{[S]_n}} \right\} \middle| \mathcal{F}_{i-1} \right).$$

Using the inequality  $\frac{1}{2}(e^t + e^{-t}) \leq e^{t^2/2}$ , we obtain, for all  $\lambda \geq 0$ ,

$$\mathbf{E} \left( \exp \left\{ \lambda \frac{\xi_i}{\sqrt{[S]_n}} \right\} \middle| \mathcal{F}_{i-1} \right) \leq \exp \left\{ \frac{\lambda^2 \xi_i^2}{2 [S]_n} \right\}.$$

Since  $\xi_i^2$  is measurable with respect to  $\mathcal{F}_{i-1}$ , it follows that

$$\sum_{i=1}^k \mathbf{E} \left( \frac{\xi_i^2}{[S]_n} \middle| \mathcal{F}_{i-1} \right) = \sum_{i=1}^k \frac{\xi_i^2}{[S]_n} \leq 1$$

for all  $k \in [1, n]$ . By Theorem 2.6 with  $V_{i-1} = \frac{\xi_i^2}{[S]_n}$ , it follows that, for all  $x, \lambda \geq 0$ ,

$$\mathbf{P} \left( \max_{1 \leq k \leq n} \frac{S_k}{\sqrt{[S]_n}} \geq x \right) \leq \exp \left\{ -\lambda x + \frac{\lambda^2}{2} \right\}. \tag{5.5}$$

The right hand side of the last inequality attends its minimum at  $\lambda = x$ . Substituting  $\lambda = x$  into (5.5), we easily get (2.24) of Remark 2.8.

## 6 Proof of Theorems 3.1 - 3.6

We make use of Corollary 2.4 to prove Theorem 3.1.

*Proof of Theorem 3.1.* From (3.1) and (3.2), it is easy to see that

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

For any  $i = 1, \dots, n$ , set

$$\xi_i = \frac{\phi_i \varepsilon_i}{\sigma \sqrt{\sum_{k=1}^n \phi_k^2}} \quad \text{and} \quad \mathcal{F}_i = \sigma \left( \phi_k, \varepsilon_k, 1 \leq k \leq i, \phi_k^2, 1 \leq k \leq n \right). \tag{6.1}$$

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

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

Notice that

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

and that

$$\begin{aligned} |\mathbf{E}(\xi_i^k | \mathcal{F}_{i-1})| &= \frac{\phi_i^2}{\sigma^k (\sum_{k=1}^n \phi_k^2)} \mathbf{E} \left( \left( \frac{\phi_i}{\sqrt{\sum_{k=1}^n \phi_k^2}} \right)^{k-2} \varepsilon_i^k \middle| \mathcal{F}_{i-1} \right) \\ &\leq \frac{\phi_i^2 \epsilon_1^{k-2}}{\sigma^k (\sum_{k=1}^n \phi_k^2)} \mathbf{E}(\varepsilon_i^k | \mathcal{F}_{i-1}) \\ &\leq \frac{1}{2} k! \epsilon_2^{k-2} \frac{\phi_i^2 \epsilon_1^{k-2}}{\sigma^{k-2} (\sum_{k=1}^n \phi_k^2)} \\ &= \frac{1}{2} k! \epsilon^{k-2} \mathbf{E}(\xi_i^2 | \mathcal{F}_{i-1}). \end{aligned}$$

Applying Corollary 2.4 to  $(\xi_i, \mathcal{F}_i)_{i=1, \dots, n}$ , we obtain the claim of Theorem 3.1.

*Proof of Theorem 3.2.* It is easy to see that the martingale differences  $(\xi_i, \mathcal{F}_i)_{i=1, \dots, n}$ , defined by (6.1), satisfy

$$\xi_i \leq U_{i-1} := \frac{|\phi_i| \epsilon}{\sigma \sqrt{\sum_{k=1}^n \phi_k^2}} \quad \text{and} \quad \mathbf{E}(\xi_i^2 | \mathcal{F}_{i-1}) = \frac{\phi_i^2}{\sum_{k=1}^n \phi_k^2}. \quad (6.2)$$

Applying Corollary 2.7 to  $(\xi_i, \mathcal{F}_i)_{i=1, \dots, n}$ , we obtain the desired inequality.

