Bernstein type inequalities for self-normalized martingales with applications

Xiequan Fan *, Shen Wang

Center for Applied Mathematics, Tianjin University, 300072 Tianjin, China.

For self-normalized martingales with conditionally symmetric differences, de la Peña [5] established the Gaussian type exponential inequalities. Bercu and Touati [2] extended de la Peña's inequalities to martingales with differences heavy on left. In this paper, we establish Bernstein type exponential inequalities for self-normalized martingales with differences bounded from below. Moreover, applications to t-statistics and autoregressive processes are discussed.

Keywords: self-normalized martingales; exponential inequalities; autoregressive processes

AMS Subject Classification: 60G42; 60E15; 60F10

1. Introduction

Let $(\xi_i)_{i\geq 1}$ be a sequence of zero-mean independent random variables satisfying $\xi_i \leq 1$ for all i. Denote $S_n = \sum_{i=1}^n \xi_i$ the partial sums of $(\xi_i)_{i\geq 1}$. Bennett [1] proved the following Bernstein type inequality: for all $x > 0$,

$$\mathbb{P}\big(S_n \geq xv^2\big) \leq \exp\bigg\{-\frac{x^2v^2}{2(1+x/3)}\bigg\}, \tag{1}$$

where $v^2 = \mathrm{Var}(S_n)$ is the variance of S_n. The importance of Bernstein type inequalities comes from the fact that they combine both the Gaussian trends and exponentially decaying rate. To see this, we rewrite the last inequality in the following form: for all $x > 0$,

$$\mathbb{P}\big(S_n \geq x\big) \leq \exp\bigg\{-\frac{x^2}{2(v^2+x/3)}\bigg\}. \tag{2}$$

It is easy to see that the last bound behaves as $\exp\{-\frac{x^2}{2v^2}\}$ for moderate $x = o(v^2)$, while it is exponentially decaying to 0 as $x \to \infty$.

The generalizations of (1) to martingales have attracted certain interest. Assume that $(\xi_i, \mathcal{F}_i)_{i=0,\cdots,n}$ is a sequence of martingale differences. If $\xi_i \leq 1$, Freedman [13] showed that (1) holds also when $\mathbb{P}\big(S_n \geq xv^2\big)$ is replaced by $\mathbb{P}\big(S_n \geq xv^2, \langle S\rangle_n \leq v^2\big)$, where $\langle S\rangle_n$ is the conditional variance of S_n. De la Peña [5], Dzhaparidze and van Zanten [8],

*Corresponding author. Email: fanxiequan@hotmail.com

and Fan et al. [12] extended Freedman's inequality to martingales with non-bounded differences. Recently, Rio [15] gave a refinements on Freedman's inequality. See also Fan et al. [10].

Despite the fact that the case of martingale is well studied, there are only a few results on Bernstein type inequalities for self-normalized martingales $S_n/[S]_n$, where $[S]_n$ is the squared variance of S_n. Among them, let us recall the following exponential inequalities of de la Peña [5]. Assume that $(\xi_i, \mathcal{F}_i)_{i=0,\cdots,n}$ is a sequence of conditionally symmetric martingale differences. Recall that ξ_i is called conditionally symmetric if $\mathcal{L}(\xi_i|\mathcal{F}_{i-1}) = \mathcal{L}(-\xi_i|\mathcal{F}_{i-1})$ for all i, where $\mathcal{L}(\xi_i|\mathcal{F}_{i-1})$ stands for the regular version of the conditional distribution of ξ_i given a σ-field $\mathcal{F}_{i-1}$. De la Peña [5] have established the following exponential inequalities for self-normalized martingales: for all $x > 0$,

$$\mathbb{P}\bigg(\frac{S_n}{[S]_n} \geq x \bigg) \leq \sqrt{\mathbb{E}\bigg[\exp\bigg\{ -\frac{1}{2}x^2[S]_n \bigg\} \bigg]}, \tag{3}$$

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

$$\mathbb{P}\bigg(\frac{S_n}{[S]_n} \geq x, [S]_n \geq y \bigg) \leq \exp\bigg\{ -\frac{1}{2}x^2 y \bigg\}, \tag{4}$$

where $[S]_n = \sum_{i=1}^n \xi_i^2$ the squared variance of S_n. In the i.i.d. case, $[S]_n/n$ usually converges almost surely to the variance of the random variables. Thus (3) and (4) can be regarded as Gaussian type inequalities.

The inequalities of de la Peña have been extended to the martingales with differences heavy on left. Recall that an integrable random variable X is called heavy on left if $\mathbb{E}X = 0$ and, for all $a > 0$, $\mathbb{E}[T_a(X)] \leq 0$, where

$$T_a(X) = \min(|X|, a)\,\mathrm{sign}(X)$$

is the truncated version of X. Clearly, conditionally symmetric martingale differences are heavy on left. Bercu and Touati [2] have obtained the following extension of de la Peña's inequity (3): for all $x > 0$,

$$\mathbb{P}\bigg(\frac{S_n}{[S]_n} \geq x \bigg) \leq \inf_{p>1} \bigg(\mathbb{E}\bigg[\exp\bigg\{ -\frac{1}{2}(p-1)x^2[S]_n \bigg\} \bigg] \bigg)^{1/p}. \tag{5}$$

They also showed that (4) holds for martingales with differences heavy on left. In the particular case $p = 2$, inequality (5) reduces to inequality (3) under the conditional symmetric assumption. Similar results for self-normalized martingales $S_n/\sqrt{[S]_n}$ can also be found in Bercu and Touati [2].

Exponential inequalities for self-normalized martingales have a lot of applications. We refer to de la Peña, Klass and Lai [6] for autoregressive processes. Bercu and Touati [2] applied such type inequalities to parameter estimations of linear regressions, autoregressive processes and branching processes. For more applications of such type inequalities, we refer to the monographs of de la Peña, Lai and Shao [7] and Bercu, Delyon and Rio [3].

In this paper, we aim to establish Bernstein type inequalities for self-normalized martingales with differences bounded from below. It is obvious that a random variable is bounded from below does not imply that it is heavy on left. Our results for self-normalized martingales are analogues to the inequalities (3)-(5). We discuss applications to t-statistics and autoregressive processes.

