Berry-Esseen bounds for self-normalized martingales

Xiequan Fan[a], Qi-Man Shao[b]

[a] *Center for Applied Mathematics, Tianjin University, Tianjin 300072, China*
[b] *Department of Statistics, The Chinese University of Hong Kong, Shatin, NT, Hong Kong*

Abstract

A Berry-Esseen bound is obtained for self-normalized martingales under the assumption of finite moments. The bound coincides with the classical Berry-Esseen bound for standardized martingales. An example is given to show the optimality of the bound. Applications to Student's statistic and autoregressive process are also discussed.

Keywords: Self-normalized process, Berry-Esseen bounds, martingales, Student's statistic, autoregressive process

2000 MSC: Primary 60G42; 60F05; Secondary 60E15

---

1. Introduction

Let $(X_i)_{i\geq 1}$ be a sequence of independent non-degenerate real-valued random variables with zero means, and let

$$S_n = \sum_{i=1}^n X_i \quad \text{and} \quad V_n^2 = \sum_{i=1}^n X_i^2$$

be the partial sum and the partial quadratic sum, respectively. The self-normalized sum is defined as S_n/V_n. The study of the asymptotic behavior of self-normalized sums has a long history. When $(X_i)_{i\geq 1}$ are i.i.d. in the domain of normal and stable law, Logan et al. [11] obtained the weak convergence for the self-normalized sum, while Giné et al. [5] proved that S_n/V_n is asymptotically normal if and only if X_1 belongs to the domain of attraction of a normal law. Under the same necessary and sufficient condition, Csörgő et al. [3] proved a self-normalized type Donsker's theorem. For general independent random variables with finite $(2+\delta)^{th}$ moments, where $0 < \delta \leq 1$, Bentkus, Bloznelis and Götze [2] (see also Bentkus and Götze [1] for i.i.d. case) have obtained the following Berry-Esseen bound : If $\mathbf{E}|X_i|^{2+\delta} < \infty$ for $\delta \in (0, 1]$, then these exists an absolute constant C such that

$$\sup_x \Big| \mathbf{P}(S_n/V_n \leq x) - \Phi(x) \Big| \ \ \leq \ \ C\, B_n^{-2-\delta} \sum_{i=1}^n \mathbf{E}|X_i|^{2+\delta},$$

where $B_n^2 = \sum_{i=1}^n \mathbf{E} X_i^2$, and $\Phi(x)$ is the standard normal distribution function. It is worth noting that the last bound coincides with the classical Berry-Esseen bound for standardized partial sums S_n/B_n and it is the best possible. For the related error of $\mathbf{P}(S_n/V_n \geq x)$ to $1-\Phi(x)$, we refer to Shao [12], Jing, Shao and Wang [9]. In these papers, self-normalized Cramér type moderate deviation theorems have been established. We also refer to de la Peña, Lai and Shao [4], Shao and Wang [13] and Shao and Zhou [14] for surveys on recent developments on self-normalized limit theory.

Despite the fact that the case for self-normalized sums of independent random variables is well studied, we are not aware of Berry-Esseen bounds for self-normalized martingales in the literature. The main purpose of this paper is to fill this gap.

We first recall some Berry-Esseen bounds for standardized martingale difference sequence. Let $(X_i, \mathcal{F}_i)_{i=0,...,n}$ be a finite sequence of martingale differences defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, where $X_0 = 0$ and $\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq ... \subseteq \mathcal{F}_n \subseteq \mathcal{F}$ are increasing σ-fields. Set

$$S_0 = 0, \qquad S_k = \sum_{i=1}^k X_i, \quad k = 1, ..., n. \tag{1}$$

Then $S = (S_k, \mathcal{F}_k)_{k=0,...,n}$ is a martingale. Let $[S]$ and $\langle S \rangle$ be, respectively, the squared variance and the conditional variance of the martingale S, that is

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

and

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

Suppose that $\mathbf{E}|X_i|^{2p} < \infty$ for some $p > 1$ and all $i = 1, ..., n$. Define

$$N_n = \sum_{i=1}^n \mathbf{E}|X_i|^{2p} + \mathbf{E}|\langle S \rangle_n - 1|^p. \tag{2}$$

When $p \in (1, 2]$, Heyde and Brown [8] (see also Theorem 3.10 of Hall and Heyde [7]) proved that there exits a constant C_p depending only on p such that

$$\sup_{x \in \mathbf{R}} \Big| \mathbf{P}(S_n \leq x) - \Phi(x) \Big| \leq C_p \, N_n^{1/(2p+1)}. \tag{3}$$

Later, Haeusler [6] gave an extension of (3) to all $p \in (1, \infty)$. Moreover, Haeusler also gave an example to justify that his bound is asymptotically the best possible. It is remarked that the $(X_i)_{1 \leq i \leq n}$ is standardized, that is, $\sum_{i=1}^n \mathbf{E} X_i^2$ is close to 1.

In this paper, we prove that the Berry-Esseen bound (3) also holds for self-normalized martingales $S_n/\sqrt{[S]_n}$ and normalized martingales $S_n/\sqrt{\langle S\rangle_n}$. Moreover, we also justify the optimality of our bounds. Applications to Student's statistic and autoregressive process are discussed.

The paper is organized as follows. Our main results are stated and discussed in Section 2. The applications are given in Section 3. Proofs of theorems are deferred to Section 4.

2. Main results

The following theorem gives a counterpart of Haeusler's result [6] for self-normalized martingales.

