Noname manuscript No.
(will be inserted by the editor)

Sharp large deviations for sums of bounded from above random variables

Xiequan Fan

Received: date / Accepted: date

Abstract We show large deviation expansions for sums of independent and bounded from above random variables. Our moderate deviation expansions are similar to those of Cramér (1938), Bahadur and Ranga Rao (1960), and Sakhanenko (1991). In particular, our results extend Talagrand's inequality from bounded random variables to random variables having finite $(2+\delta)$th moments, where $\delta \in (0, 1]$. As a consequence, we obtain an improvement of Hoeffding's inequality. Applications to linear regression, self-normalized large deviations and t-statistic are also discussed.

Keywords sharp large deviations · Cramér large deviations · Talagrand's inequality · Hoeffding's inequality · sums of independent random variables

Mathematics Subject Classification (2000) 60F10 · 60F05 · 60E15 · 60G50

1 Introduction

Let $(\xi_i)_{i\geq 1}$ be a sequence of independent non-degenerate random variables (r.v.s) satisfying $\mathbf{E}\xi_i = 0$. The study of sharp large deviations has a long history. Many interesting asymptotic expansions have been established in Cramér [8], Bahadur and Ranga Rao [1], Petrov [17,18], Saulis and Statulevičius [21], Sakhanenko [22], Nagaev [16], Bercu and Rouault [5], Borovkov and Mogulskii [6], Petrov and Robinson [19] and [12]. See also Grama and Haeusler [13] and [11] for martingales. For self-normalized sums, we refer to Shao [23] and Jing, Shao and Wang [15], where the authors have established Cramér type large deviations for self-normalized r.v.s under finite $(2 + \delta)$th moments, where $\delta \in (0, 1]$.

In this paper, we consider the sharp large deviations for sums of bounded from above r.v.s $\xi_i \leq A$ for all i, where A is a positive constant. Without loss of generality, we take $A = 1$, otherwise we consider ξ_i/A instead of ξ_i. Thus $\xi_i \leq 1$ for all i. Let $S_n = \sum_{i=1}^{n} \xi_i$. Denote

$$\sigma_i^2 = \mathbf{E}\xi_i^2 \qquad \text{and} \qquad \sigma^2 = \sum_{i=1}^{n} \sigma_i^2.$$

X. Fan (✉)
Center for Applied Mathematics, Tianjin University, 300072 Tianjin, China;
Regularity Team, INRIA Saclay, Palaiseau, 91120, France
E-mail: fanxiequan@hotmail.com

The celebrated Bennett inequality [2] states that: If $\xi_i \geq -1$ for all $1 \leq i \leq n$, then for all $x > 0$,

$$\mathbf{P}(S_n \geq x\sigma) \leq B(x,\sigma) := \left(\frac{x+\sigma}{\sigma}\right)^{-\sigma x - \sigma^2} e^{x\sigma}. \tag{1}$$

However, Bennett's inequality is not tight enough. One of the improvements on Bennett's inequality is Hoeffding's inequality (cf. (2.8) of [14]), which states that for all $x > 0$,

$$\begin{aligned} \mathbf{P}(S_n \geq x\sigma) \ &\leq \ H_n(x,\sigma) \qquad (2)\\ &:= \left\{ \left(\frac{\sigma}{x+\sigma}\right)^{x\sigma+\sigma^2} \left(\frac{n}{n-x\sigma}\right)^{n-x\sigma} \right\}^{\frac{n}{n+\sigma^2}} \mathbf{1}_{\{x \leq \frac{n}{\sigma}\}} \\ &\leq \ B(x,\sigma), \qquad (3) \end{aligned}$$

where (and hereafter) by convention $\infty^0 = 1$ applied when $x = n$. Considering the following distribution law

$$\mathbf{P}(\eta_i = 1) = \frac{\sigma^2/n}{1+\sigma^2/n} \quad \text{and} \quad \mathbf{P}(\eta_i = -\sigma^2/n) = \frac{1}{1+\sigma^2/n}, \tag{4}$$

Hoeffding showed that (2) is the best that can be obtained from the following exponential Markov inequality

$$\mathbf{P}(S_n \geq x\sigma) \leq \inf_{\lambda \geq 0} \mathbf{E} e^{\lambda(S_n - x\sigma)}, \qquad x \geq 0.$$

Indeed, it is easy to see that $\mathbf{E}\eta_i = 0, \eta_i \leq 1$, $\mathbf{E}\eta_i^2 = \sigma^2/n$ and

$$\inf_{\lambda \geq 0} \mathbf{E} \exp\left\{\lambda(\sum_{i=1}^n \eta_i - x\sigma)\right\} = H_n(x,\sigma)$$

for all $0 \leq x \leq \frac{n}{\sigma}$.

Notice that $\lim_{\sigma\to\infty} \mathbf{P}(S_n > x\sigma) = 1 - \Phi(x)$ and $\lim_{\sigma\to\infty} H_n(x,\sigma) = e^{-x^2/2}$, where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

is the standard normal distribution function. The central limit theorem (CLT) suggests that Hoeffding's inequality (2) can be substantially refined by adding a missing factor $\Theta(x)$ as $\sigma \to \infty$, where

$$\Theta(x) = \Big(1 - \Phi(x)\Big) \exp\left\{\frac{x^2}{2}\right\} = O\left(\frac{1}{x}\right), \qquad x \to \infty. \tag{5}$$

The factor $\Theta(x)$ satisfies

$$\frac{1}{\sqrt{2\pi}(1+x)} \leq \Theta(x) \leq \frac{1}{\sqrt{\pi}(1+x)}, \qquad x \geq 0, \tag{6}$$

and $\sqrt{2\pi}\,\Theta(x)$ is known as Mill's ratio.

For sums of bounded r.v.s $-B \leq \xi_i \leq 1$ for some constant $B \geq 1$ and all $1 \leq i \leq n$, Talagrand [24] proved the following inequalities: For all $0 \leq x \leq \frac{\sigma}{CB}$,

$$\begin{aligned} \mathbf{P}(S_n \geq x\sigma) &\leq \left(\Theta(x) + C\frac{B}{\sigma}\right) \inf_{\lambda \geq 0} \mathbf{E} e^{\lambda(S_n - x\sigma)} \qquad (7)\\ &\leq \left(\Theta(x) + C\frac{B}{\sigma}\right) H_n(x,\sigma), \qquad (8) \end{aligned}$$

where $C > 0$ is an absolute constant. See also Sakhanenko [22] for a result similar to (7) but $\Theta(x)$ is replaced by $\Theta(\alpha(x))$, where $\alpha(x) = \{2\sup_\lambda [x\lambda - \log \mathbf{E} e^{\lambda S_n}]\}^{1/2}$. By (6), Talagrand's inequality (8) improves Hoeffding's inequality (2) by adding a factor $\Theta(x)[1 + o(1)]$ of order $\frac{1}{1+x}$ in the range $0 \le x = o(\frac{\sigma}{B})$ as $\frac{B}{\sigma} \to 0$. In the i.i.d. case, this range reduces to $0 \le x = o(\sqrt{n}), n \to \infty$.

In this paper, we extend Talagrand's inequality (7) from bounded r.v.s to r.v.s having finite $(2+\delta)$th moments, $\delta \in (0, 1]$. In particular, we improve Talagrand's inequality to an *equality*, which will imply simple large deviation expansions. Moreover, an improvement of Hoeffding's inequality (2) under finite $(2+\delta)$th moments is also given.

The paper is organized as follows. In Section 2, we present the extension of Talagrand's inequality. In Section 3, we apply our result to linear regression. In Section 4, we give two lower bounds for self-normalized moderate deviations and t-statistic. In Section 5, we prepare some auxiliary results. In Sections 6 and 7, we prove the main theorems.

2 An extension of Talagrand's inequality

Throughout the paper, we make use of the following notations: $a \wedge b = \min\{a, b\}$, $a \vee b = \max\{a, b\}$, $a_+ = a \vee 0$, and θ stands for some values satisfying $|\theta| \le 1$. Moreover, we denote C and C_δ, probably supplied with some indices, a generic positive absolute constant and a generic positive constant depending only on δ, respectively.

Our main result is the following theorem, which extends Talagrand's inequality (7) from bounded r.v.s to r.v.s having finite $(2+\delta)$th moments, $\delta \in (0, 1]$.