The paper is organized as follows. We present our main results in Section 2. In Section 3, we discuss the applications, and prove our main results in Section 4.

2. Main results

Let $(\xi_i, \mathcal{F}_i)_{i=0,\cdots,n}$ be a finite sequence of real-valued square integrable martingale differences defined on a probability space $(\Omega, \mathcal{F}, P)$, where $\xi_0 = 0$ and $\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq \ldots \subseteq \mathcal{F}_n \subseteq \mathcal{F}$ are increasing σ-fields. So by definition, we have $\mathbb{E}[\xi_i|\mathcal{F}_{i-1}] = 0, i = 1, \ldots, n$. Set

$$S_0 = 0 \quad \text{and} \quad S_k = \sum_{i=1}^{k} \xi_i \tag{6}$$

for $k = 1, \ldots, n$. Then $S = (S_k, \mathcal{F}_k)_{k=1,\ldots,n}$ is a martingale. Let $[S]$ and $\langle S \rangle$ be, respectively, the squared variance and the conditional variance of the martingale S, that is

$$[S]_0 = 0, \qquad [S]_k = \sum_{i=1}^{k} \xi_i^2,$$

and

$$\langle S \rangle_0 = 0, \qquad \langle S \rangle_k = \sum_{i=1}^{k} \mathbb{E}[\xi_i^2|\mathcal{F}_{i-1}], \quad k = 1, ..., n. \tag{7}$$

Our main result is the following Bernstein type inequalities for self-normalized martingales with differences bounded from below. It is worth to be mentioned that the inequalities are new even for independent random variables.

THEOREM 2.1 *Assume that $\xi_i \geq -1$ for all $i \in [1, n]$. Then for all $x > 0$,*

$$\begin{aligned} \mathbb{P}\Big(\frac{S_n}{[S]_n} \geq x \Big) &\leq \inf_{p>1} \left(\mathbb{E}\left[\exp\left\{ -(p-1)\Big(x - \log(1+x)\Big)[S]_n \right\} \mathbf{1}_{\{S_n \geq x[S]_n\}} \right] \right)^{1/p} &(8) \\ &\leq \inf_{p>1} \left(\mathbb{E}\left[\exp\left\{ -(p-1)\frac{x^2}{2(1+x)}[S]_n \right\} \mathbf{1}_{\{S_n \geq x[S]_n\}} \right] \right)^{1/p}, &(9) \end{aligned}$$

and, for all $y > 0$,

$$\begin{aligned} \mathbb{P}\Big(\frac{S_n}{[S]_n} \geq x,\ [S]_n \geq y \Big) &\leq \exp\Big\{ -(x - \log(1+x))y \Big\} &(10) \\ &\leq \exp\Big\{ -\frac{x^2 y}{2(1+x)} \Big\}. &(11) \end{aligned}$$

Clearly, inequality (9) implies that for all $x > 0$,

$$\mathbb{P}\left(\frac{S_n}{[S]_n} \geq x\right) \;\; \leq \;\; \inf_{p>1}\left(\mathbb{E}\left[\exp\left\{-(p-1)\frac{x^2}{2(1+x)}[S]_n\right\}\right]\right)^{1/p},$$

which is an analogue to de la Peña's inequality and the inequality of Bercu and Touati (cf. the inequalities (3) and (5)).

Similarly, for normalized martingales $S_n/\langle S\rangle_n$, we have the following Bernstein type exponential inequalities.

Theorem 2.2 *Assume that $\xi_i \geq -1$ for all $i \in [1, n]$. Then for all $x > 0$,*

$$\begin{aligned}
&\mathbb{P}\left(\frac{S_n}{\langle S\rangle_n} \leq -x\right) \\
&\leq \inf_{p>1}\left(\mathbb{E}\left[\exp\left\{-(p-1)\Big((1+x)\log(1+x)-x\Big)\langle S\rangle_n\right\}\mathbf{1}_{\{S_n\leq -x\langle S\rangle_n\}}\right]\right)^{1/p} \quad (12)\\
&\leq \inf_{p>1}\left(\mathbb{E}\left[\exp\left\{-(p-1)\frac{x^2}{2(1+x/3)}\langle S\rangle_n\right\}\mathbf{1}_{\{S_n\leq -x\langle S\rangle_n\}}\right]\right)^{1/p}. \quad (13)
\end{aligned}$$

Inequality (13) implies that for all $x > 0$ and $p > 1$,

$$\mathbb{P}\left(\frac{S_n}{\langle S\rangle_n} \leq -x\right) \;\; \leq \;\; \left(\mathbb{E}\left[\exp\left\{-(p-1)\frac{x^2}{2(1+x/3)}\langle S\rangle_n\right\}\right]\right)^{1/p}. \quad (14)$$

In the independent case, the r.h.s of the bound (14) is decreasing in p. Indeed, when $(\xi_i)_{i=1,\cdots,n}$ are independent random variables, we have $\langle S\rangle_n = \mathrm{Var}(S_n)$, where $\mathrm{Var}(S_n)$ stands for the variance of S_n. Then the bound (14) is decreasing in p.

For more exponential inequalities similar to that of Theorem 2.2, we refer to Theorem 1.3 of de la Peña [5]. In particular, de la Peña proved (13) with $p = 2$. Moreover, de la Peña also proved the following Bernstein type exponential inequalities: for all $x, y > 0$,

$$\begin{aligned}
\mathbb{P}\left(\frac{S_n}{\langle S\rangle_n} \leq -x,\ \langle S\rangle_n \geq y\right) \;\; &\leq \;\; \exp\left\{-\Big((1+x)\log(1+x)-x\Big)y\right\} \quad (15)\\
&\leq \;\; \exp\left\{-\frac{x^2 y}{2(1+x/3)}\right\}. \quad (16)
\end{aligned}$$

It is easy to see that the inequalities (15) and (16) are respectively the counterparts of (10) and (11) for $S_n/\langle S\rangle_n$.

Notice that in the independent case, the bounds (14) and (16) are exactly Bernstein's bound (1). Thus (14) and (16) can be regarded as Bernstein type inequalities for martingales.