Theorem 2.1. *Suppose that $\mathbf{E}|X_i|^{2p} < \infty$ for some $p > 1$ and all $i = 1, ..., n$. Then there exits a constant C_p depending only on p such that*

$$\sup_{x\in\mathbf{R}} \left|\mathbf{P}\Big(\frac{S_n}{\sqrt{[S]_n}} \leq x\Big) - \Phi\left(x\right)\right| \leq C_p\, N_n^{1/(2p+1)}, \tag{4}$$

where N_n is defined by (2). Moreover, there exit a sequence of martingale differences $(X_i, \mathcal{F}_i)_{i=0,...,n}$ and a positive constant c_p depending only on p such that

$$\sup_{x\in\mathbf{R}} \left|\mathbf{P}\Big(\frac{S_n}{\sqrt{[S]_n}} \leq x\Big) - \Phi\left(x\right)\right| N_n^{-1/(2p+1)} \geq c_p. \tag{5}$$

Clearly, inequality (5) shows that the bound (4) is asymptotically the best possible.

For a stationary martingale difference sequence, the term $\sum_{i=1}^{n} \mathbf{E}|X_i|^{2p}$ is of order n^{1-p}. Then inequality (4) implies the following corollary.

Corollary 2.1. *Let $(X_i, \mathcal{F}_i)_{i\geq 1}$ be a stationary martingale difference sequence. Suppose that $\mathbf{E}|X_1|^{2p} < \infty$ for some $p > 1$. Then there exits a constant c_p, which does not depend on n, such that*

$$\sup_{x\in\mathbf{R}} \left|\mathbf{P}\Big(\frac{S_n}{\sqrt{[S]_n}} \leq x\Big) - \Phi\left(x\right)\right| \leq c_p \left(n^{1-p} + \mathbf{E}|\langle S\rangle_n - 1|^p\right)^{1/(2p+1)}. \tag{6}$$

The next theorem gives a Berry-Esseen bound for normalized martingales $S_n/\sqrt{\langle S\rangle_n}$.

Theorem 2.2. *Under the assumptions of Theorem 2.1, the inequalities (4) and (5) hold when $S_n/\sqrt{[S]_n}$ is replaced by $S_n/\sqrt{\langle S\rangle_n}$.*

For a stationary martingale difference sequence, the following result is a consequence of the last theorem.

Corollary 2.2. *Assume the conditions of Corollary 2.1. Inequality (6) holds when $S_n/\sqrt{[S]_n}$ is replaced by $S_n/\sqrt{\langle S\rangle_n}$.*

3. Applications

3.1. Application to Student's t-statistic

The study of self-normalized partial sums originates from Student's t-statistic. The Student's t-statistic T_n is defined by

$$T_n = \sqrt{n}\,\overline{X}_n/\widehat{\sigma},$$

where

$$\overline{X}_n = \frac{S_n}{n} \quad \text{and} \quad \widehat{\sigma}^2 = \sum_{i=1}^{n} \frac{(X_i - \overline{X}_n)^2}{n-1}.$$

It is known that for all $x \geq 0$,

$$\mathbf{P}\Big(T_n > x\Big) = \mathbf{P}\bigg(\frac{S_n}{\sqrt{[S]_n}} > x\Big(\frac{n}{n+x^2-1}\Big)^{1/2}\bigg).$$

When $(X_i)_{i\geq 1}$ is a sequence of i.i.d. random variables, Bentkus and Götze [1] proved that if $\mathbb{E}|X_i|^{2+\delta} < \infty$ for all $i = 1, ..., n$ and some $\delta \in (0, 1]$, then

$$\sup_{x\in\mathbf{R}} \Big|\mathbf{P}(T_n \leq x) - \Phi(x)\Big| \quad = \quad O\Big(n^{-\delta/2}\Big). \tag{7}$$

For martingales, we have the following analogue.

Corollary 3.1. *Let $(X_i, \mathcal{F}_i)_{i\geq 1}$ be a stationary martingale difference sequence. Suppose that $\mathbf{E}|X_1|^{2p} < \infty$ for some $p > 1$. Then there exits a constant C_p, which does not depend on n, such that (6) holds when $\mathbf{P}(S_n/\sqrt{[S]_n} \leq x)$ is replaced by $\mathbf{P}(T_n \leq x)$.*

3.2. Application to autoregressive process

Consider the autoregressive process given by

$$Y_{n+1} = \theta Y_n + \varepsilon_{n+1}, \quad n \geq 0,$$

where Y_n and ε_n represent the observation and the driven noise, respectively. The parameter θ is unknown and needs to be estimated at stage n from the data $Y_i, i \leq n$. For sake of simplicity, we assume that $Y_0 = 0$. We also assume that $(\varepsilon_n)_{n\geq 0}$ is a stationary martingale difference sequence with $\mathbf{E}[\varepsilon_i^2|\varepsilon_1, ..., \varepsilon_{i-1}] = \sigma^2$ a.s. for a positive constant σ. We can estimate the unknown parameter θ by the least-squares estimator given by

$$\widehat{\theta}_n = \frac{\sum_{i=1}^{n} Y_i Y_{i+1}}{\sum_{i=1}^{n} Y_i^2}.$$

It is well known that $(\widehat{\theta}_n - \theta)\sqrt{\Sigma_{i=1}^{n} Y_i^2}$ converges in distribution to a normal law, see Theorem 3 of Lai and Wei [10]. By Theorem 2.2, we have the following Berry-Esseen bound for the least-squares estimator $\widehat{\theta}_n$.