Theorem 1 *Assume that*

$$\xi_i \le 1,$$

and that there exist two constants $B \ge 1$ and $\delta \in (0, 1]$ such that

$$\mathbf{E}|\xi_i|^{2+\delta} \le B^\delta \mathbf{E}\xi_i^2, \qquad i \ge 1. \tag{9}$$

Then for all $0 \le x \le \frac{\sigma}{C_\delta B}$,

$$\mathbf{P}(S_n \ge x\sigma) = \left(\Theta(x) + \theta C \Big(\frac{B}{\sigma}\Big)^\delta\right) \inf_{\lambda \ge 0} \mathbf{E} e^{\lambda(S_n - x\sigma)}. \tag{10}$$

In particular, in the i.i.d. case, it implies that for all $0 \le x = o(n^{\delta/2})$,

$$\frac{\mathbf{P}(S_n \ge x\sigma)}{\Theta(x) \inf_{\lambda \ge 0} \mathbf{E} e^{\lambda(S_n - x\sigma)}} = 1 + o(1) \tag{11}$$

as $n \to \infty$.

By inspecting the proof of Theorem 1, we can see that Theorem 1 holds true for $C_\delta = \frac{3}{2} 6^{1/\delta}$ and $C = 127.75$.

To show the tightness of equality (10), let $S_n' = \varepsilon_1 + ... + \varepsilon_n$ be the sums of independent Rademacher r.v.s, i.e. $\mathbf{P}(\varepsilon_i = \pm 1) = \frac{1}{2}$ for all i. We display the tail probabilities and the simulation of

$$R(x, n) = \frac{\mathbf{P}(S_n' \ge x\sqrt{n})}{\Theta(x) \inf_{\lambda \ge 0} \mathbf{E} e^{\lambda(S_n' - x\sigma)}} = \frac{\mathbf{P}(S_n' \ge x\sqrt{n})}{\Theta(x) H_n(x, \sqrt{n})}$$

in Figure 1, which shows that $R(x, n)$ is very close to 1 for large n's.

In the following corollary, we give an improvement on Hoeffding's inequality (2).

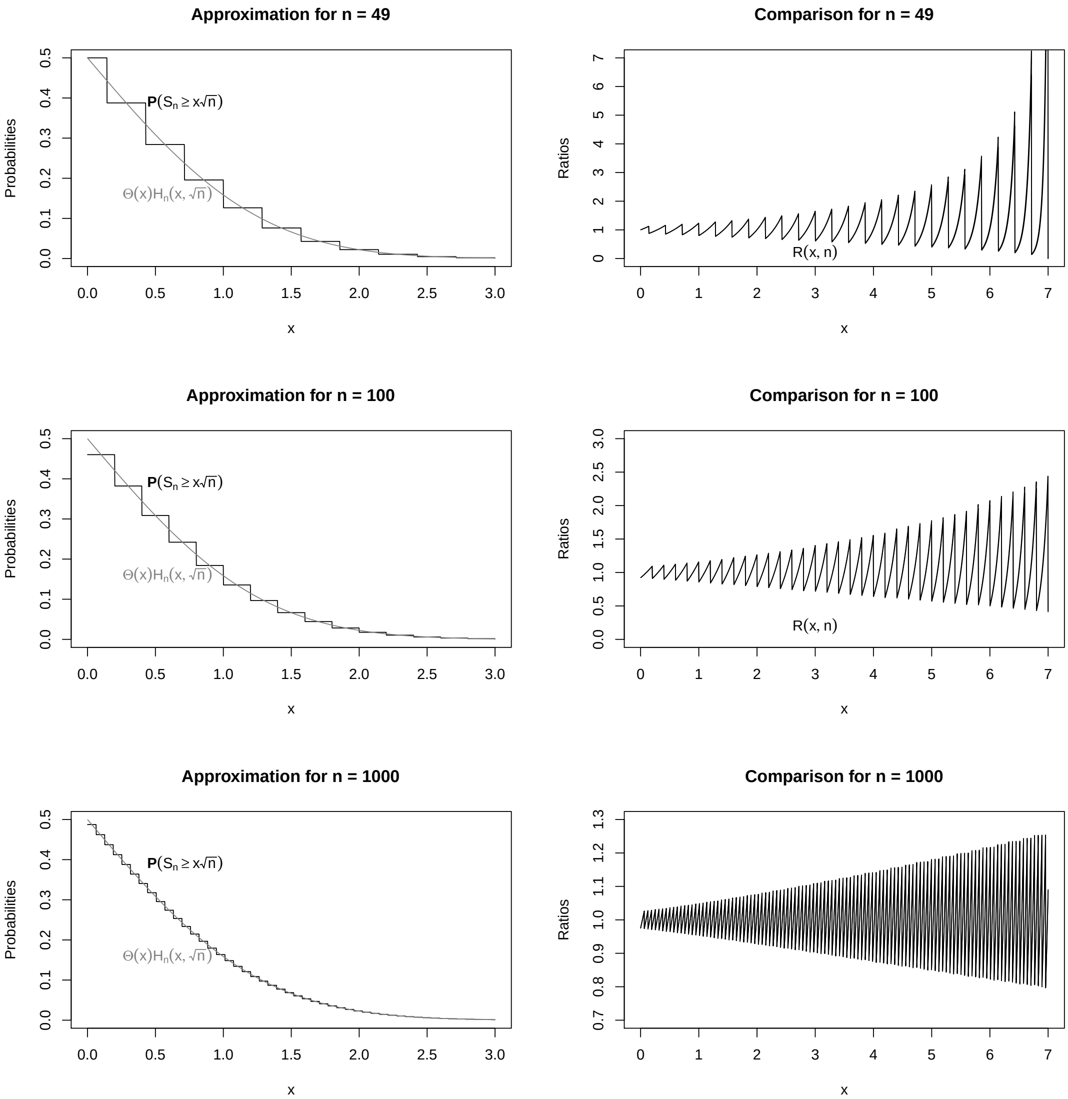

Fig. 1 Tail probabilities and ratios $R(x,n)$ are displayed as a function of x and various n.

Corollary 1 *Assume condition of Theorem 1. Then for all* $0 \leq x \leq \frac{\sigma}{C_\delta B}$,

$$\mathbf{P}(S_n \geq x\sigma) \leq \left(\left(\Theta(x) + C\Big(\frac{B}{\sigma}\Big)^{\delta} \right) \wedge 1 \right) H_n(x, \sigma). \tag{12}$$

It is clear that inequality (12) improves Hoeffding's bound $H_n(x,\sigma)$ by adding a missing factor $\big(\Theta(x) + C\big(\frac{B}{\sigma}\big)^{\delta}\big) \wedge 1$. By (6), this factor is of order of $\Theta(x)[1+o(1)]$ in the range $0 \leq x = o\big(\big(\sigma/B\big)^{\delta}\big)$ as $B/\sigma \to 0$. In the i.i.d. case, this range reduces to $0 \leq x = o(n^{\delta/2})$ as $n \to \infty$.

For r.v.s ξ_i without moments of order larger than 2, some improvements of Hoeffding's inequality (2) can be found in Bentkus [3] and Bentkus, Kalosha and van Zuijlen [4]. See also Pinelis [20] for an improvements of Bennett-Hoeffding's inequality (3) which is larger than

Hoeffding's inequality (2). In particular, when $\xi_i \leq 1$ for all $1 \leq i \leq n$, Bentkus [3] showed that

$$\mathbf{P}\left(S_n \geq x\right) \leq \frac{e^2}{2}\, \mathbf{P}^o\Big(\sum_{i=1}^n \eta_i \geq x\Big), \tag{13}$$

where η_i are i.i.d. with distribution (4) and $\mathbf{P}^o\left(\sum_{i=1}^n \eta_i \geq x\right)$ is the log-concave hull of $\mathbf{P}(\sum_{i=1}^n \eta_i \geq x)$, i.e. $\mathbf{P}^o$ is the minimum log-concave function such that $\mathbf{P}^o \geq \mathbf{P}$. Applying (10) to $\mathbf{P}\left(\sum_{i=1}^n \eta_i \geq x\right)$ with $B = \max\{1, \frac{\sigma^2}{n}\}$ and $\delta = 1$, we have for all $0 \leq x \leq \frac{1}{C}\min\{\frac{n}{\sigma}, \sigma\}$,

$$\mathbf{P}\Big(\sum_{i=1}^n \eta_i \geq x\sigma\Big) = \left(\Theta(x) + O(1)\max\left\{\frac{1}{\sigma}, \frac{\sigma}{n}\right\}\right) H_n(x,\sigma). \tag{14}$$

By the inequalities (13) and (14), we find that (12) refines Bentkus' constant $\frac{e^2}{2}$ (≈ 3.6945) to $1 + o(1)$ for all $0 \leq x = o(\big(\max\{\frac{1}{\sigma}, \frac{\sigma}{n}\}\big)^{\delta})$ as $\sigma \to 0$.