The following deviation inequality for self-normalized martingales has its independent interest.

Theorem 2.3 *Assume that $\xi_i \geq -1$ for all $i \in [1, n]$. Then for all $b > 0, M \geq 1$ and $x > 0$,*

$$\mathbb{P}\left(\frac{S_n}{\sqrt{[S]_n}} \geq x,\, b \leq \sqrt{[S]_n} \leq bM\right) \leq \sqrt{e}\Big(1+2(1+x)\ln M\Big)\exp\left\{-\frac{x^2}{2(1+x/b)}\right\}. \quad (17)$$

Similarly, when $[S]_n$ in the left hand side of (17) is replaced by $\langle S\rangle_n$, we have the following inequality for normalized martingales. Such type inequalities are due to Liptser and Spokoiny [14].

THEOREM 2.4 *Assume that $\xi_i \geq -1$ for all $i \in [1,n]$. Then for all $b > 0, M \geq 1$ and $x > 0$,*

$$\begin{aligned}
&\mathbb{P}\Big(\frac{S_n}{\sqrt{\langle S\rangle_n}} \leq -x,\, b \leq \sqrt{\langle S\rangle_n} \leq bM\Big) \\
&\leq \sqrt{e}\Big(1+2(1+x)\ln M\Big)\exp\Big\{-\frac{x^2}{2(1+x/(3b))}\Big\}. \qquad (18)
\end{aligned}$$

It is interesting to see that in the independent case, inequality (18) with $b=\sqrt{\mathrm{Var}(S_n)}$ and $M=1$ reduces to exactly Bennett's inequality, up to an absolute constant $\sqrt{e}$. Thus the bound (18) is rather tight.

3. Applications

3.1. *Student's t-statistics*

Consider Student's t-statistic T_n defined by

$$T_n = \sqrt{n}\,\bar{\xi}\Big/\Big(\frac{1}{n-1}\sum_{j=1}^{n}(\xi_j-\bar{\xi})^2\Big)^{1/2},$$

where $\bar{\xi} = \sum_{i=1}^n \xi_i/n$. Clearly, T_n and $S_n/\sqrt{[S]_n}$ are closely related via the following identity:

$$T_n = \frac{S_n}{\sqrt{[S]_n}}\Big(\frac{n-1}{n-(S_n/\sqrt{[S]_n})^2}\Big)^{1/2}. \qquad (19)$$

Since $x/(n-x^2)^{1/2}$ is increasing on $(-\sqrt{n},\sqrt{n})$, it follows from (19) that

$$\{T_n \geq x\} = \Big\{\frac{S_n}{\sqrt{[S]_n}} \geq x\Big(\frac{n}{n+x^2-1}\Big)^{1/2}\Big\}. \qquad (20)$$

The above fact was pointed out by Efron [9]. With the help of (20), the following large deviation type result for t-statistic is an immediate consequence of Theorem 2.3.

THEOREM 3.1 *Assume that $\xi_i \geq -1$ for all $i \in [1,n]$. Then for all $b > 0, M \geq 1$ and $x > 0$,*

$$\begin{aligned}
&\mathbb{P}\big(T_n \geq x,\, b \leq \sqrt{[S]_n} \leq bM\big) \\
&\leq \sqrt{e}\Big(1+2x\Big(\frac{n}{n+x^2-1}\Big)^{1/2}\ln M\Big)\exp\Big\{-\frac{x^2\big(\frac{n}{n+x^2-1}\big)}{2\big(1+x\big(\frac{n}{n+x^2-1}\big)^{1/2}/b\big)}\Big\}. \qquad (21)
\end{aligned}$$

3.2. *Autoregressive processes*

The model of autoregressive process can be expressed as follows: for all $n \geq 0$, by

$$X_{n+1} = \theta X_n + \varepsilon_{n+1} \tag{22}$$

where X_n and ε_n are the observations and driven noises, respectively. We assume that (ε_n) is a sequence of i.i.d. centered random variables with variation $\sigma^2 > 0$ and $X_0 = \varepsilon_0$. We can estimate the unknown parameter θ by the least-squares estimator given by, for all $n \geq 1$,

$$\hat{\theta}_n = \frac{\sum_{k=1}^n X_{k-1} X_k}{\sum_{k=1}^n X_{k-1}^2} \tag{23}$$

Bercu and Touati [2] has established the convergence rate of $\hat{\theta}_n - \theta$ when X_0 and (ε_n) are normal random variables. Here, we would like to give a convergence rate of $\hat{\theta}_n - \theta$ for the case that the driven noises (ε_n) are bounded. Applying Theorem 2.2 and de la Peña's inequality (16), we have the following exponential inequalities.

Theorem 3.2 *Assume that $|\varepsilon_i| \leq C$ for some positive constant C and all i. If $|\theta| < 1$, then for all $x > 0$,*

$$\mathbb{P}\Big(|\hat{\theta}_n - \theta| \geq x \Big) \leq 2 \inf_{p>1} \left(\mathbb{E}\left[\exp\left\{ -(p-1) \frac{x^2}{2\big(\sigma^2 + \frac{x\,C^2}{3(1-|\theta|)}\big)} \sum_{k=1}^n X_{k-1}^2 \right\} \right] \right)^{1/p}, \tag{24}$$

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

$$\mathbb{P}\Big(|\hat{\theta}_n - \theta| \geq x,\ \sum_{k=1}^n X_{k-1}^2 \geq y \Big) \leq 2 \exp\left\{ - \frac{x^2 y}{2\big(\sigma^2 + \frac{x\,C^2}{3(1-|\theta|)}\big)} \right\}. \tag{25}$$

Inequality (25) is similar to an exponential inequalities of de la Peña, Klass and Lai [6], which states that when (ε_n) are the standard normal random variables, it holds for all $x, y > 0$,

$$\mathbb{P}\Big(|\hat{\theta}_n - \theta| \geq x,\ \sum_{k=1}^n X_{k-1}^2 \geq y \Big) \leq 2 \exp\left\{ -\frac{1}{2} x^2 y \right\}. \tag{26}$$

By Theorem 2.4, we obtain the following result.