Theorem 3.1. *Suppose that* $\mathbf{E}|\varepsilon_1|^{2p} < \infty$ *for some* $p > 1$. *If* $|\theta| < 1$, *then*

$$\sup_{x \in \mathbf{R}} \Big| \mathbf{P}\Big((\widehat{\theta}_n - \theta) \sqrt{\Sigma_{i=1}^n Y_i^2} \le x\sigma \Big) - \Phi(x) \Big|$$
$$= O\Bigg(n^{1-p} + n^{-p} \mathbf{E} \Big| \sum_{i=1}^n (Y_i^2 - \mathbf{E} Y_i^2) \Big|^p \Bigg)^{1/(2p+1)}, \qquad (8)$$

where $Y_n = \sum_{i=1}^n \theta^{n-i} \varepsilon_i$.

4. Proofs of theorems

4.1. Proof of Theorem 2.1

We assume that $N_n \le 1$. Otherwise, (4) is trivial.

Firstly, we give a lower bound for $\mathbf{P}(S_n \le x\sqrt{[S]_n}) - \Phi(x), x \le 0$. Let $\varepsilon_n \in (0, 1/2]$ be a positive number, whose exact value will be chosen later. It is easy to see that for $x \le 0$,

$$\begin{aligned} \mathbf{P}(S_n \le x\sqrt{[S]_n}) - \Phi(x) &\ge \mathbf{P}(S_n \le x\sqrt{[S]_n},\ [S]_n < 1 + \varepsilon_n) - \Phi(x) \\ &\ge \mathbf{P}(S_n \le x\sqrt{1+\varepsilon_n},\ [S]_n < 1 + \varepsilon_n) - \Phi(x) \\ &\ge \mathbf{P}(S_n \le x\sqrt{1+\varepsilon_n}) - \mathbf{P}([S]_n \ge 1 + \varepsilon_n) - \Phi(x) \\ &= I_1 + I_2 - I_3, \end{aligned} \qquad (9)$$

where

$$\begin{aligned} I_1 &= \mathbf{P}(S_n \le x\sqrt{1+\varepsilon_n}) - \Phi(x\sqrt{1+\varepsilon_n}), \\ I_2 &= \Phi(x\sqrt{1+\varepsilon_n}) - \Phi(x), \\ I_3 &= \mathbf{P}([S]_n \ge 1 + \varepsilon_n). \end{aligned}$$

Next, we estimate I_1, I_2 and I_3. By Haeusler's inequality [6] (see also (3) when $p \in (1, 2]$), we get the following estimation for I_1 :

$$I_1 \ge -C_{p,1} \Big(\sum_{i=1}^n \mathbf{E}|X_i|^{2p} + \mathbf{E}|\langle S \rangle_n - 1|^p \Big)^{1/(2p+1)}. \qquad (10)$$

By one-term Taylor's expansion, we have the following estimation for I_2 :

$$\begin{aligned} I_2 &\ge -c_1 e^{-x^2/2} |x| (\sqrt{1+\varepsilon_n} - 1) \\ &\ge -c_2\, \varepsilon_n. \end{aligned} \qquad (11)$$

For I_3, by Markov's inequality, it follows that

$$\begin{aligned}
I_3 &= \mathbf{P}([S]_n - \langle S\rangle_n + \langle S\rangle_n - 1 \geq \varepsilon_n) \\
&\leq \mathbf{P}\Big([S]_n - \langle S\rangle_n \geq \frac{\varepsilon_n}{2}\Big) + \mathbf{P}\Big(\langle S\rangle_n - 1 \geq \frac{\varepsilon_n}{2}\Big) \\
&\leq c_3 \varepsilon_n^{-p}\Big(\mathbf{E}|[S]_n - \langle S\rangle_n|^p + \mathbf{E}|\langle S\rangle_n - 1|^p\Big). \qquad (12)
\end{aligned}$$

We distinguish two cases to estimate I_3. Notice that $([S]_i - \langle S\rangle_i, \mathcal{F}_i)_{i=0,\dots,n}$ is also a martingale.

Case 1: If $p \in (1,2]$, by the inequality of von Bahr-Esseen [15], it follows that

$$\begin{aligned}
\mathbf{E}|[S]_n - \langle S\rangle_n|^p &\leq c_4 \sum_{i=1}^n \mathbf{E}|X_i^2 - \mathbf{E}[X_i^2|\mathcal{F}_{i-1}]|^p \\
&\leq 2c_4 \sum_{i=1}^n \mathbf{E}[|X_i|^{2p} + |\mathbf{E}[X_i^2|\mathcal{F}_{i-1}]|^p\,] \\
&\leq c_5 \sum_{i=1}^n \mathbf{E}|X_i|^{2p}. \qquad (13)
\end{aligned}$$

Returning to (12), we have

$$I_3 \leq c_6\ \varepsilon_n^{-p}\Big(\sum_{i=1}^n \mathbf{E}|X_i|^{2p} + \mathbf{E}|\langle S\rangle_n - 1|^p\Big). \qquad (14)$$

Case 2: If $p > 2$, by Rosenthal's inequality (cf. Theorem 2.12 of Hall and Heyde [7]), we have

$$\mathbf{E}|[S]_n - \langle S\rangle_n|^p \leq C_{p,2}\Big(\mathbf{E}\Big(\sum_{i=1}^n \mathbf{E}[X_i^4|\mathcal{F}_{i-1}]\Big)^{p/2} + \sum_{i=1}^n \mathbf{E}|X_i|^{2p}\Big). \qquad (15)$$