*Proof of Theorem 3.4.* From (3.1) and (3.2), it is easy to see that

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

For any  $i = 1, \dots, n$ , set

$$\xi_i = \phi_i \varepsilon_i \quad \text{and} \quad \mathcal{F}_i = \sigma(\phi_k, \varepsilon_k, 1 \leq k \leq i, \phi_{i+1}). \quad (6.3)$$

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

$$\mathbf{E}(e^{\lambda \xi_i} | \mathcal{F}_{i-1}) \leq e^{c|\lambda \phi_i|^\alpha} \quad \text{for all } i \in [1, n].$$

Applying Theorem 2.6 to  $(\xi_i, \mathcal{F}_i)_{i=1, \dots, n}$ , we obtain, for all  $x, \lambda, v \geq 0$ ,

$$\mathbf{P} \left( \pm (\theta_n - \theta) \sum_{k=1}^n \phi_k^2 \geq x \text{ and } \sum_{k=1}^n |\phi_k|^\alpha \leq v^\alpha \right) \leq \exp \left\{ -\lambda x + c \lambda^\alpha v^\alpha \right\}. \quad (6.4)$$

The right hand side of the last inequality takes its minimum at

$$\lambda = \lambda(x) = \left( \frac{x}{c \alpha v^\alpha} \right)^{\frac{1}{\alpha-1}}.$$

Substituting  $\lambda = \lambda(x)$  into (6.4), we obtain the desired inequality.

*Proof of Theorem 3.6.* From (3.7) and (3.8), it is easy to see that

$$(\theta'_n - \theta) \sum_{k=1}^n X_{k-1}^2 = \sum_{k=1}^n X_{k-1} \varepsilon_k.$$

For any  $i = 1, \dots, n$ , set

$$\xi_i = X_{i-1}\varepsilon_i \quad \text{and} \quad \mathcal{F}_i = \sigma\left(X_0, \varepsilon_k, 1 \leq k \leq i\right). \quad (6.5)$$

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

$$|\xi_i| \leq U_{i-1} := X_{i-1}\epsilon \quad \text{and} \quad \mathbf{E}((X_{k-1}\varepsilon_i)^2 | \mathcal{F}_{i-1}) = X_{k-1}^2 \mathbf{E}\varepsilon_i^2 \leq X_{k-1}^2 \sigma^2.$$

Applying Corollary 2.7 to  $(\xi_i, \mathcal{F}_i)$ , we obtain the desired inequality.

## References

- [1] Azuma, K., 1967. Weighted sum of certain independent random variables. *Tohoku Math. J.* **19**, No. 3, 357–367. MR-0221571
- [2] Bennett, G., 1962. Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* **57**, No. 297, 33–45.
- [3] Bentkus, V., 2003. An inequality for tail probabilities of martingales with differences bounded from one side. *J. Theoret. Probab.* **16**, No. 1, 161–173. MR-1956826
- [4] Bercu, B. and Touati, A., 2008. Exponential inequalities for self-normalized martingales with applications. *Ann. Appl. Probab.* **18**, 1848–1869. MR-2462551
- [5] Bercu, B., 2008. Inégalités exponentielles pour les martingales. Journées ALEA, 1–33.
- [6] Bernstein, S., 1927. *Theorem of Probability*. Moscow.
- [7] Courbot, B., 1999. Rates of convergence in the functional CLT for martingales. *C. R. Acad. Sci. Paris* **328**, 509–513. MR-1679980
- [8] de la Peña, V. H., 1999. A general class of exponential inequalities for martingales and ratios. *Ann. Probab.* **27**, No. 1, 537–564. MR-1681153
- [9] de la Peña, V. H., Klass, M. J. and Lai, Z. L., 2004. Self-normalized processes: exponential inequalities, moment bounds and iterated logarithm laws. *Ann. Probab.* **32**, No. 3, 1902–1933. MR-2073181
- [10] de la Peña, V. H., Klass, M. J. and Lai, Z. L., 2007. Pseudo-maximization and self-normalized processes. *Probab. Surveys* **4**, 172–192. MR-2368950
- [11] Dedecker, J. and Fan, X., 2015. Deviation inequalities for separately Lipschitz functionals of iterated random functions. *Stochastic Process. Appl.* **125**, 60–90. MR-3274692
- [12] Delyon, B., 2009. Exponential inequalities for sums of weakly dependent variables. *Electronic J. Probab.* **14**, No. 28, 752–779. MR-2495559
- [13] Dzharaparidze, K., and van Zanten, J. H., 2001. On Bernstein-type inequalities for martingales. *Stochastic Process. Appl.* **93**, 109–117. MR-1819486
- [14] Fan, X., Grama, I. and Liu, Q., 2012. Hoeffding’s inequality for supermartingales. *Stochastic Process. Appl.* **122**, 3545–3559. MR-2956116
- [15] Fan, X., Grama, I. and Liu, Q., 2012. Large deviation exponential inequalities for supermartingales. *Electron. Commun. Probab.* **17**, no. 59, 1–8. MR-3005732
- [16] Freedman, D. A., 1975. On tail probabilities for martingales. *Ann. Probab.* **3**, No. 1, 100–118. MR-0380971
- [17] Fuk, D. Kh., 1973. Some probabilistic inequalities for martingales. *Siberian. Math. J.* **14**, No. 1, 185–193. MR-0326835
- [18] Haeusler, E., 1984. An exact rate of convergence in the functional central limit theorem for special martingale difference arrays. *Probab. Theory Relat. Fields.* **65**, No. 4, 523–534. MR-0736144
- [19] Hitczenko, P., 1990. Upper bounds for the  $L_p$ -norms of martingales. *Probab. Theory Related Fields* **86**, 225–238. MR-1065280
- [20] Hoeffding, W., 1963. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13–30. MR-0144363
- [21] Jing, B. Y., Shao, Q. M. and Wang, Q., 2003. Self-normalized Cramér-type large deviations for independent random variables. *Ann. Probab.* **31**, 2167–2215. MR-2016616
- [22] Khan, R. A., 2009.  $L_p$ -version of the Dubins-Savage inequality and some exponential inequalities. *J. Theor. Probab.* **22**, 348–364. MR-2501324
- [23] Lesigne, E. and Volný, D., 2001. Large deviations for martingales. *Stochastic Process. Appl.* **96**, 143–159. MR-1856684
- [24] Liptser, R. and Spokoiny, V., 2000. Deviation probability bound for martingales with applications to statistical estimation. *Statist. Probab. Lett.* **46**, 347–357. MR-1743992