Theorem 3.3 *Assume that $|\varepsilon_i| \leq C$ for some positive constant C and all i. If $|\theta| < 1$, then for all $b > 0, M \geq 1$ and $x > 0$,*

$$\begin{aligned} &\mathbb{P}\Big(\big|\widehat{\theta}_n - \theta\big| \sqrt{\Sigma_{k=1}^n X_{k-1}^2} \geq x,\ b \leq \sqrt{\Sigma_{k=1}^n X_{k-1}^2} \leq bM \Big) \\ &\leq 2\sqrt{e} \Big(1 + 2(1 + \frac{x}{\sigma}) \ln M \Big) \exp\left\{ - \frac{x^2}{2\big(\sigma^2 + \frac{x\,C^2}{3(1-|\theta|)}\big)} \right\}. \end{aligned} \tag{27}$$

4. Proofs of Theorems

4.1. *Preliminary lemmas*

The following technical lemma is from Fan et al. [11]. For reader's convenience, we shall give a proof following [11].

LEMMA 4.1 *Assume that $\xi_i \geq -1$ for all $i \in [1, n]$. For any $\lambda \in [0,1)$, denote*

$$U_n(\lambda) = \exp\Big\{\lambda S_n + (\lambda + \log(1-\lambda))[S]_n\Big\}.$$

Then $(U_i(\lambda), \mathcal{F}_i)_{i=0,\cdots,n}$ is a supermartingale, and satisfies that for all $\lambda \in [0,1)$,

$$\mathbb{E}\big[U_n(\lambda)\big] \leq 1. \tag{28}$$

Proof. Assume $\xi_i \geq -1$ and $\lambda \in [0,1)$, then $\lambda\xi_i \geq -\lambda > -1$. Since the function

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

is increasing in x, we obtain

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

Therefore, we have

$$\exp\big\{\lambda\xi_i + \xi_i^2(\lambda + \log(1-\lambda))\big\} \leq 1 + \lambda\xi_i.$$

Since $\mathbb{E}[\xi_i|\mathcal{F}_{i-1}] = 0$, it follows that

$$\mathbb{E}\Big[\exp\big\{\lambda\xi_i + (\lambda + \log(1-\lambda))\xi_i^2\big\} \,\Big|\, \mathcal{F}_{i-1}\Big] \leq 1.$$

For all $\lambda \in [0,1)$ and $n \geq 0$, we have

$$U_n(\lambda) = U_{n-1}(\lambda)\exp\Big\{\lambda\xi_n + (\lambda + \log(1-\lambda))\xi_n^2\Big\}.$$

Hence, we deduce that for all $\lambda \in [0,1)$,

$$\begin{aligned} \mathbb{E}[U_n(\lambda)|\mathcal{F}_{n-1}] &= U_{n-1}(\lambda)\mathbb{E}\Big[\exp\Big\{\lambda\xi_n + (\lambda + \log(1-\lambda))\xi_n^2\Big\}\Big|\mathcal{F}_{n-1}\Big] \\ &\leq U_{n-1}(\lambda), \end{aligned}$$

which means $(U_i(\lambda), \mathcal{F}_i)_{i=0,\cdots,n}$ is a positive supermartingale. Moreover, it holds

$$\mathbb{E}[U_n(\lambda)] \leq \mathbb{E}[U_{n-1}(\lambda)] \leq ... \leq \mathbb{E}[U_1(\lambda)] \leq 1.$$

This completes the proof of Lemma 4.1. ■

In order to prove Theorem 2.2, we need the following lemma of Freedman [13].

Lemma 4.2 *Assume that $\xi_i \geq -1$ for all $i \in [1, n]$. Denote*

$$W_n(\lambda) = \exp\Big\{ -\lambda S_n - (e^{\lambda} - 1 - \lambda)\langle S\rangle_n \Big\}, \quad \lambda \geq 0.$$

Then $(W_i(\lambda), \mathcal{F}_i)_{i=0,\cdots,n}$ is a supermartingale, and satisfies that

$$\mathbb{E}\big[W_n(\lambda)\big] \leq 1. \tag{29}$$

4.2. *Proof of Theorem 2.1*

We follow the method of Bercu and Touati [2]. Let $A_n = \{S_n \geq x[S]_n\}, x > 0$. By Markov's inequality, Hölder's inequality and Lemma 4.1, we have for all $\lambda \in [0, 1)$ and $q > 1$,

$$\begin{aligned}
\mathbb{P}(A_n) &\leq \mathbb{E}\bigg[\exp\bigg\{\frac{\lambda}{q}\Big(S_n - x[S]_n\Big)\bigg\}\mathbf{1}_{A_n}\bigg] \\
&= \mathbb{E}\bigg[\exp\bigg\{\frac{1}{q}\Big(\lambda S_n + (\lambda + \log(1-\lambda))[S]_n\Big)\bigg\}\exp\bigg\{\frac{1}{q}\Big(-\lambda - \log(1-\lambda) - \lambda x\Big)[S]_n\bigg\}\mathbf{1}_{A_n}\bigg] \\
&\leq \bigg(\mathbb{E}\bigg[\exp\bigg\{\frac{p}{q}\Big(-\lambda - \log(1-\lambda) - \lambda x\Big)[S]_n\bigg\}\mathbf{1}_{A_n}\bigg]\bigg)^{1/p}\Big(\mathbb{E}\big[U_n(\lambda)\big]\Big)^{1/q} \\
&\leq \bigg(\mathbb{E}\bigg[\exp\bigg\{-\frac{p}{q}\Big(\lambda + \log(1-\lambda) + \lambda x\Big)[S]_n\bigg\}\mathbf{1}_{A_n}\bigg]\bigg)^{1/p},
\end{aligned} \tag{30}$$

where $p = 1 + p/q$. Consequently, as $p/q = p - 1$, we can deduce from (30) that

$$\mathbb{P}(A_n) \leq \inf_{p>1}\bigg(\mathbb{E}\bigg[\exp\Big\{-(p-1)(\lambda + \log(1-\lambda) + \lambda x)[S]_n\Big\}\mathbf{1}_{A_n}\bigg]\bigg)^{1/p}.$$