Inequality (12) implies the following Cramér type large deviations.

Corollary 2 *Assume condition of Theorem 1. Then for all* $0 \leq x \leq \frac{\sigma}{C_\delta B}$*,*

$$\mathbf{P}(S_n \geq x\sigma) \leq \Big(1 - \Phi\left(\check{x}\right)\Big)\left[1 + C\left(1 + \check{x}\right)\left(\frac{B}{\sigma}\right)^{\delta}\right], \tag{15}$$

where $\check{x} = \frac{x}{\sqrt{1+\frac{x}{3\sigma}}}$ *and satisfies*

$$\check{x} = x\left(1 - \frac{x}{6\sigma} + o(\frac{x}{\sigma})\right) \qquad as \;\; \frac{x}{\sigma} \to 0.$$

In particular, in the i.i.d. case, it implies that for all $0 \leq x = o(n^{\delta/2})$*,*

$$\mathbf{P}(S_n \geq x\sigma) \leq \Big(1 - \Phi(\check{x})\Big)\Big[1 + o(1)\Big]. \tag{16}$$

The interesting feature of the bound (16) is that it closely recovers the shape of the standard normal tail for all $0 \leq x = o(n^{\delta/2})$ as $n \to \infty$.

3 Application to linear regression

The linear regression model is given by

$$X_k = \theta\phi_k + \varepsilon_k, \quad k \geq 1, \tag{17}$$

where X_k, ϕ_k and ε_k are, respectively, the response variable, the positive covariate and the noise. Let $(\varepsilon_k)_{k\geq 1}$ be a sequence of i.i.d. random variables, with finite variance $\mathbf{E}\varepsilon_k^2 = \sigma_1^2 > 0$. Our interest is to estimate the unknown parameter θ, based on the random variables $(X_k)_{k\geq 1}$ and $(\phi_k)_{k\geq 1}$. The well-known least squares estimator θ_n is given by

$$\theta_n = \frac{\sum_{k=1}^n \phi_k X_k}{\sum_{k=1}^n \phi_k^2}. \tag{18}$$

Consider the self-normalized approximation $(\theta_n - \theta)\sqrt{\Sigma_{k=1}^n \phi_k^2}$. In the real-world applications, for instance considering the impact of the footprint size ϕ_k on the height X_k, it is plausible that $a \leq \phi_k \leq b$ for two positive absolute constants a and b. If $X_k \geq 0$, then we also have

$$\varepsilon_k = X_k - \theta\varphi_k \geq -\theta b \geq -c$$

for a positive absolute constant c. By Theorem 1 and Corollary 1, we have the following result.

Theorem 2 *Assume that there exist three positive absolute constants a, b and c such that for all $k \geq 1$,*

$$a \leq \phi_k \leq b, \qquad \varepsilon_k \geq -c.$$

We also assume that $\mathbf{E}|\varepsilon_k|^{2+\delta} < \infty$ for an absolute constant $\delta \in (0,1]$. Then for all $0 \leq x = o(\sqrt{n})$,

$$\mathbf{P}\Big((\theta - \theta_n)\sqrt{\Sigma_{k=1}^n \phi_k^2} \geq x\sigma_1\Big) = \Big(\Theta(x) + \theta C \frac{1}{n^{\delta/2}}\Big) \inf_{\lambda \leq 0} \mathbf{E}\exp\Big\{\lambda\Big(\sum_{i=1}^n \frac{\phi_k \varepsilon_k}{\sqrt{\sum_{k=1}^n \phi_k^2}} + x\sigma_1\Big)\Big\}$$
$$\leq \Big(\Theta(x) + C\frac{1}{n^{\delta/2}}\Big) H_n(x, \frac{\sqrt{n}a\sigma_1}{bc}).$$

Proof. From (17) and (18), it is easy to see that

$$(\theta - \theta_n)\sqrt{\Sigma_{k=1}^n \phi_k^2} = \sum_{k=1}^n \frac{-\phi_k \varepsilon_k}{\sqrt{\sum_{k=1}^n \phi_k^2}}. \tag{19}$$

Set

$$\xi_i = -\frac{\phi_k \varepsilon_k \sqrt{n} a}{bc\sqrt{\sum_{k=1}^n \phi_k^2}}.$$

Then it is easy to verify that

$$\xi_i \leq 1, \qquad \sum_{i=1}^n \mathbf{E}\xi_i^2 = \frac{na^2\sigma_1^2}{b^2c^2} \quad \text{and} \quad (\theta - \theta_n)\sqrt{\Sigma_{k=1}^n \phi_k^2}\frac{\sqrt{n}a}{bc} = \sum_{i=1}^n \xi_i.$$

By Theorem 1 and Corollary 1, it follows that for all $0 \leq x = o(\sqrt{n})$,

$$\mathbf{P}\Big((\theta - \theta_n)\sqrt{\Sigma_{k=1}^n \phi_k^2} \geq x\sigma_1\Big) = \Big(\Theta(x) + \theta C \frac{1}{n^{\delta/2}}\Big) \inf_{\lambda \geq 0} \mathbf{E}\exp\Big\{\lambda\Big(\sum_{i=1}^n \xi_i - x\frac{\sqrt{n}a\sigma_1}{bc}\Big)\Big\}$$
$$= \Big(\Theta(x) + \theta C \frac{1}{n^{\delta/2}}\Big) \inf_{\lambda \leq 0} \mathbf{E}\exp\Big\{\lambda\Big(\sum_{i=1}^n \frac{\phi_k \varepsilon_k}{\sqrt{\sum_{k=1}^n \phi_k^2}} + x\sigma_1\Big)\Big\}$$
$$\leq \Big(\Theta(x) + C\frac{1}{n^{\delta/2}}\Big) H_n(x, \frac{\sqrt{n}a\sigma_1}{bc}),$$

which completes the proof of theorem. □

4 Applications to self-normalized deviations and t-statistic

Limit theorems for self-normalized sums S_n/V_n, $V_n^2 = \sum_{i=1}^n \xi_i^2$, put a totally new countenance on classical limit theorems. It is well known that self-normalized limit theorems require much fewer moment conditions than that of normalized limit theorems. For example, under finite $(2+\delta)$th moments, $\delta \in (0,1]$, Jing, Shao and Wang [15] showed that

$$\frac{\mathbf{P}(S_n/V_n > x)}{1 - \Phi(x)} = \exp\Big\{O(1)(1+x)^{2+\delta}\varepsilon_n^\delta\Big\} \tag{20}$$

uniformly for all $0 \leq x = o(\min\{\varepsilon_n^{-1},\ \kappa_n^{-1}\})$, where

$$\varepsilon_n^\delta = \sum_{i=1}^n \mathbf{E}|\xi_i/\sigma|^{2+\delta} \quad \text{and} \quad \kappa_n^2 = \max_{1 \leq i \leq n} \mathbf{E}(\xi_i/\sigma)^2.$$

The proof of lower bound for self-normalized Cramér type large deviations is based on the following observation: For any $x, \lambda > 0$,

$$\Big\{S_n/V_n \geq x\Big\} \supseteq \Big\{S_n \geq \frac{x^2+\lambda^2 V_n^2}{2\lambda}\Big\} = \Big\{\sum_{i=1}^n \zeta_i(\lambda) \geq \frac{x^2}{2}\Big\},$$

where

$$\zeta_i(\lambda) = \lambda\xi_i - \frac{1}{2}\lambda^2\xi_i^2.$$

Thus $\mathbf{P}\big(S_n/V_n \geq x\big) \geq \mathbf{P}\big(\sum_{i=1}^n \zeta_i(\lambda) \geq \frac{x^2}{2}\big)$. Notice that $(\lambda\xi_i - \frac{1}{2}\lambda^2\xi_i^2)_{i\geq 1}$ is also a sequence of independent random variables and satisfies $\lambda\xi_i - \frac{1}{2}\lambda^2\xi_i^2 \leq 1$. By an argument similar to the proof of Theorem 1, we have the following lower bound on the tail probabilities of self-normalized sums.