- [25] Liu, Q. and 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**, 3101–3132. MR-2568267
- [26] McDiarmid, C., 1989. On the method of bounded differences. *Surveys in combi.* MR-1036755
- [27] Nagaev, S. V., 1979. Large deviations of sums of independent random variables. *Ann. Probab.* **7**, No. 5, 745–789. MR-0542129
- [28] Pinelis, I., 1994. Optimum bounds for the distributions of martingales in Banach spaces. *Ann. Probab.* **22** 1679–1706. MR-1331198
- [29] Pinelis, I., 2006a. On normal domination of (super)martingales. *Electronic J. Probab.* **11**, 1049–1070. MR-2268536
- [30] Pinelis, I., 2006b. Binomial upper bounds on generalized moments and tail probabilities of (super)martingales with differences bounded from above. *High Dimensional probab.* **51**, 33–52. MR-2387759
- [31] Rio, E., 2013a. On McDiarmid’s concentration inequality. *Electron. Commun. Probab.* **18**, no. 44, 1–11. MR-3070910
- [32] Rio, E., 2013b. Extensions of the Hoeffding-Azuma inequalities. *Electron. Commun. Probab.* **18**, no. 54, 1–6. MR-3078017
- [33] Sason, I., 2013. Tightened exponential bounds for discrete time, conditionally symmetric martingales with bounded jumps. *Statist. Probab. Lett.* **83**, 1928–1936. MR-3069899
- [34] Shorack, G. R. and Wellner, J., 1986. *Empirical processes with application to statistics*, Wiley, New York. MR-0838963
- [35] van de Geer, S., 1995. Exponential inequalities for martingales, with application to maximum likelihood estimation for counting processes. *Ann. Statist.* **23**, 1779–1801. MR-1370307
- [36] van de Geer, S., 2002. On Hoeffding’s inequality for dependent random variables. In *Empirical Process Techniques for Dependent Data* (H. Dehling, T. Mikosch and M. Sørensen, eds.) 161–170. Birkhäuser, Boston.
- [37] Wang, Q. Y. and Jing, B. Y., 1999. An exponential nonuniform Berry-Esseen bound for self-normalized sums, *Ann. Probab.*, **27**, No. 4, 2068–2088. MR-1742902

**Acknowledgments.** We would like to thank M. Ledoux, D. Chafaï and the referee for their helpful comments and suggestions. Fan was partially supported by the Postgraduate Study Abroad Program sponsored by China Scholarship Council. Liu was partially supported by the National Natural Science Foundation of China (no. 11171044, no. 11101039, and no. 11401590).