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

$$\underline{\lambda} = \lambda(x) = \frac{x}{1+x},$$

therefore we obtain

$$\mathbb{P}(A_n) \leq \inf_{p>1}\bigg(\mathbb{E}\bigg[\exp\Big\{-(p-1)(x - \log(1+x))[S]_n\Big\}\bigg]\mathbf{1}_{A_n}\bigg)^{1/p}.$$

Using the following inequality

$$x - \log(1+x) \geq \frac{x^2}{2(1+x)}, \qquad x > 0, \tag{31}$$

we deduce that

$$
\begin{aligned}
\mathbb{P}(A_n) &\leq \inf_{p>1} \left(\mathbb{E}\left[\exp\left\{ -(p-1)(x-\log(1+x))[S]_n \right\} \mathbf{1}_{A_n} \right] \right)^{1/p} \\
&\leq \inf_{p>1} \left(\mathbb{E}\left[\exp\left\{ -(p-1)\frac{x^2}{2(1+x)}[S]_n \right\} \mathbf{1}_{A_n} \right] \right)^{1/p},
\end{aligned}
$$

which gives the first two desired inequalities. ∎

Next we prove the last two desired inequalities. Denote $B_n = \{S_n \geq x[S]_n,\ [S]_n \geq y\}$. By an argument similar to the proof of (30), we deduce that for all $q > 1$,

$$
\begin{aligned}
\mathbb{P}(B_n) &\leq \left(\mathbb{E}\left[\exp\left\{ \frac{p}{q}\Big(-\underline{\lambda} - \log(1-\underline{\lambda}) - \underline{\lambda} x \Big)[S]_n \right\} \mathbf{1}_{B_n} \right] \right)^{1/p} \\
&\leq \left(\mathbb{E}\left[\exp\left\{ -\frac{p}{q}\Big(\underline{\lambda} + \log(1-\underline{\lambda}) + \underline{\lambda} x \Big) y \right\} \right] \right)^{1/p} \\
&= \exp\left\{ -\frac{1}{q}\Big(x - \log(1+x) \Big) y \right\}.
\end{aligned}
$$

Therefore, by (31), it holds

$$
\begin{aligned}
\mathbb{P}(B_n) &\leq \inf_{q>1} \exp\left\{ -\frac{1}{q}\Big(x - \log(1+x) \Big) y \right\} \\
&= \exp\left\{ -\Big(x - \log(1+x) \Big) y \right\} \\
&\leq \exp\left\{ -\frac{x^2\, y}{2(1+x)} \right\},
\end{aligned}
$$

which gives the last two desired inequalities.

4.3. *Proof of Theorem 2.2*

For all $x > 0$, denote by

$$
D_n = \{-S_n \geq x\langle S\rangle_n\}.
$$

By exponential Markov's inequality, we deduce that for all $\lambda \in [0, 3)$ and $q > 1$,

$$
\begin{aligned}
\mathbb{P}(D_n) &\leq \mathbb{E}\left[\exp\left\{ \frac{\lambda}{q}\Big(-S_n - x\langle S\rangle_n \Big) \right\} \mathbf{1}_{D_n} \right] \\
&= \mathbb{E}\left[\exp\left\{ -\frac{\lambda}{q} S_n - \frac{e^\lambda - 1 - \lambda}{q}\langle S\rangle_n \right\} \exp\left\{ \Big(\frac{e^\lambda - 1 - \lambda}{q} - \frac{\lambda x}{q} \Big)\langle S\rangle_n \right\} \mathbf{1}_{D_n} \right].
\end{aligned}
$$

Using Hölder's inequality and Lemma 4.2, we have for all $\lambda \in [0, 3)$ and $q > 1$,

$$
\begin{aligned}
\mathbb{P}(D_n) &\leq \left(\mathbb{E}\Big[\exp\Big\{\Big(\frac{p(e^\lambda - 1 - \lambda)}{q} - \frac{p\lambda x}{q}\Big)\langle S\rangle_n\Big\}\mathbf{1}_{D_n}\Big]\right)^{1/p}\Big(\mathbb{E}\big[W_n(t)\big]\Big)^{1/q} \\
&\leq \left(\mathbb{E}\Big[\exp\Big\{\frac{p}{q}\Big(e^\lambda - 1 - \lambda - \lambda x\Big)\langle S\rangle_n\Big\}\mathbf{1}_{D_n}\Big]\right)^{1/p}, \qquad (32)
\end{aligned}
$$

where $p = 1 + p/q$. Consequently, as $p/q = p - 1$, we can deduce from (32) that

$$
\mathbb{P}(D_n) \leq \inf_{p>1}\left(\mathbb{E}\Big[\exp\Big\{(p-1)\Big(e^\lambda - 1 - \lambda - \lambda x\Big)\langle S\rangle_n\Big\}\mathbf{1}_{D_n}\Big]\right)^{1/p}. \qquad (33)
$$

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

$$
\overline{\lambda} = \lambda(x) := \log(1 + x).
$$

Substituting $\overline{\lambda} = \lambda(x)$ in (33), we obtain

$$
\mathbb{P}(D_n) \leq \inf_{p>1}\left(\mathbb{E}\Big[\exp\Big\{-(p-1)\Big((1+x)\log(1+x) - x\Big)\langle S\rangle_n\Big\}\mathbf{1}_{D_n}\Big]\right)^{1/p}. \qquad (34)
$$