Theorem 3 *Assume that there exist two constants $B>0$ and $\delta \in (0,1]$ such that*

$$\mathbf{E}|\xi_i|^{2+\delta} \leq B^\delta \mathbf{E}\xi_i^2, \qquad i \geq 1.$$

Then for all $0 \leq x = o(\frac{\sigma}{B})$,

$$\begin{aligned}\mathbf{P}\Big(S_n/V_n \geq x\Big) &\geq \mathbf{P}\Big(\sum_{i=1}^n \zeta_i(\lambda) \geq \frac{x^2}{2}\Big)\\ &= \exp\Bigg\{-\frac{x^2}{2} + \Psi_n(\lambda)\Bigg\}\Bigg(\Theta(x) + \theta C\Big(\frac{B}{\sigma}\Big)^\delta\Bigg),\end{aligned} \tag{21}$$

where

$$\Psi_n(\lambda) = \sum_{i=1}^n \log \mathbf{E}\, e^{\zeta_i(\lambda)} \tag{22}$$

and $\lambda \geq 0$ is defined by the following equation

$$\sum_{i=1}^n \frac{\mathbf{E}\,\zeta_i(\lambda) e^{\zeta_i(\lambda)}}{\mathbf{E}\, e^{\zeta_i(\lambda)}} = \frac{x^2}{2}.$$

Moreover, it holds for all $0 \leq x = o(\frac{\sigma}{B})$,

$$\Psi_n(\lambda) = O(1) x^{2+\delta}\Big(\frac{B}{\sigma}\Big)^\delta, \tag{23}$$

where $O(1)$ is bounded by an absolute constant.

The self-normalized sums are closely related to Student's t-statistic. Student's t-statistic T_n is defined by the following formula

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

where

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

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

$$\mathbf{P}\Big(T_n \geq x\Big) = \mathbf{P}\Big(S_n/V_n \geq x\Big(\frac{n}{n+x^2-1}\Big)^{1/2}\Big).$$

See Efron [9]. By the last equality, once we have an estimation for the tail probabilities of self-normalized sums S_n/V_n, we have a similar estimation for the tail probabilities of T_n. So Theorem 3 implies the following lower bound on tail probabilities of T_n.

Theorem 4 *Assume the condition of Theorem 3. Then for all* $0 \leq x = o(\frac{\sigma}{B})$,

$$\mathbf{P}\Big(T_n \geq x\Big) \geq \exp\left\{-\frac{x^2}{2}\Big(\frac{n}{n+x^2-1}\Big) + \Psi_n(\lambda)\right\}\left(\Theta\Big(x(\frac{n}{n+x^2-1})^{1/2}\Big) + \theta C\Big(\frac{B}{\sigma}\Big)^{\delta}\right),$$

where $\Psi_n(\lambda)$ *is defined by (22) and* $\lambda \geq 0$ *is defined by the following equation*

$$\sum_{i=1}^{n} \frac{\mathbf{E}\,\zeta_i(\lambda)e^{\zeta_i(\lambda)}}{\mathbf{E}\,e^{\zeta_i(\lambda)}} = \frac{x^2}{2}\Big(\frac{n}{n+x^2-1}\Big).$$

5 Auxiliary Results

Consider the positive random variable

$$Z_n(\lambda) = \prod_{i=1}^{n} \frac{e^{\lambda \xi_i}}{\mathbf{E}e^{\lambda \xi_i}}, \qquad \lambda \geq 0,$$

so that $\mathbf{E}Z_n(\lambda) = 1$ (the Esscher transformation). Introduce the conjugate probability measure $\mathbf{P}_\lambda$ defined by

$$d\mathbf{P}_\lambda = Z_n(\lambda)d\mathbf{P}. \tag{24}$$

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

$$b_i(\lambda) = \mathbf{E}_\lambda \xi_i = \frac{\mathbf{E}\xi_i e^{\lambda \xi_i}}{\mathbf{E}e^{\lambda \xi_i}}, \quad i = 1, ..., n,$$

and

$$\eta_i(\lambda) = \xi_i - b_i(\lambda), \quad i = 1, ..., n,$$

we obtain the following decomposition:

$$S_k = B_k(\lambda) + Y_k(\lambda), \quad k = 1, ..., n, \tag{25}$$

where

$$B_k(\lambda) = \sum_{i=1}^{k} b_i(\lambda) \qquad \text{and} \qquad Y_k(\lambda) = \sum_{i=1}^{k} \eta_i(\lambda).$$

In the proof of our main result, we shall need a two-sided bound of $B_n(\lambda)$. To this end, we need some technical lemmas.

For a random variable bounded from above, the following inequality is well-known.

Lemma 1 *Assume* $\xi \leq 1$. *Denote by* σ *the standard variance of* ξ. *Then for all* $\lambda \geq 0$,

$$\mathbf{E}e^{\lambda \xi} \leq Be(\lambda, \sigma^2),$$

where

$$Be(\lambda, t) = \frac{t}{1+t}\exp\{\lambda\} + \frac{1}{1+t}\exp\{-\lambda t\}.$$

A proof of the inequality can be found in Bennett [2]. This inequality is sharp, and it attains to equality when ξ has the distribution law:

$$\mathbf{P}(\xi = 1) = \frac{\sigma^2}{1+\sigma^2} \qquad \text{and} \qquad \mathbf{P}(\xi = -\sigma^2) = \frac{1}{1+\sigma^2}.$$

Lemma 2 *Assume* $\mathbf{E}|\xi|^{2+\delta} \leq B^\delta \mathbf{E}\xi^2$ *for some constants* B *and* $\delta \in (0,1]$. *Then*

$$\mathbf{E}\xi^2 \leq B^2. \tag{26}$$

Proof. Using Jensen's inequality, we deduce that

$$(\mathbf{E}\xi^2)^{(2+\delta)/2} \leq \mathbf{E}|\xi|^{2+\delta} \leq B^\delta \mathbf{E}\xi^2,$$

which implies (26). □

Lemma 3 *Assume* $\mathbf{E}|\xi|^{2+\delta} \leq B^\delta \mathbf{E}\xi^2$ *for some constants* B *and* $\delta \in (0,1]$. *Then for all* $\lambda \geq 0$,

$$\mathbf{E}\xi^2 e^{\lambda\xi} \geq (1 - B^\delta \lambda^\delta)\mathbf{E}\xi^2. \tag{27}$$

Proof. Using the following inequality $t^2 e^t \geq t^2(1 - |t|^\delta), t \in \mathbf{R}$, we deduce that for all $\lambda \geq 0$,

$$\mathbf{E}(\lambda\xi)^2 e^{\lambda\xi} \geq \mathbf{E}(\lambda\xi)^2 - \mathbf{E}|\lambda\xi|^{2+\delta} \geq \lambda^2(\mathbf{E}\xi^2 - (\lambda B)^\delta \mathbf{E}\xi^2), \tag{28}$$

which gives the desired inequality. □

In the following lemma, we give a two-sided bound for $B_n(\lambda)$.

Lemma 4 *Assume* $\xi_i \leq 1$ *for all* i. *Then for all* $\lambda \geq 0$,

$$B_n(\lambda) \leq (e^\lambda - 1)\sigma^2.$$

If ξ_i *satisfies* $\mathbf{E}|\xi_i|^{2+\delta} \leq B^\delta \mathbf{E}\xi_i^2$ *for some constants* B *and* $\delta \in (0,1]$ *for all* i, *then for all* $\lambda \geq 0$,

$$B_n(\lambda) \geq \left(1 - \frac{B^\delta \lambda^\delta}{1+\delta}\right) \lambda \sigma^2 e^{-\lambda}.$$

Proof. By Jensen's inequality, we have for all $\lambda \geq 0$, $\mathbf{E}e^{\lambda\xi_i} \geq e^{\lambda \mathbf{E}\xi_i} = 1$. Notice that $\mathbf{E}\xi_i e^{\lambda\xi_i} = \mathbf{E}\xi_i(e^{\lambda\xi_i} - 1) \geq 0$ for $\lambda \geq 0$. Then by the fact $\xi_i \leq 1$, we obtain the upper bound of $B_n(\lambda)$: For all $\lambda \geq 0$,

$$\begin{aligned}
B_n(\lambda) &\leq \sum_{i=1}^n \mathbf{E}\xi_i e^{\lambda\xi_i} = \sum_{i=1}^n \int_0^\lambda \mathbf{E}\xi_i^2 e^{t\xi_i}\, dt \\
&\leq \sum_{i=1}^n \int_0^\lambda \sigma_i^2\, e^t dt \\
&= (e^\lambda - 1)\sigma^2.
\end{aligned}$$