Noting that $X_i^4 = (X_i^2)^{(p-2)/(p-1)}(|X_i|^{2p})^{1/(p-1)}$ for $p > 2$, we have by Hölder's inequality

$$\mathbf{E}[X_i^4|\mathcal{F}_{i-1}] \leq \Big(\mathbf{E}[|X_i|^{2p}|\mathcal{F}_{i-1}]\Big)^{1/(p-1)}\Big(\mathbf{E}[X_i^2|\mathcal{F}_{i-1}]\Big)^{(p-2)/(p-1)},$$

and hence

$$\begin{aligned}
\sum_{i=1}^n \mathbf{E}[X_i^4|\mathcal{F}_{i-1}] &\leq \sum_{i=1}^n \Big(\mathbf{E}[|X_i|^{2p}|\mathcal{F}_{i-1}]\Big)^{1/(p-1)}\Big(\mathbf{E}[X_i^2|\mathcal{F}_{i-1}]\Big)^{(p-2)/(p-1)} \\
&\leq \Big(\sum_{i=1}^n \mathbf{E}[|X_i|^{2p}|\mathcal{F}_{i-1}]\Big)^{1/(p-1)}\Big(\langle S\rangle_n\Big)^{(p-2)/(p-1)}.
\end{aligned}$$

By the inequality

$$(a+b)^q \le 2^q(a^q+b^q), \qquad a,b\ge 0 \ \text{ and } \ q>0,$$

and the fact that $(p^2-2p)/(2p-2)\le p$, it follows that for $p>2$,

$$\begin{aligned}\Big(\sum_{i=1}^n \mathbf{E}[X_i^4|\mathcal{F}_{i-1}]\Big)^{p/2} &\le \Big(\sum_{i=1}^n \mathbf{E}[|X_i|^{2p}|\mathcal{F}_{i-1}]\Big)^{p/(2p-2)}\Big(\langle S\rangle_n\Big)^{(p^2-2p)/(2p-2)} \\ &\le 2^p\Big(\sum_{i=1}^n \mathbf{E}[|X_i|^{2p}|\mathcal{F}_{i-1}]\Big)^{p/(2p-2)}\Big(1+|\langle S\rangle_n-1|^{(p^2-2p)/(2p-2)}\Big) \\ &\le 2^p\Big(\sum_{i=1}^n \mathbf{E}[|X_i|^{2p}|\mathcal{F}_{i-1}]\Big)^{p/(2p-2)} \\ &\quad +2^p\Big(\sum_{i=1}^n \mathbf{E}[|X_i|^{2p}|\mathcal{F}_{i-1}]\Big)^{p/(2p-2)}|\langle S\rangle_n-1|^{p(p-2)/(2p-2)}.\end{aligned} \tag{16}$$

As to the second term on the r.h.s. of the last inequality, we use the inequality

$$x^a y^{1-a}\le x+y, \qquad x,y\ge 0 \text{ and } a\in[0,1],$$

and hence

$$\begin{aligned}\Big(\sum_{i=1}^n \mathbf{E}[X_i^4|\mathcal{F}_{i-1}]\Big)^{p/2} &\le 2^p\Big(\sum_{i=1}^n \mathbf{E}[|X_i|^{2p}|\mathcal{F}_{i-1}]\Big)^{p/(2p-2)} \\ &\quad +2^p\Big(\sum_{i=1}^n \mathbf{E}[|X_i|^{2p}|\mathcal{F}_{i-1}]+|\langle S\rangle_n-1|^p\Big).\end{aligned}$$

Thus,

$$\begin{aligned}&\mathbf{E}\Big(\sum_{i=1}^n \mathbf{E}[X_i^4|\mathcal{F}_{i-1}]\Big)^{p/2} \\ &\le 2^p\Big[\mathbf{E}\Big(\sum_{i=1}^n \mathbf{E}[|X_i|^{2p}|\mathcal{F}_{i-1}]\Big)^{p/(2p-2)}+\sum_{i=1}^n\mathbf{E}|X_i|^{2p}+\mathbf{E}|\langle S\rangle_n-1|^p\Big] \\ &\le 2^p\Big[\Big(\sum_{i=1}^n \mathbf{E}|X_i|^{2p}\Big)^{p/(2p-2)}+\sum_{i=1}^n\mathbf{E}|X_i|^{2p}+\mathbf{E}|\langle S\rangle_n-1|^p\Big].\end{aligned} \tag{17}$$

Returning to (15), we get for $p>2$,

$$\mathbf{E}|[S]_n-\langle S\rangle_n|^p\le C_{p,3}\bigg(\Big(\sum_{i=1}^n\mathbf{E}|X_i|^{2p}\Big)^{p/(2p-2)}+\mathbf{E}|\langle S\rangle_n-1|^p\bigg).$$

From (12) and the last inequality, we obtain for $p > 2$,

$$I_3 \leq C_{p,4}\ \varepsilon_n^{-p}\bigg(\Big(\sum_{i=1}^{n}\mathbf{E}|X_i|^{2p}\Big)^{p/(2p-2)} + \mathbf{E}|\langle S\rangle_n - 1|^p\bigg). \tag{18}$$

By the inequalities (14) and (18), we always have for $p > 1$,

$$I_3 \leq C_{p,5}\ \varepsilon_n^{-p}\bigg(\sum_{i=1}^{n}\mathbf{E}|X_i|^{2p} + \Big(\sum_{i=1}^{n}\mathbf{E}|X_i|^{2p}\Big)^{p/(2p-2)} + \mathbf{E}|\langle S\rangle_n - 1|^p\bigg). \tag{19}$$

Combining (9), (10), (11) and (19) together, we deduce that for $p > 1$,

$$\begin{aligned}
&\mathbf{P}(S_n \leq x\sqrt{[S]_n}) - \Phi(x)\\
&\geq -C_{p,1}\,\Big(\sum_{i=1}^{n}\mathbf{E}|X_i|^{2p} + \mathbf{E}|\langle S\rangle_n - 1|^p\Big)^{1/(2p+1)} - \ c_2\,\varepsilon_n\\
&\quad - C_{p,5}\,\varepsilon_n^{-p}\bigg(\sum_{i=1}^{n}\mathbf{E}|X_i|^{2p} + \Big(\sum_{i=1}^{n}\mathbf{E}|X_i|^{2p}\Big)^{p/(2p-2)} + \mathbf{E}|\langle S\rangle_n - 1|^p\bigg).
\end{aligned}$$