Using the inequality

$$
e^\lambda - 1 - \lambda \leq \frac{\lambda^2}{2(1 - \lambda/3)}, \qquad \lambda \in [0, 3), \qquad (35)
$$

we get for all $x \geq 0$,

$$
\begin{aligned}
(1+x)\log(1+x) - x &= \inf_{\lambda \geq 0}\Big(e^\lambda - 1 - \lambda - \lambda x\Big) \\
&\leq \inf_{\lambda \geq 0}\Big(\frac{\lambda^2}{2(1-\lambda/3)} - \lambda x\Big) \\
&= \exp\left\{-\frac{x^2}{1 + x/3 + \sqrt{1 + 2x/3}}\right\} \\
&\leq \exp\left\{-\frac{x^2}{2(1 + x/3)}\right\},
\end{aligned}
$$

where the last inequality follows from the fact $\sqrt{1 + 2x/3} \leq 1 + x/3$. Thus, from (34), we obtain for all $x \geq 0$,

$$
\begin{aligned}
\mathbb{P}(D_n) &\leq \inf_{p>1}\left(\mathbb{E}\Big[\exp\Big\{-(p-1)\Big((1+x)\log(1+x) - x\Big)\langle S\rangle_n\Big\}\mathbf{1}_{D_n}\Big]\right)^{1/p} \\
&\leq \inf_{p>1}\left(\mathbb{E}\Big[\exp\Big\{-(p-1)\frac{x^2}{1 + x/3 + \sqrt{1 + 2x/3}}\langle S\rangle_n\Big\}\mathbf{1}_{D_n}\Big]\right)^{1/p} \\
&\leq \inf_{p>1}\left(\mathbb{E}\Big[\exp\Big\{-(p-1)\frac{x^2}{2(1 + x/3)}\langle S\rangle_n\Big\}\mathbf{1}_{D_n}\Big]\right)^{1/p}.
\end{aligned}
$$

This proves (12) and (14). ∎

4.4. *Proof of Theorem 2.3*

The proof of Theorem 2.3 is based on a modified method of Liptser and Spokoiny [14]. Given $a > 1$, introduce the geometric series $b_k = ba^k$ and define random events

$$C_k = \left\{ \frac{S_n}{\sqrt{[S]_n}} \geq x,\ b_k \leq \sqrt{[S]_n} < b_{k+1} \right\}, \quad k = 0, 1, \ldots, K,$$

where K stands for the integer part of $\log_a M$. Clearly, it holds

$$\left\{ \frac{S_n}{\sqrt{[S]_n}} \geq x,\ b \leq \sqrt{[S]_n} \leq bM \right\} \subseteq \bigcup_{k=0}^{K} C_k, \tag{36}$$

which leads to

$$\mathbb{P}\left(\frac{S_n}{\sqrt{[S]_n}} \geq x,\ b \leq \sqrt{[S]_n} \leq bM \right) \leq \sum_{k=0}^{K} \mathbb{P}(C_k). \tag{37}$$

Notice that

$$\lambda + \log(1-\lambda) \geq -\frac{\lambda^2}{2(1-\lambda)}, \quad \lambda \in [0, 1).$$

For any $\lambda \in [0, 1)$, the last inequality and (28) together implies that

$$\mathbb{E}\Big[\exp\Big\{ \lambda S_n - \frac{\lambda^2}{2(1-\lambda)} [S]_n \Big\} \mathbf{1}_{C_k} \Big] \leq 1.$$

Next, taking $\lambda_k = x/(x + b_k)$, for any $x > 0$, we obtain

$$\begin{aligned}
1 \ &\geq\ \mathbb{E}\left[\exp\left\{ \frac{x}{x+b_k} S_n - \frac{x^2}{2b_k(x+b_k)} [S]_n \right\} \mathbf{1}_{C_k} \right] \\
&\geq\ \mathbb{E}\left[\exp\left\{ \frac{x^2}{x+b_k} \sqrt{[S]_n} - \frac{x^2}{2b_k(x+b_k)} [S]_n \right\} \mathbf{1}_{C_k} \right] \\
&\geq\ \mathbb{E}\left[\exp\left\{ \inf_{b_k \leq c < b_{k+1}} \left(\frac{x^2 c}{x+b_k} - \frac{x^2 c^2}{2b_k(x+b_k)} \right) \right\} \mathbf{1}_{C_k} \right] \\
&\geq\ \mathbb{E}\left[\exp\left\{ \frac{x^2 b_{k+1}}{x+b_k} - \frac{x^2 b_{k+1}^2}{2b_k(x+b_k)} \right\} \mathbf{1}_{C_k} \right],
\end{aligned}$$

which implies that

$$\begin{aligned}
\mathbb{P}(C_k) \ &\leq\ \exp\left\{ -\frac{x^2}{1+x/b_k} \Big(a - \frac{a^2}{2} \Big) \right\} \\
&\leq\ \exp\left\{ -\frac{x^2}{1+x/b} \Big(a - \frac{a^2}{2} \Big) \right\}.
\end{aligned}$$

Finally, we may pick a to make the right-hand side of the last bound as small as possible. This leads to the choice $a = 1 + \frac{1}{1+x}$, so that

$$x^2\left(a - \frac{a^2}{2}\right) = x^2\left\{1 + \frac{1}{1+x} - \frac{1}{2}\left(1 + \frac{1}{1+x}\right)^2\right\} \geq \frac{1}{2}(x^2 - 1).$$

Since $\log(1+\frac{1}{1+x}) \geq \frac{1}{2(1+x)}$ for $x \geq 0$, we obtain $\log_a M \leq 2(1+x)\ln M$, and (17) follows by (37). ■

4.5. *Proof of Theorem 2.4*

The proof of Theorem 2.4 is similar to the proof of Theorem 2.3. Given $a > 1$, introduce the geometric series $b_k = ba^k$ and define random events

$$H_k = \left\{\frac{-S_n}{\sqrt{\langle S\rangle_n}} \geq x,\ b_k \leq \sqrt{\langle S\rangle_n} < b_{k+1}\right\}, \quad k = 0, 1, \ldots, K,$$

where K stands for the integer part of $\log_a M$. Clearly, it holds

$$\left\{\frac{-S_n}{\sqrt{\langle S\rangle_n}} \geq x,\ b \leq \sqrt{\langle S\rangle_n} \leq bM\right\} \subseteq \bigcup_{k=0}^{K} H_k, \tag{38}$$

which leads to

$$\mathbb{P}\left(\frac{-S_n}{\sqrt{\langle S\rangle_n}} \geq x,\ b \leq \sqrt{\langle S\rangle_n} \leq bM\right) \leq \sum_{k=0}^{K} \mathbb{P}(H_k). \tag{39}$$