If ξ_i satisfies $\mathbf{E}|\xi_i|^{2+\delta} \leq B^\delta \mathbf{E}\xi_i^2$, by Lemma 3, it follows that for all $\lambda \geq 0$,

$$\begin{aligned}
\sum_{i=1}^n \mathbf{E}\xi_i e^{\lambda\xi_i} &= \int_0^\lambda \sum_{i=1}^n \mathbf{E}\xi_i^2 e^{t\xi_i} dt \\
&\geq \int_0^\lambda (1 - B^\delta t^\delta) dt \sum_{i=1}^n \mathbf{E}\xi_i^2 \\
&= \left(1 - \frac{B^\delta \lambda^\delta}{1+\delta}\right) \lambda\sigma^2.
\end{aligned}$$

Therefore, we get the lower bound of $B_n(\lambda)$: For all $\lambda \geq 0$,

$$B_n(\lambda) = \sum_{i=1}^n \frac{\mathbf{E}\xi_i e^{\lambda\xi_i}}{\mathbf{E}e^{\lambda\xi_i}} \geq \left(1 - \frac{B^\delta \lambda^\delta}{1+\delta}\right) \lambda\sigma^2 e^{-\lambda},$$

which completes the proof of Lemma 4. □

Next, we give an upper bound for the cumulant function

$$\Psi_n(\lambda) = \sum_{i=1}^{n} \log \mathbf{E} e^{\lambda \xi_i}, \qquad \lambda \geq 0. \tag{29}$$

Lemma 5 *Assume $\xi_i \leq 1$ for all i. Then for all $\lambda \geq 0$,*

$$\Psi_n(\lambda) \leq n \log \left(\frac{1}{1+\sigma^2/n} \exp\left\{-\lambda \sigma^2/n\right\} + \frac{\sigma^2/n}{1+\sigma^2/n} \exp\{\lambda\} \right).$$

Proof. Since the function

$$f(\lambda, t) = \log \left(\frac{1}{1+t} \exp\left\{-\lambda t\right\} + \frac{t}{1+t} \exp\{\lambda\} \right), \qquad \lambda, t \geq 0,$$

has a negative second derivative in $t > 0$, then for any fixed $\lambda \geq 0$, $-f(\lambda, t)$ is convex in $t \geq 0$. Therefore, by Lemma 1 and Jensen's inequality, we get for all $\lambda \geq 0$,

$$\begin{aligned}
\Psi_n(\lambda) &\leq \sum_{i=1}^{n} f(\lambda, \sigma_i^2) \\
&\leq n f(\lambda, \sigma^2/n) \\
&= n \log \left(\frac{1}{1+\sigma^2/n} \exp\left\{-\lambda \sigma^2/n\right\} + \frac{\sigma^2/n}{1+\sigma^2/n} \exp\{\lambda\} \right).
\end{aligned}$$

This completes the proof of Lemma 5. □

Denote the variance of $Y_n(\lambda)$ by $\overline{\sigma}^2(\lambda) = \mathbf{E}_\lambda Y_n^2(\lambda),\ \lambda \geq 0$. By the relation between $\mathbf{E}$ and $\mathbf{E}_\lambda$, it is obvious that

$$\overline{\sigma}^2(\lambda) = \sum_{i=1}^{n} \left(\frac{\mathbf{E}\xi_i^2 e^{\lambda \xi_i}}{\mathbf{E} e^{\lambda \xi_i}} - \frac{(\mathbf{E}\xi_i e^{\lambda \xi_i})^2}{(\mathbf{E} e^{\lambda \xi_i})^2} \right), \qquad \lambda \geq 0.$$

The following lemma gives some estimations of $\overline{\sigma}^2(\lambda)$.

Lemma 6 *Assume $\xi_i \leq 1$, and that $\mathbf{E}|\xi_i|^{2+\delta} \leq B^\delta \mathbf{E}\xi_i^2$ for some constants B, $\delta \in (0, 1]$ and all i. Then for all $\lambda \geq 0$,*

$$e^{-2\lambda} \left(1 - B^\delta \lambda^\delta - B^2 (e^\lambda - 1)^2\right) \sigma^2 \leq \overline{\sigma}_n^2(\lambda) \leq e^\lambda \sigma^2. \tag{30}$$

Moreover, if $B \geq 1$, it holds

$$\overline{\sigma}^2(\lambda) \geq (1 - 5B^\delta \lambda^\delta)_+ \sigma^2. \tag{31}$$

Proof. Since $\mathbf{E} e^{\lambda \xi_i} \geq 1, \lambda \geq 0$, and $\xi_i \leq 1$, we get for all $\lambda \geq 0$,

$$\overline{\sigma}^2(\lambda) \leq \sum_{i=1}^{n} \mathbf{E}\xi_i^2 e^{\lambda \xi_i} \leq \sum_{i=1}^{n} \mathbf{E}\xi_i^2 e^{\lambda} = e^\lambda \sigma^2.$$

This gives the upper bound of $\overline{\sigma}^2(\lambda)$. Next, we consider the lower bound of $\overline{\sigma}^2(\lambda)$. It is easy to see that for all $\lambda \geq 0$,

$$\mathbf{E}\xi_i e^{\lambda \xi_i} = \int_0^\lambda \mathbf{E}\xi_i^2 e^{t \xi_i} dt \leq \int_0^\lambda e^t \mathbf{E}\xi_i^2 dt = \left(e^\lambda - 1\right) \mathbf{E}\xi_i^2, \tag{32}$$

Using Lemma 3 and (32), we obtain for all $\lambda \geq 0$,

$$\begin{aligned}\overline{\sigma}^2(\lambda) &\geq \sum_{i=1}^n \frac{\mathbf{E}\xi_i^2 e^{\lambda\xi_i} - (\mathbf{E}\xi_i e^{\lambda\xi_i})^2}{e^{2\lambda}} \\ &\geq \frac{(1-B^\delta\lambda^\delta)\sigma^2 - (e^\lambda-1)^2\sum_{i=1}^n(\mathbf{E}\xi_i^2)^2}{e^{2\lambda}} \\ &\geq e^{-2\lambda}\left(1-B^\delta\lambda^\delta - B^2(e^\lambda-1)^2\right)\sigma^2,\end{aligned}$$

which gives the first lower bound of $\overline{\sigma}^2(\lambda)$. Noting that $\overline{\sigma}^2(\lambda) \geq 0$ and $B \geq 1$, by a simple calculation, we obtain (31). □

For the random variable $Y_n(\lambda)$, $\lambda \geq 0$, we have the following result on the rate of convergence to the standard normal law.

Lemma 7 *Assume that $\xi_i \leq 1$, and that $\mathbf{E}|\xi_i|^{2+\delta} \leq B^\delta \mathbf{E}\xi_i^2$ for some constants B, $\delta \in (0,1]$ and all i. Then for all $\lambda \geq 0$,*

$$\sup_{y\in\mathbf{R}}\left|\mathbf{P}_\lambda\left(\frac{Y_n(\lambda)}{\overline{\sigma}(\lambda)} \leq y\right) - \Phi(y)\right| \leq \frac{2^{2+\delta}\widetilde{C}_\delta e^\lambda}{\overline{\sigma}_n^{2+\delta}(\lambda)}\sum_{i=1}^n \mathbf{E}|\xi_i|^{2+\delta}.$$

Proof. Notice that $Y_n(\lambda) = \sum_{i=1}^n \eta_i(\lambda)$ is the sum of independent r.v.s $\eta_i(\lambda)$ and $\mathbf{E}_\lambda\eta_i(\lambda) = 0$. Using the well-known rate of convergence in the central limit theorem (cf. e.g. [17], p. 115), we get for all $\lambda \geq 0$,

$$\sup_{y\in\mathbf{R}}\left|\mathbf{P}_\lambda\left(\frac{Y_n(\lambda)}{\overline{\sigma}(\lambda)} \leq y\right) - \Phi(y)\right| \leq \frac{\widetilde{C}_\delta}{\overline{\sigma}_n^{2+\delta}(\lambda)}\sum_{i=1}^n \mathbf{E}_\lambda|\eta_i|^{2+\delta}. \tag{33}$$

Using the inequality $(a+b)^{1+q} \leq 2^q(a^{1+q}+b^{1+q})$ for $a,b,q \geq 0$, we deduce that for all $\lambda \geq 0$,

$$\begin{aligned}\sum_{i=1}^n \mathbf{E}_\lambda|\eta_i|^{2+\delta} &\leq 2^{1+\delta}\sum_{i=1}^n \mathbf{E}_\lambda(|\xi_i|^{2+\delta} + |\mathbf{E}_\lambda\xi_i|^{2+\delta}) \\ &\leq 2^{2+\delta}\sum_{i=1}^n \mathbf{E}_\lambda|\xi_i|^{2+\delta} \ \leq\ 2^{2+\delta}\sum_{i=1}^n \mathbf{E}|\xi_i|^{2+\delta}e^{\lambda\xi_i} \\ &\leq 2^{2+\delta}e^\lambda\sum_{i=1}^n \mathbf{E}|\xi_i|^{2+\delta}.\end{aligned}$$

Therefore, we obtain for all $\lambda \geq 0$,

$$\sup_{y\in\mathbf{R}}\left|\mathbf{P}_\lambda\left(\frac{Y_n(\lambda)}{\overline{\sigma}(\lambda)} \leq y\right) - \Phi(y)\right| \leq \frac{2^{2+\delta}\widetilde{C}_\delta e^\lambda}{\overline{\sigma}^{2+\delta}(\lambda)}\sum_{i=1}^n \mathbf{E}|\xi_i|^{2+\delta}.$$

This completes the proof of Lemma 7. □

We are now ready to prove the main technical result of this section.