Taking

$$\varepsilon_n = \bigg(\sum_{i=1}^{n}\mathbf{E}|X_i|^{2p} + \Big(\sum_{i=1}^{n}\mathbf{E}|X_i|^{2p}\Big)^{p/(2p-2)} + \mathbf{E}|\langle S\rangle_n - 1|^p\bigg)^{1/(p+1)}, \tag{20}$$

we obtain for $x \leq 0$ and $p > 1$,

$$\begin{aligned}
&\mathbf{P}(S_n \leq x\sqrt{[S]_n}) - \Phi(x)\\
&\geq -C_{p,1}\Big(\sum_{i=1}^{n}\mathbf{E}|X_i|^{2p} + \mathbf{E}|\langle S\rangle_n - 1|^p\Big)^{1/(2p+1)}\\
&\quad - C_{p,6}\,\bigg(\sum_{i=1}^{n}\mathbf{E}|X_i|^{2p} + \Big(\sum_{i=1}^{n}\mathbf{E}|X_i|^{2p}\Big)^{p/(2p-2)} + \mathbf{E}|\langle S\rangle_n - 1|^p\bigg)^{1/(p+1)}\\
&\geq -C_{p,7}\Big(\sum_{i=1}^{n}\mathbf{E}|X_i|^{2p} + \mathbf{E}|\langle S\rangle_n - 1|^p\Big)^{1/(2p+1)},
\end{aligned} \tag{21}$$

where the last line follows from the fact that $p/((2p-2)(p+1)) \geq 1/(2p+1)$ and $N_n \leq 1$.

Secondly, we give an upper bound for $\mathbf{P}(S_n \leq x\sqrt{[S]_n}) - \Phi(x), x \leq 0$. It is obvious that for $x \leq 0$,

$$\mathbf{P}(S_n \leq x\sqrt{[S]_n}) - \Phi(x)$$

$$
\begin{aligned}
&\leq \mathbf{P}(S_n \leq x\sqrt{[S]_n},\ [S]_n > 1-\varepsilon_n) - \Phi(x) \\
&\quad + \mathbf{P}(S_n \leq x\sqrt{[S]_n},\ [S]_n \leq 1-\varepsilon_n) \\
&\leq \mathbf{P}(S_n \leq x\sqrt{1-\varepsilon_n},\ [S]_n > 1-\varepsilon_n) - \Phi(x) + \mathbf{P}([S]_n \leq 1-\varepsilon_n) \\
&\leq \mathbf{P}(S_n \leq x\sqrt{1-\varepsilon_n}) - \Phi(x\sqrt{1-\varepsilon_n}) + \Phi(x\sqrt{1-\varepsilon_n}) - \Phi(x) \\
&\quad + \mathbf{P}([S]_n \leq 1-\varepsilon_n) \\
&= I_4 + I_5 + I_6.
\end{aligned}
$$

Following the same lines as in the proof of (21), we get for $x \leq 0$ and $p > 1$,

$$
\mathbf{P}(S_n \leq x\sqrt{[S]_n}) - \Phi(x) \ \leq \ C_{p,8}\Big(\sum_{i=1}^n \mathbf{E}|X_i|^{2p} + \mathbf{E}|\langle S\rangle_n - 1|^p\Big)^{1/(2p+1)}. \tag{22}
$$

Combining (21) and (22) together, we get for $p > 1$,

$$
\sup_{x\leq 0}\Big|\mathbf{P}(S_n \leq x\sqrt{[S]_n}) - \Phi(x)\Big| \ \leq \ C_{p,8}\Big(\sum_{i=1}^n \mathbf{E}|X_i|^{2p} + \mathbf{E}|\langle S\rangle_n - 1|^p\Big)^{1/(2p+1)}. \tag{23}
$$

Notice that $(-S_k, \mathcal{F}_k)_{k=0,...,n}$ is also a martingale. Applying the last inequality to $(-S_k, \mathcal{F}_k)_{k=0,...,n}$, we get

$$
\begin{aligned}
&\sup_{x>0}\Big|\mathbf{P}(S_n \leq x\sqrt{[S]_n}) - \Phi(x)\Big| \\
&= \sup_{x>0}\Big|\mathbf{P}(S_n \leq x\sqrt{[S]_n}) - 1 + 1 - \Phi(x)\Big| \\
&= \sup_{x>0}\Big|\Phi(-x) - \mathbf{P}(-S_n < -x\sqrt{[S]_n})\Big| \\
&\leq C_{p,9}\Big(\sum_{i=1}^n \mathbf{E}|X_i|^{2p} + \mathbf{E}|\langle S\rangle_n - 1|^p\Big)^{1/(2p+1)}.
\end{aligned} \tag{24}
$$

Combining the inequalities (23) and (24) together, we obtain

$$
\sup_{x\in\mathbf{R}}\Big|\mathbf{P}(S_n \leq x\sqrt{[S]_n}) - \Phi(x)\Big| \leq C_{p,10}\Big(\sum_{i=1}^n \mathbf{E}|X_i|^{2p} + \mathbf{E}|\langle S\rangle_n - 1|^p\Big)^{1/(2p+1)}, \tag{25}
$$

which gives the desired inequality (4).