Notice that

$$e^{\lambda} - 1 - \lambda \leq \frac{\lambda^2}{2(1-\lambda/3)}, \quad \lambda \in [0, 3).$$

For any $\lambda \in [0, 3)$, the last inequality and Lemma 4.2 together implies that

$$\mathbb{E}\Big[\exp\Big\{\lambda(-S_n) - \frac{\lambda^2}{2(1-\lambda/3)}\langle S\rangle_n\Big\}\mathbf{1}_{H_k}\Big] \leq 1.$$

Next, taking $\lambda_k = x/(b_k + x/3)$, for any $x > 0$, we obtain

$$\begin{aligned}
1 &\geq \mathbb{E}\Big[\exp\Big\{\frac{x}{b_k + x/3}(-S_n) - \frac{x^2}{b_k(b_k + x/3)}\langle S\rangle_n\Big\}\mathbf{1}_{H_k}\Big] \\
&\geq \mathbb{E}\Big[\exp\Big\{\frac{x^2}{b_k + x/3}\sqrt{\langle S\rangle_n} - \frac{x^2}{b_k(b_k + x/3)}\langle S\rangle_n\Big\}\mathbf{1}_{H_k}\Big] \\
&\geq \mathbb{E}\Big[\exp\Big\{\inf_{b_k \leq c < b_{k+1}}\Big(\frac{x^2 c}{b_k + x/3} - \frac{x^2 c^2}{b_k(b_k + x/3)}\Big)\Big\}\mathbf{1}_{H_k}\Big] \\
&\geq \mathbb{E}\Big[\exp\Big\{\frac{x^2 b_{k+1}}{b_k + x/3} - \frac{x^2 b_{k+1}^2}{b_k(b_k + x/3)}\Big\}\mathbf{1}_{H_k}\Big],
\end{aligned}$$

which implies that

$$\begin{aligned}
\mathbb{P}(H_k) &\leq \exp\Big\{ -\frac{x^2}{1+\frac{x}{3b_k}}a + \frac{x^2}{(1+\frac{x}{3b_k})}\frac{a^2}{2}\Big)\Big\} \\
&\leq \exp\Big\{ -\frac{x^2}{1+\frac{x}{3b_k}}\Big(a - \frac{a^2}{2}\Big)\Big\} \\
&\leq \exp\Big\{ -\frac{x^2}{1+\frac{x}{3b}}\Big(a - \frac{a^2}{2}\Big)\Big\}.
\end{aligned}$$

Finally, taking $a = 1 + \frac{1}{1+x}$, we obtain the desired inequality from (39), with an argument similar to the proof of Theorem 2.3. ■

4.6. *Proof of Theorem 3.2*

By (22), we have $X_k = \sum_{i=0}^{k} \theta^{k-i}\varepsilon_i$. Since $|\theta| < 1$ and $|\varepsilon_i| \leq C$, we deduce that for all k,

$$|X_k| \leq C\sum_{i=0}^{k}|\theta|^{k-i} \leq \frac{C}{1-|\theta|}.$$

From (22) and (23), it is easy to see that for all $n \geq 1$,

$$\hat{\theta}_n - \theta = \frac{\sum_{k=1}^{n} X_{k-1}\varepsilon_k}{\sum_{k=1}^{n} X_{k-1}^2}. \tag{40}$$

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

$$\xi_i = X_{i-1}\varepsilon_i(1-|\theta|)/C^2 \ \text{ and } \ \mathcal{F}_i = \sigma\big(\varepsilon_k, 0 \leq k \leq i\big).$$

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

$$|\xi_i| \leq 1$$

and

$$\langle S\rangle_n = \sum_{k=1}^{n} \mathbb{E}\big[\xi_i^2 \big| \mathcal{F}_{i-1}\big] = \frac{\sigma^2(1-|\theta|)^2}{C^4} \sum_{k=1}^{n} X_{i-1}^2.$$

Thus we have

$$\hat{\theta}_n - \theta = \frac{(1-|\theta|)\sigma^2}{C^2} \frac{S_n}{\langle S\rangle_n}.$$

Applying inequality (14) to $(\xi_i, \mathcal{F}_i)_{i=1,\dots,n}$, we deduce that for all $x > 0$,

$$\begin{aligned}
&\mathbb{P}\Big(\hat{\theta}_n - \theta \leq -x\Big) \\
&= \mathbb{P}\Big(\frac{S_n}{\langle S\rangle_n} \leq -\frac{x\, C^2}{(1-|\theta|)\sigma^2}\Big) \\
&\leq \inf_{p>1} \bigg(\mathbb{E}\bigg[\exp\bigg\{-(p-1)\frac{x^2}{2(\sigma^2 + xC^2/(3(1-|\theta|)))} \sum_{k=1}^{n} X_{k-1}^2\bigg\}\bigg]\bigg)^{1/p}. \qquad (41)
\end{aligned}$$

Similarly, applying inequality (14) to $(-\xi_i, \mathcal{F}_i)_{i=1,\dots,n}$, we have for all $x > 0$,

$$\mathbb{P}\Big(\hat{\theta}_n - \theta \geq x\Big) \leq \inf_{p>1} \bigg(\mathbb{E}\bigg[\exp\bigg\{-(p-1)\frac{x^2}{2(\sigma^2 + xC^2/(3(1-|\theta|)))} \sum_{k=1}^{n} X_{k-1}^2\bigg\}\bigg]\bigg)^{1/p}. \quad (42)$$

Combining (41) and (42) together, we obtain for all $x > 0$,

$$\mathbb{P}\Big(|\hat{\theta}_n - \theta| \geq x\Big) \leq 2\inf_{p>1} \bigg(\mathbb{E}\bigg[\exp\bigg\{-(p-1)\frac{x^2}{2(\sigma^2 + xC^2/(3(1-|\theta|)))} \sum_{k=1}^{n} X_{k-1}^2\bigg\}\bigg]\bigg)^{1/p},$$

which gives the first desired inequality. Applying de la Peña's inequality (16) to $(\xi_i, \mathcal{F}_i)_{i=1,\dots,n}$, we get for all $x, y > 0$,

$$\begin{aligned}
\mathbb{P}\Big(\hat{\theta}_n - \theta \leq -x,\ \sum_{k=1}^{n} X_{k-1}^2 \geq y\Big) &= \mathbb{P}\Big(\frac{S_n}{\langle S\rangle_n} \leq -\frac{x\, C^2}{(1-|\theta|)\sigma^2},\ \langle S\rangle_n \geq y\frac{(1-|\theta|)^2\sigma^2}{C^4}\Big) \\
&\leq \exp\bigg\{-\frac{x^2 y}{2(\sigma^2 + xC^2/(3(1-|\theta|)))}\bigg\}. \qquad (43)
\end{aligned}$$