Theorem 5 *Assume that $\xi_i \leq 1$, and that $\mathbf{E}|\xi_i|^{2+\delta} \leq B^\delta\mathbf{E}\xi_i^2$ for some constants B and $\delta \in (0,1]$ for all i. For an $x \geq 0$, if there exists a positive $\overline{\lambda}$ such that $\Psi_n'(\overline{\lambda}) = x\sigma$, then*

$$\mathbf{P}(S_n \geq x\sigma) = \left(\Theta\left(\overline{\lambda}\overline{\sigma}(\overline{\lambda})\right) + \theta\varepsilon_x\right)\inf_{\lambda\geq 0}\mathbf{E}e^{\lambda(S_n - x\sigma)}, \tag{34}$$

where

$$\varepsilon_x = \frac{2^{3+\delta}\widetilde{C}_\delta e^{\overline{\lambda}}}{\overline{\sigma}^{2+\delta}(\overline{\lambda})}\sum_{i=1}^n \mathbf{E}|\xi_i|^{2+\delta}.$$

Proof. According to the definition of the conjugate probability measure (cf. (24)), we have the following representation of $\mathbf{P}(S_n \geq x\sigma)$: For given $x, \lambda \geq 0$,

$$\begin{aligned}\mathbf{P}(S_n \geq x\sigma) &= \mathbf{E}_\lambda(Z_n(\lambda)^{-1}\mathbf{1}_{\{S_n \geq x\sigma\}}) \\ &= \mathbf{E}_\lambda(e^{-\lambda S_n + \Psi_n(\lambda)}\mathbf{1}_{\{S_n \geq x\sigma\}}) \\ &= \mathbf{E}_\lambda(e^{-\lambda x\sigma + \Psi_n(\lambda) - \lambda Y_n(\lambda) - \lambda B_n(\lambda) + \lambda x\sigma}\mathbf{1}_{\{Y_n(\lambda)+B_n(\lambda)-x\sigma \geq 0\}}) \\ &= e^{-\lambda x\sigma + \Psi_n(\lambda)}\mathbf{E}_\lambda(e^{-\lambda[Y_n(\lambda)+B_n(\lambda)-x\sigma]}\mathbf{1}_{\{Y_n(\lambda)+B_n(\lambda)-x\sigma \geq 0\}}).\end{aligned}$$

Setting $U_n(\lambda) = \lambda(Y_n(\lambda) + B_n(\lambda) - x\sigma)$, we get

$$\mathbf{P}(S_n \geq x\sigma) = e^{-\lambda x\sigma + \Psi_n(\lambda)} \int_0^\infty e^{-t}\mathbf{P}_\lambda(0 < U_n(\lambda) \leq t)dt. \tag{35}$$

For an $x \geq 0$, if there exists a $\overline{\lambda} = \overline{\lambda}(x)$ such that $\Psi_n'(\overline{\lambda}) = x\sigma$, then the exponential function $e^{-\lambda x\sigma + \Psi_n(\lambda)}$ in (35) attains its minimum at $\lambda = \overline{\lambda}$. Since $B_n(\overline{\lambda}) = \Psi_n'(\overline{\lambda}) = x\sigma$, we have $U_n(\overline{\lambda}) = \overline{\lambda}Y_n(\overline{\lambda})$ and

$$e^{-\overline{\lambda}x\sigma + \Psi_n(\overline{\lambda})} = \inf_{\lambda \geq 0} e^{-\lambda x\sigma + \Psi_n(\lambda)} = \inf_{\lambda \geq 0} \mathbf{E}e^{\lambda(S_n - x\sigma)}. \tag{36}$$

Using Lemma 7, we deduce that

$$\begin{aligned}\int_0^\infty e^{-t}\mathbf{P}_{\overline{\lambda}}(0 < U_n(\overline{\lambda}) \leq t)dt &= \int_0^\infty e^{-\overline{\lambda}y\overline{\sigma}(\overline{\lambda})}\mathbf{P}_{\overline{\lambda}}\left(0 < U_n(\overline{\lambda}) \leq \overline{\lambda}y\overline{\sigma}(\overline{\lambda})\right)\overline{\lambda}\overline{\sigma}(\overline{\lambda})dy \\ &= \int_0^\infty e^{-\overline{\lambda}y\overline{\sigma}(\overline{\lambda})}\mathbf{P}(0 < \mathcal{N}(0,1) \leq y)\overline{\lambda}\overline{\sigma}(\overline{\lambda})dy + \theta\varepsilon_x \\ &= \int_0^\infty e^{-\overline{\lambda}y\overline{\sigma}(\overline{\lambda})}d\Phi(y) + \theta\varepsilon_x \\ &= \Theta\left(\overline{\lambda}\overline{\sigma}(\overline{\lambda})\right) + \theta\varepsilon_x,\end{aligned} \tag{37}$$

where $\mathcal{N}(0,1)$ stands for the standard normal r.v. Therefore, from (35) and (36), it follows that

$$\mathbf{P}(S_n \geq x\sigma) = \left(\Theta\left(\overline{\lambda}\overline{\sigma}(\overline{\lambda})\right) + \theta\varepsilon_x\right) \inf_{\lambda \geq 0} \mathbf{E}e^{\lambda(S_n - x\sigma)}.$$

This completes the proof of Theorem 5. □

6 Proof of Theorem 1

In the spirit of Talagrand [24], we would like to make use of $\Theta(x)$ to approximate $\Theta(\overline{\lambda}\overline{\sigma}(\overline{\lambda}))$ in Theorem 5. The proof of Theorem 1 is a continuation of the proof of Theorem 5.
Proof of Theorem 1. Since $|\Theta'(x)| \leq \frac{1}{\sqrt{\pi}(x^2 \vee 1)}$, we deduce that

$$\left|\Theta\left(\overline{\lambda}\overline{\sigma}(\overline{\lambda})\right) - \Theta(x)\right| \leq \frac{1}{\sqrt{\pi}} \frac{|x - \overline{\lambda}\overline{\sigma}(\overline{\lambda})|}{(\overline{\lambda}^2\overline{\sigma}^2(\overline{\lambda}) \wedge x^2) \vee 1}. \tag{38}$$

Using Lemma 4, we have for all $0 \leq \overline{\lambda} \leq \frac{1}{B}$,

$$\left(1 - \frac{B^\delta\overline{\lambda}^\delta}{1+\delta}\right) e^{-\overline{\lambda}}\overline{\lambda}\sigma \leq \frac{B_n(\overline{\lambda})}{\sigma} = x \leq (e^{\overline{\lambda}} - 1)\sigma. \tag{39}$$