Next we give a proof of (5). We follow the example of Haeusler [6]. Let $(\alpha_n)_{n\geq 1}$ be a sequence of positive numbers such that $\alpha_n \to 0$ as $n \to \infty$. Define function $f_n : \mathbf{R} \to [0, \infty)$ as follows

$$
f_n(x) = \begin{cases} x^{-1}, & \text{if } \frac{1}{2}\sqrt{\alpha_n} \leq x < \infty, \\ 0, & \text{otherwise.} \end{cases}
$$

Furthermore, let $X_1, ..., X_{n-1}$ be independent and normally distributed random variables with mean 0 and variance $1/(n-1)$. Denote ν_x the one-point mass concentration at x. Define the random variable X_n such that its conditional distribution, given $X_1, ..., X_{n-1}$, is

$$\mathbf{P}(X_n \in \cdot | S_{n-1} = x) = \frac{1}{2}\nu_{-\alpha_n f_n(x)}(\cdot) + \frac{1}{2}\nu_{\alpha_n f_n(x)}(\cdot),$$

where $S_{n-1} = \sum_{i=1}^{n-1} X_i$. Denote $\mathcal{F}_i$ the natural filtration of $X_1, ..., X_n$, that is $\mathcal{F}_0$ being the trivial $\sigma-$field and $\mathcal{F}_i = \sigma\{X_1, ..., X_i\}, i = 1, ..., n$. Clearly, $(X_i, \mathcal{F}_i)_{i=0,...,n}$ is a finite sequence of martingale differences. Moreover, it holds

$$\sum_{i=1}^{n-1} \mathbf{E}|X_i|^{2p} = \frac{n-1}{(n-1)^p}\mathbf{E}|\mathcal{N}(0,1)|^{2p} \sim C_{p,11} n^{1-p}$$

and

$$\begin{aligned}
\mathbf{E}|X_n|^{2p} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |y|^{2p}\mathbf{P}(X_n \in dy|\mathcal{N}(0,1) = x)\mathbf{P}(\mathcal{N}(0,1) \in dx) \\
&= \frac{1}{\sqrt{2\pi}}\int_{\frac{1}{2}\sqrt{\alpha_n}}^{\infty} \Big|\frac{\alpha_n}{x}\Big|^{2p} e^{-\frac{1}{2}x^2} dx \\
&\sim C_{p,12}\, \alpha_n^{p+\frac{1}{2}} \qquad (26)
\end{aligned}$$

for some constants $0 < C_{p,11}, C_{p,12} < \infty$, where $\mathcal{N}(0,1)$ is a standard random variable. Similarly, we have

$$\sum_{i=1}^{n-1} \mathbf{E}[X_i^2|\mathcal{F}_{i-1}] = \sum_{i=1}^{n-1} \mathbf{E}X_i^2 = 1$$

and

$$\begin{aligned}
\mathbf{E}|\langle S\rangle_n - 1|^p &= \mathbf{E}|\mathbf{E}[X_n^2|\mathcal{F}_{n-1}]|^p \\
&= \int_{-\infty}^{\infty}\Big|\int_{-\infty}^{\infty} y^2\mathbf{P}(X_n \in dy|\mathcal{N}(0,1) = x)\Big|^p\mathbf{P}(\mathcal{N}(0,1) \in dx) \\
&= \frac{1}{\sqrt{2\pi}}\int_{\frac{1}{2}\sqrt{\alpha_n}}^{\infty} \Big|\frac{\alpha_n}{x}\Big|^{2p} e^{-\frac{1}{2}x^2} dx \\
&\sim C_{p,12}\, \alpha_n^{p+\frac{1}{2}}. \qquad (27)
\end{aligned}$$

Thus

$$N_n \sim C_p\, \alpha_n^{p+\frac{1}{2}},$$

where C_p is a positive constant depending only on p. On the other hand, we have

$$
\begin{aligned}
\mathbf{P}\Big(\frac{S_n}{\sqrt{[S]_n}} \leq 0\Big) &= \mathbf{P}\Big(X_n + S_{n-1} \leq 0\Big) \\
&= \int_{-\infty}^{\infty} \mathbf{P}(X_n \leq -x | \mathcal{N}(0,1) = x) \mathbf{P}(\mathcal{N}(0,1) \in dx) \\
&= \int_{-\infty}^{\frac{1}{2}\sqrt{\alpha_n}} \mathbf{P}(X_n \leq -x | \mathcal{N}(0,1) = x) \mathbf{P}(\mathcal{N}(0,1) \in dx) \\
&\quad + \int_{\frac{1}{2}\sqrt{\alpha_n}}^{\infty} \mathbf{P}(X_n \leq -x | \mathcal{N}(0,1) = x) \mathbf{P}(\mathcal{N}(0,1) \in dx) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{1}{2}x^2} dx + \frac{1}{\sqrt{2\pi}} \int_{\frac{1}{2}\sqrt{\alpha_n}}^{\sqrt{\alpha_n}} \frac{1}{2} e^{-\frac{1}{2}x^2} dx \\
&= \Phi(0) + \frac{1}{4\sqrt{2\pi}} \sqrt{\alpha_n} \Big(1 + o(1)\Big).
\end{aligned}
$$

Hence, we deduce that

$$
\begin{aligned}
& \sup_{x \in \mathbf{R}} \Big| \mathbf{P}\Big(\frac{S_n}{\sqrt{[S]_n}} \leq x\Big) - \Phi\,(x) \Big| \, N_n^{-1/(2p+1)} \\
\geq & \Big| \mathbf{P}\Big(\frac{S_n}{\sqrt{[S]_n}} \leq 0\Big) - \Phi\,(0) \Big| \, (C_p\, \alpha_n^{p+\frac{1}{2}})^{-1/(2p+1)} \Big(1 + o(1)\Big) \\
= & \frac{1}{4\sqrt{2\pi}} \sqrt{\alpha_n}\, (C_p\, \alpha_n^{p+\frac{1}{2}})^{-1/(2p+1)} \Big(1 + o(1)\Big) \\
\sim & \frac{1}{4\sqrt{2\pi}}\, (C_p\,)^{-1/(2p+1)}.
\end{aligned}
$$

This completes the proof of Theorem 2.1.