Similarly, applying de la Peña's inequality (16) to $(-\xi_i, \mathcal{F}_i)_{i=1,\dots,n}$, we have for all $x, y > 0$,

$$\mathbb{P}\Big(\hat{\theta}_n - \theta \geq x,\ \sum_{k=1}^{n} X_{k-1}^2 \geq y\Big) \leq \exp\bigg\{-\frac{x^2 y}{2(\sigma^2 + xC^2/(3(1-|\theta|)))}\bigg\}. \qquad (44)$$

Combining (43) and (44) together, we obtain for all $x, y > 0$,

$$\mathbb{P}\Big(|\hat{\theta}_n - \theta| \geq x,\ \sum_{k=1}^{n} X_{k-1}^2 \geq y\Big) \leq 2\exp\bigg\{-\frac{x^2 y}{2(\sigma^2 + xC^2/(3(1-|\theta|)))}\bigg\},$$

which gives the second desired inequality. ■

4.7. *Proof of Theorem 3.3*

Recall the notations in the proof of Theorem 3.2. It is easy to see that

$$(\widehat{\theta}_n - \theta)\sqrt{\Sigma_{k=1}^n X_{k-1}^2} = \frac{\sum_{k=1}^n X_{k-1}\varepsilon_k}{\sqrt{\sum_{k=1}^n X_{k-1}^2}} = \sigma \frac{S_n}{\sqrt{\langle S\rangle_n}}.$$

Therefore, by Theorem 2.4, for all $b > 0, M \geq 1$ and $x > 0$,

$$\begin{aligned}
&\mathbb{P}\Big((\widehat{\theta}_n - \theta)\sqrt{\Sigma_{k=1}^n X_{k-1}^2} \leq -x,\ b \leq \sqrt{\Sigma_{k=1}^n X_{k-1}^2} \leq bM \Big) \\
&\leq \mathbb{P}\Big(\frac{S_n}{\sqrt{\langle S\rangle_n}} \leq -\frac{x}{\sigma},\ b\frac{(1-|\theta|)\sigma}{C^2} \leq \sqrt{\langle S\rangle_n} \leq bM\frac{(1-|\theta|)\sigma}{C^2} \Big) \\
&\leq \sqrt{e}\Big(1 + 2(1+\frac{x}{\sigma}) \ln M \Big) \exp\Big\{ - \frac{x^2}{2(\sigma^2 + xC^2/(3b(1-|\theta|)))} \Big\}.
\end{aligned}$$

Similarly, the same bound holds for the tail probabilities

$$\mathbb{P}\Big((\widehat{\theta}_n - \theta)\sqrt{\Sigma_{k=1}^n X_{k-1}^2} \geq x,\ b \leq \sqrt{\Sigma_{k=1}^n X_{k-1}^2} \leq bM \Big).$$

Hence, we have for all $b > 0, M \geq 1$ and $x > 0$,

$$\begin{aligned}
&\mathbb{P}\Big(|\widehat{\theta}_n - \theta|\sqrt{\Sigma_{k=1}^n X_{k-1}^2} \geq x,\ b \leq \sqrt{\Sigma_{k=1}^n X_{k-1}^2} \leq bM \Big) \\
&\leq 2\sqrt{e}\Big(1 + 2(1+\frac{x}{\sigma}) \ln M \Big) \exp\Big\{ - \frac{x^2}{2(\sigma^2 + xC^2/(3b(1-|\theta|)))} \Big\},
\end{aligned}$$

which gives the desired inequality. ■

Acknowledgements

We would like to thank the editor and two referees for their valuable comments and remarks. The work has been partially supported by the National Natural Science Foundation of China (grants numbers 11601375, 11501146 and 11626250).

References

[1] Bennett G. Probability inequalities for the sum of independent random variables. J Amer Statist Assoc. 1962;57(297):33-45.
[2] Bercu B, Touati A. Exponential inequalities for self-normalized martingales with applications. Ann Appl Probab. 2008;18:1848-1869.
[3] Bercu B, Delyon B, Rio E. Concentration inequalities for sums and martingales. New York: Springer; 2015.
[4] Bernstein S. The Theory of Probabilities. Moscow, Leningrad (in Russian); 1946.
[5] de la Peña VH. A general class of exponential inequalities for martingales and ratios. Ann Probab. 1999;27(1):537-564.

[6] de la Peña VH, Klass MJ, Lai TL. Pseudo-maximization and self-normalized processes. Probab. Surveys 2007;4:172-192.

[7] de la Peña VH, Lai TL, Shao QM. Self-normalized processes: Limit theory and Statistical Applications. Springer; 2008.

[8] Dzhaparidze K, van Zanten JH, On Bernstein-type inequalities for martingales. Stochastic Process Appl. 2001;93:109-117.

[9] Efron B. Student's t-test under symmetry conditions. J Amer Statist Assoc. 1969;64:1278-1302.

[10] Fan X, Grama I, Liu Q. Hoeffding's inequality for supermartingales. Stochastic Process Appl 2012;122:3545C3559.

[11] Fan X, Grama I, Liu Q. Exponential inequalities for martingales with applications. Electron J Probab. 2015;20(1):1-22.

[12] Fan X, Grama I, Liu Q. Martingale inequalities of type Dzhaparidze and van Zanten. Statistics. 2017;51(6):1200-1213.

[13] Freedman DA, On tail probabilities for martingales. Ann Probab. 1975;3(1):100-118.

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

[15] Rio E. New deviation inequalities for martingales with bounded increments. Stochastic Process Appl. 2017;137(5):1637-1648.