By the estimation of $\overline{\sigma}(\overline{\lambda})$ in Lemma 6, it follows that for all $0 \leq \overline{\lambda}B \leq 6^{-1/\delta}$,

$$
\begin{aligned}
\left|x-\overline{\lambda}\overline{\sigma}(\overline{\lambda})\right| &\leq \overline{\lambda}\sigma\left[\Big(\frac{e^{\overline{\lambda}}-1}{\overline{\lambda}}-\sqrt{1-5B^{\delta}\overline{\lambda}^{\delta}}\Big)\vee\Big(e^{\frac{\overline{\lambda}}{2}}-\Big(1-\frac{B^{\delta}\overline{\lambda}^{\delta}}{1+\delta}\Big)e^{-\overline{\lambda}}\Big)\right] \\
&\leq \overline{\lambda}\sigma\left[\Big(1+\frac{1}{2}\overline{\lambda}Be^{\overline{\lambda}B}-(1-4B^{\delta}\overline{\lambda}^{\delta})\Big)\vee\Big(\frac{3}{2}\overline{\lambda}B+\frac{1}{8}\overline{\lambda}^{2}B^{2}e^{\frac{\overline{\lambda}B}{2}}+\overline{\lambda}^{\delta}B^{\delta}e^{-\overline{\lambda}B}\Big)\right] \\
&\leq \overline{\lambda}\sigma\left[\Big(4B^{\delta}\overline{\lambda}^{\delta}+\frac{1}{2}e^{\frac{1}{6}}\overline{\lambda}B\Big)\vee\Big(\frac{3}{2}\overline{\lambda}B+\frac{1}{48}e^{\frac{1}{12}}\overline{\lambda}B+\overline{\lambda}^{\delta}B^{\delta}\Big)\right] \\
&\leq 4.6\,(\overline{\lambda}B)^{\delta}\overline{\lambda}\sigma \qquad (40)
\end{aligned}
$$

and

$$
\begin{aligned}
\overline{\lambda}^{2}\overline{\sigma}^{2}(\overline{\lambda})\wedge x^{2} &\geq \Big(\overline{\lambda}^{2}\sigma^{2}(1-5B^{\delta}\overline{\lambda}^{\delta})\Big)\wedge\Big(\overline{\lambda}^{2}\sigma^{2}\Big(1-\frac{B^{\delta}\overline{\lambda}^{\delta}}{1+\delta}\Big)^{2}\Big)e^{-2\overline{\lambda}} \\
&\geq \frac{1}{6}\,\overline{\lambda}^{2}\sigma^{2}\wedge e^{-\frac{1}{3}}\Big(\frac{5}{6}\Big)^{2}\overline{\lambda}^{2}\sigma^{2} \\
&\geq \frac{1}{6}\,\overline{\lambda}^{2}\sigma^{2}. \qquad (41)
\end{aligned}
$$

Hence, inequality (38) implies that for all $0 \leq \overline{\lambda} \leq \frac{1}{6^{1/\delta}B}$,

$$
\begin{aligned}
\Big|\Theta\left(\overline{\lambda}\overline{\sigma}(\overline{\lambda})\right)-\Theta(x)\Big| &\leq \frac{27.6}{\sqrt{\pi}(\overline{\lambda}\sigma)^{1-\delta}}\Big(\frac{B}{\sigma}\Big)^{\delta}\mathbf{1}_{\{\overline{\lambda}\sigma\geq\sqrt{6}\}}+\frac{4.6}{\sqrt{\pi}}(\overline{\lambda}\sigma)^{1+\delta}\Big(\frac{B}{\sigma}\Big)^{\delta}\mathbf{1}_{\{\overline{\lambda}\sigma<\sqrt{6}\}} \\
&\leq \frac{27.6}{\sqrt{\pi}}\Big(\frac{B}{\sigma}\Big)^{\delta}. \qquad (42)
\end{aligned}
$$

Therefore, by Lemma 6 and condition (9), it is easy to see that for all $0 \leq \overline{\lambda} \leq \frac{1}{6^{1/\delta}B}$,

$$
\frac{e^{\overline{\lambda}}}{\overline{\sigma}^{2+\delta}(\overline{\lambda})}\sum_{i=1}^{n}\mathbf{E}|\xi_i|^{2+\delta} \leq \frac{e^{1/6}}{(5/6)^{1+\delta/2}}\,\frac{\sum_{i=1}^{n}\mathbf{E}|\xi_i|^{2+\delta}}{\sigma^{2+\delta}} \leq 1.71\Big(\frac{B}{\sigma}\Big)^{\delta}. \qquad (43)
$$

By (39) and the inequality $e^{-x} \geq 1-x$ for all $x \geq 0$, it follows that

$$
\Big(1-2B^{\delta}\overline{\lambda}^{\delta}\Big)\,\overline{\lambda} \leq \frac{x}{\sigma}
$$

and

$$
0 \leq \overline{\lambda} \leq \frac{1}{6^{1/\delta}B} \quad \text{ for all } \quad 0 \leq x\frac{B}{\sigma} \leq \frac{2}{3}6^{-1/\delta}. \qquad (44)
$$

Combining (34), (42) and (43) together, we have for all $0 \leq x \leq \frac{2}{3}6^{-1/\delta}\frac{\sigma}{B}$,

$$
\mathbf{P}(S_n \geq x\sigma) = \left(\Theta(x)+\theta\Big(\frac{27.6}{\sqrt{\pi}}+27.36\widetilde{C}_{\delta}\Big)\Big(\frac{B}{\sigma}\Big)^{\delta}\right)\inf_{\lambda\geq 0}\mathbf{E}e^{\lambda(S_n-x\sigma)},
$$

where $\widetilde{C}_{\delta}$ is the smallest constant such that (33) holds. By Theorem 2.1 of Chen and Shao [7], we have $\widetilde{C}_{\delta} \leq 4.1$. Thus $\frac{27.6}{\sqrt{\pi}}+27.36\widetilde{C}_{\delta} \leq 127.75$. This completes the proof of Theorem 1. □

Proof of Corollary 1. Using (36) and Lemma 5, we get for all $x \geq 0$,

$$\begin{aligned}
&\inf_{\lambda\geq 0} \mathbf{E}e^{\lambda(S_n - x\sigma)} \\
&= \inf_{\lambda\geq 0} \mathbf{E}e^{-x\lambda\sigma+\Psi_n(\lambda)} \\
&\leq \inf_{\lambda\geq 0} \exp\left\{-\lambda x\sigma + n\log\left(\frac{1}{1+\sigma^2/n}\exp\left\{-\lambda\sigma^2/n\right\} + \frac{\sigma^2/n}{1+\sigma^2/n}\exp\{\lambda\}\right)\right\} \\
&= H_n(x,\sigma). \qquad (45)
\end{aligned}$$

Notice that $\inf_{\lambda\geq 0}\mathbf{E}e^{\lambda(S_n-x\sigma)} \leq \mathbf{E}e^{0\cdot(S_n-x\sigma)} = 1$. Thus, the desired inequality follows from Theorem 1 and (45). □

Proof of Corollary 2. Since $\Theta(x)$ is decreasing in $x \geq 0$, we deduce that $\Theta(x) \leq \Theta(\check{x})$ where $\check{x} = \frac{x}{\sqrt{1+\frac{x}{3\sigma}}}$. Notice that Hoeffding's bound is less than Bernstein's bound, i.e.

$$H_n(x,\sigma) \leq \exp\left\{-\frac{\check{x}^2}{2}\right\}$$

(cf. Remark 2.1 of [10]). Therefore, from (12), we have for all $0 \leq x \leq \frac{\sigma}{C_\delta B}$,

$$\begin{aligned}
\mathbf{P}(S_n \geq x\sigma) &\leq \left(\Theta(\check{x}) + C\Big(\frac{B}{\sigma}\Big)^\delta\right)\exp\left\{-\frac{\check{x}^2}{2}\right\} \\
&= 1-\Phi(\check{x}) + C\Big(\frac{B}{\sigma}\Big)^\delta \exp\left\{-\frac{\check{x}^2}{2}\right\}.
\end{aligned}$$

Using (6), we obtain for all $0 \leq x \leq \frac{\sigma}{C_\delta B}$,

$$\mathbf{P}(S_n \geq x\sigma) \leq \Big(1-\Phi(\check{x})\Big)\left[1 + C\,(1+\check{x})\Big(\frac{B}{\sigma}\Big)^\delta\right].$$