4.2. Proof of Theorem 2.2

First, we give a lower bound for $\mathbf{P}(S_n \leq x\sqrt{\langle S \rangle_n}) - \Phi(x), x \leq 0$. Let $\varepsilon_n \in (0, 1/2]$ be a positive number, whose exact value will be chosen later. It is easy to see that for $x \leq 0$,

$$
\begin{aligned}
\mathbf{P}(S_n \leq x\sqrt{\langle S \rangle_n}) - \Phi(x) &\geq \mathbf{P}(S_n \leq x\sqrt{\langle S \rangle_n},\ \langle S \rangle_n < 1 + \varepsilon_n) - \Phi(x) \\
&\geq \mathbf{P}(S_n \leq x\sqrt{1 + \varepsilon_n},\ \langle S \rangle_n < 1 + \varepsilon_n) - \Phi(x) \\
&\geq \mathbf{P}(S_n \leq x\sqrt{1 + \varepsilon_n}) - \mathbf{P}(\langle S \rangle_n \geq 1 + \varepsilon_n) - \Phi(x) \\
&= I_1 + I_2 - I_7,
\end{aligned}
$$

where

$$
I_7 = \mathbf{P}(\langle S \rangle_n \geq 1 + \varepsilon_n).
$$

For I_7, by Markov's inequality, it follows that

$$I_7 \leq \varepsilon_n^{-p} \mathbf{E} |\langle S \rangle_n - 1|^p. \tag{28}$$

Combining the estimations (10), (11) and (28) together, we have for $p > 1$,

$$\begin{aligned} &\mathbf{P}(S_n \leq x\sqrt{[S]_n}) - \Phi(x) \\ &\geq -C_{p,1}\Big(\sum_{i=1}^n \mathbf{E}|X_i|^{2p} + \mathbf{E}|\langle S \rangle_n - 1|^p \Big)^{1/(2p+1)} - \, c_2\, \varepsilon_n \, - \, \varepsilon_n^{-p} \mathbf{E}|\langle S \rangle_n - 1|^p. \end{aligned}$$

Next we carry out an argument as the proof of Theorem 2.1 with

$$\varepsilon_n = \Big(\mathbf{E}|\langle S \rangle_n - 1|^p \Big)^{1/(p+1)}, \tag{29}$$

what we obtain is

$$\sup_{x \in \mathbf{R}} \Big| \mathbf{P}\Big(\frac{S_n}{\sqrt{\langle S \rangle_n}} \leq x \Big) - \Phi\left(x\right) \Big| \leq C_{p,2}\Big(\sum_{i=1}^n \mathbf{E}|X_i|^{2p} + \mathbf{E}|\langle S \rangle_n - 1|^p \Big)^{1/(2p+1)}, \tag{30}$$

that is inequality (4) holds when $S_n/\sqrt{[S]_n}$ is replaced by $S_n/\sqrt{\langle S \rangle_n}$. The proof of optimality is similar to the proof of (5). This completes the proof of Theorem 2.2.

4.3. Proof of Theorem 3.1

The proof of theorem is based on Theorem 2.2. It is easy to see that

$$\frac{1}{\sigma}\,(\widehat{\theta}_n - \theta)\sqrt{\Sigma_{i=1}^n Y_i^2} = \frac{\sum_{i=1}^n Y_i \varepsilon_{i+1}}{\sigma \sqrt{\sum_{i=1}^n Y_i^2}}.$$

Notice that $Y_n = \sum_{i=1}^n \theta^{n-i} \varepsilon_i$. Set

$$X_i = \frac{Y_i \varepsilon_{i+1}}{\sigma \sqrt{\sum_{i=1}^n \mathbf{E} Y_i^2}} \qquad \text{and} \qquad \mathcal{F}_i = \sigma\{\varepsilon_k,\ 1 \leq k \leq i+1\}.$$

Then it is easy to see that $(X_i, \mathcal{F}_i)_{i=0,\ldots,n}$ is a sequence of martingale differences, and that

$$\frac{1}{\sigma}\,(\widehat{\theta}_n - \theta)\sqrt{\Sigma_{i=1}^n Y_i^2} = \frac{S_n}{\sqrt{\langle S \rangle_n}}.$$

Moreover, we have

$$\mathbf{E} Y_n^2 = \sum_{i=1}^n \theta^{2(n-i)} \sigma^2 = \frac{1 - \theta^{2\,n}}{1 - \theta^2} \sigma^2$$

and

$$\sum_{i=1}^{n}\mathbf{E}Y_i^2=\sum_{i=1}^{n}\frac{1-\theta^{2\,i}}{1-\theta^2}\sigma^2.$$

By Rosenthal's inequality, we also have

$$\begin{aligned}\mathbf{E}|Y_n|^{2p} &\leq C_p\Big((\mathbf{E}Y_n^2)^p+\sum_{i=1}^{n}\mathbf{E}|\theta^{n-i}\varepsilon_i|^{2p}\Big)\\ &\leq C_p\bigg(\Big(\frac{1-\theta^{2\,n}}{1-\theta^2}\Big)^p\sigma^{2p}+\frac{1-|\theta|^{2p\,n}}{1-|\theta|^{2p}}\mathbf{E}|\varepsilon_1|^{2p}\bigg)\end{aligned}$$

and

$$\sum_{i=1}^{n}\mathbf{E}|Y_i|^{2p}\leq C_p\,\sum_{i=1}^{n}\bigg(\Big(\frac{1-\theta^{2\,i}}{1-\theta^2}\Big)^p\sigma^{2p}+\frac{1-|\theta|^{2p\,i}}{1-|\theta|^{2p}}\mathbf{E}|\varepsilon_1|^{2p}\bigg).$$

Thus

$$\frac{\sum_{i=1}^{n}\mathbf{E}|X_i|^{2p}}{(\sum_{i=1}^{n}\mathbf{E}X_i^2)^p}\leq C_p\,\sum_{i=1}^{n}\bigg(\Big(\frac{1-\theta^{2\,i}}{1-\theta^2}\Big)^p\sigma^{2p}+\frac{1-|\theta|^{2p\,i}}{1-|\theta|^{2p}}\mathbf{E}|\varepsilon_1|^{2p}\bigg)\Big/\Big(\sum_{i=1}^{n}\frac{1-\theta^{2\,i}}{1-\theta^2}\sigma^2\Big)^p. \tag{31}$$

If $|\theta|<1$, inequality (31) implies that

$$\begin{aligned}\frac{\sum_{i=1}^{n}\mathbf{E}|X_i|^{2p}}{(\sum_{i=1}^{n}\mathbf{E}X_i^2)^p} &\leq C_p\,\sum_{i=1}^{n}\bigg(\frac{\sigma^{2p}}{(1-\theta^2)^p}+\frac{\mathbf{E}|\varepsilon_1|^{2p}}{1-|\theta|^{2p}}\bigg)\Big/\Big(\sum_{i=1}^{n}\sigma^2\Big)^p\\ &\leq C_p\bigg(\frac{1}{(1-\theta^2)^p}+\frac{1}{1-|\theta|^{2p}}\frac{\mathbf{E}|\varepsilon_1|^{2p}}{\sigma^{2p}}\bigg)n^{1-p}.\end{aligned}$$

It is obvious that

$$\langle S\rangle_n=\frac{\sum_{i=1}^{n}Y_i^2}{\sum_{i=1}^{n}\mathbf{E}Y_i^2}. \tag{32}$$

By Theorem 2.2, we obtain

$$\begin{aligned}&\sup_{x\in\mathbf{R}}\Big|\mathbf{P}\Big((\widehat{\theta}_n-\theta)\sqrt{\Sigma_{i=1}^{n}Y_i^2}\leq x\sigma\Big)-\Phi\left(x\right)\Big|\\ &\leq C_{p,\theta}\left(\frac{\sum_{i=1}^{n}\mathbf{E}|X_i|^{2p}}{(\sum_{i=1}^{n}\mathbf{E}X_i^2)^p}+\mathbf{E}\bigg|\frac{\sum_{i=1}^{n}Y_i^2}{\sum_{i=1}^{n}\mathbf{E}Y_i^2}-1\bigg|^p\right)^{1/(2p+1)}\\ &=O\bigg(n^{1-p}+n^{-p}\mathbf{E}\Big|\sum_{i=1}^{n}(Y_i^2-\mathbf{E}Y_i^2)\Big|^p\bigg)^{1/(2p+1)}.\end{aligned}$$

This completes the proof of theorem.

Acknowledgements

The research is partially supported by Hong Kong RGC GRF 14302515. Fan has been partially supported by the National Natural Science Foundation of China (Grant nos. 11601375 and 11626250).

Referees

[1] Bentkus, V., Götze, F. (1996). The Berry-Esseen bound for Student's statistic. Ann. Probab. 24(1): 491-503.

[2] Bentkus, V., Bloznelis, M., Götze, F. (1996). A Berry-Esséen bound for Student's statistic in the non-iid case. J. Theor. Probab. 9(3): 765-796.

[3] Csörgő, M., Szyszkowicz, B., Wang, Q. (2003). Donsker's theorem for self-normalized partial sums processes. Ann. Probab. 31 1228-1240.

[4] de la Peña, V.H., Lai, T.L., Shao, Q.M. (2009). Self-normalized processes: limit theory and statistical applications. Springer, New York.

[5] Giné, E., Götze, F., Mason, D.M. (1997). When is the Student t-statistic asymptotically standard normal? Ann. Probab. 25 1514-1531.

[6] Haeusler, E. (1988). On the rate of convergence in the central limit theorem for martingales with discrete and continuous time. Ann. Probab. 16(1), 275-299.

[7] Hall, P., Heyde, C.C. (1980). Martingale Limit Theory and its Applications. Academic, New York.

[8] Heyde, C.C., Brown, B.M. (1970). On the depature from normality of a certain class of martingales. Ann. Math. Statist. 41, 2161-2165.

[9] Jing, B.Y., Shao, Q.M., Wang, Q. (2003). Self-normalized Cramér-type large deviations for independent random variables. Ann. Probab. 31(4): 2167-2215.

[10] Lai, T.L., Wei, C.Z. (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. Ann. Statist. 10(1): 154-166.

[11] Logan, B.F., Mallows, C.L., Rice, S.O., Shepp, L.A. (1973). Limit distributions of self-normalized sums. Ann. Probab. 1: 788-809.

[12] Shao, Q.M. (1999). A Cramér type large deviation result for Student's $t-$statistic. J. Theor. Probab. 12(2): 385-398.

[13] Shao, Q.M., Wang, Q.M. (2013). Self-normalized limit theorems: A survey. Probab. Survey 10, 69-93.

[14] Shao, Q.M., Zhou, W.X. (2017). Self-normalization: Taming a wild population in a heavy-tailed world. Appl. Math. J. Chinese Univ. 32, 253-269.

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