This completes the proof of Corollary 2. □

7 Proof of Theorem 3

Consider the positive random variable

$$H_n(\lambda) = \prod_{i=1}^n \frac{e^{\zeta_i(\lambda)}}{\mathbf{E}e^{\zeta_i(\lambda)}}, \qquad \lambda \geq 0,$$

so that $\mathbf{E}H_n(\lambda) = 1$. Introduce the following new conjugate probability measure $\mathbf{P}_\lambda$ defined by

$$d\mathbf{P}_\lambda = H_n(\lambda)d\mathbf{P}. \qquad (46)$$

In this section, denote by $\mathbf{E}_\lambda$ the expectation with respect to $\mathbf{P}_\lambda$ defined by (46). According to (46), we have the following representation: For all $0 \leq x = o(\frac{\sigma}{B})$,

$$\begin{aligned}
\mathbf{P}\Big(\sum_{i=1}^n \zeta_i(\lambda) \geq \frac{x^2}{2}\Big) &= \mathbf{E}_\lambda\Big[H_n(\lambda)^{-1}\mathbf{1}_{\{\sum_{i=1}^n \zeta_i(\lambda)\geq \frac{x^2}{2}\}}\Big] \\
&= \mathbf{E}_\lambda\Big[\exp\Big\{-\sum_{i=1}^n \eta_i(\lambda) - B_n(\lambda) + \Psi_n(\lambda)\Big\}\mathbf{1}_{\{\sum_{i=1}^n \zeta_i(\lambda)\geq \frac{x^2}{2}\}}\Big]
\end{aligned}$$

$$= \mathbf{E}_\lambda\Big[\exp\Big\{-\sum_{i=1}^n \eta_i(\lambda) - B_n(\lambda) + \Psi_n(\lambda)\Big\}\mathbf{1}_{\{\sum_{i=1}^n \zeta_i(\lambda)\geq \frac{x^2}{2}\}}\Big]$$
$$= \mathbf{E}_\lambda\Big[\exp\Big\{-\sum_{i=1}^n \eta_i(\lambda) - B_n(\lambda) + \Psi_n(\lambda)\Big\}\mathbf{1}_{\{\sum_{i=1}^n \eta_i(\lambda)\geq \frac{x^2}{2}-B_n(\lambda)\}}\Big] \quad (47)$$

where

$$\eta_i(\lambda) = \zeta_i(\lambda) - \mathbf{E}_\lambda\zeta_i(\lambda)$$

and

$$B_n(\lambda) = \sum_{i=1}^n \mathbf{E}_\lambda\zeta_i(\lambda) = \sum_{i=1}^n \frac{\mathbf{E}\,\zeta_i(\lambda)e^{\zeta_i(\lambda)}}{\mathbf{E}\,e^{\zeta_i(\lambda)}}$$

Let $\overline{\lambda} = \overline{\lambda}(x)$ be the smallest positive solution of the following equation

$$B_n(\lambda) = \frac{x^2}{2}.$$

Recall

$$Y_n(\lambda) = \sum_{i=1}^n \eta_i(\lambda).$$

From (47), we obtain

$$\mathbf{P}\Big(\sum_{i=1}^n \zeta_i(\overline{\lambda}) \geq \frac{x^2}{2}\Big) = \exp\left\{-\frac{x^2}{2} + \Psi_n(\overline{\lambda})\right\}\mathbf{E}_{\overline{\lambda}}[e^{-Y_n(\overline{\lambda})}\mathbf{1}_{\{Y_n(\overline{\lambda})\geq 0\}}]$$
$$= \exp\left\{-\frac{x^2}{2} + \Psi_n(\overline{\lambda})\right\}\int_0^\infty e^{-y}\mathbf{P}_{\overline{\lambda}}(0 \leq Y_n(\overline{\lambda}) \leq y)dy. \quad (48)$$

Set $F_n(y) = \mathbf{P}_{\overline{\lambda}}(Y_n(\overline{\lambda}) \leq y)$. Recall $\widetilde{C}_\delta \leq 4.1$ (see Theorem 2.1 of Chen and Shao [7]). By an argument similar to the proof of Lemma 7, we have for all $0 \leq x = o(\frac{\sigma}{B})$,

$$\sup_{y\in\mathbf{R}}\Big|F_n(y) - \Phi(y/\overline{\lambda}\,\sigma)\Big| \leq C_1\Big(\frac{B}{\sigma}\Big)^\delta.$$

Hence, for all $0 \leq x = o(\frac{\sigma}{B})$,

$$\left|\int_0^\infty e^{-y}\mathbf{P}_{\overline{\lambda}}(0 \leq Y_n(\overline{\lambda}) \leq y)dy - \int_0^\infty e^{-y}\mathbf{P}(0 \leq \mathcal{N}(0,1) \leq y/\overline{\lambda}\,\sigma)dy\right| \leq 2C_1\Big(\frac{B}{\sigma}\Big)^\delta,$$

where $\mathcal{N}(0,1)$ is the standard normal r.v. By a simple calculation, it follows that

$$\int_0^\infty e^{-y}\mathbf{P}(0 \leq \mathcal{N}(0,1) \leq y/\overline{\lambda}\,\sigma)dy = \Theta(\overline{\lambda}\,\sigma).$$

From (48), we have for all $0 \leq x = o(\frac{\sigma}{B})$,

$$\mathbf{P}\Big(\sum_{i=1}^n \zeta_i(\overline{\lambda}) \geq \frac{x^2}{2}\Big) = \exp\left\{-\frac{x^2}{2} + \Psi_n(\overline{\lambda})\right\}\left(\Theta(\overline{\lambda}\,\sigma) + \theta C_2\Big(\frac{B}{\sigma}\Big)^\delta\right). \quad (49)$$

Next, we would like to substitute x for $\overline{\lambda}\sigma$ in the item $\Theta(\overline{\lambda}\sigma)$. By an argument similar to the proof of (42), we get for all $0 \leq x = o(\frac{\sigma}{B})$,

$$\Big|\Theta(\overline{\lambda}\sigma) - \Theta(x)\Big| \leq C_3\Big(\frac{B}{\sigma}\Big)^\delta. \quad (50)$$

So we have for all $0 \leq x = o(\frac{\sigma}{B})$,

$$\mathbf{P}\Big(\sum_{i=1}^{n} \zeta_i(\overline{\lambda}) \geq \frac{x^2}{2}\Big) = \exp\left\{ -\frac{x^2}{2} + \Psi_n(\overline{\lambda}) \right\} \left(\Theta(x) + \theta C\Big(\frac{B}{\sigma}\Big)^{\delta} \right), \tag{51}$$

which gives the desired expansion of tail probabilities.

In the sequel we shall give an estimation for $\Psi_n(\overline{\lambda})$. To this end, we need the following useful lemma (cf. Lemma 6.2 of Jing, Shao and Wang [15]).

Lemma 8 *Let X be a random variable with $\mathbf{E}X = 0$ and $\mathbf{E}X^2 < \infty$. For $\lambda > 0$, let $\zeta = \lambda X - \frac{1}{2}(\lambda X)^2$. Then for $\lambda > 0$,*

$$\mathbf{E}e^{\zeta} = 1 + O(1)\varepsilon_\lambda, \tag{52}$$

$$\mathbf{E}\zeta e^{\zeta} = \frac{1}{2}\lambda^2 \mathbf{E}X^2 + O(1)\varepsilon_\lambda, \tag{53}$$

where

$$\varepsilon_\lambda = \lambda^2 \mathbf{E}[X^2 \mathbf{1}_{\{|\lambda X|>1\}}] + \lambda^3 \mathbf{E}[|X|^3 \mathbf{1}_{\{|\lambda X|\leq 1\}}].$$

Notice that

$$\varepsilon_\lambda \leq \mathbf{E}|\lambda X|^{2+\delta} = \lambda^{2+\delta}\mathbf{E}|X|^{2+\delta}.$$

Since $\mathbf{E}|\xi_i|^{2+\delta} \leq B^{\delta}\mathbf{E}\xi_i^2$, by Lemmas 2 and 8, for any $\lambda = o(B^{-1})$,

$$\Psi_n(\lambda) = O(1)\sum_{i=1}^{n} \lambda^{2+\delta}\mathbf{E}|\xi_i|^{2+\delta} = O(1)\lambda^{2+\delta}B^{\delta}\sigma^2 \tag{54}$$

and

$$\sum_{i=1}^{n} \frac{\mathbf{E}\,\zeta_i(\lambda)e^{\zeta_i(\lambda)}}{\mathbf{E}\,e^{\zeta_i(\lambda)}} = \frac{1}{2}\lambda^2\sigma^2 + O(1)\lambda^{2+\delta}B^{\delta}\sigma^2.$$

Recall that $\overline{\lambda} = \overline{\lambda}(x) > 0$ is the smallest positive solution of the following equation

$$\sum_{i=1}^{n} \frac{\mathbf{E}\,\zeta_i(\lambda)e^{\zeta_i(\lambda)}}{\mathbf{E}\,e^{\zeta_i(\lambda)}} = \frac{x^2}{2},$$

we have for all $0 \leq x = o(\frac{\sigma}{B})$,

$$\overline{\lambda} = O(1)\frac{x}{\sigma}.$$

Thus, from (54), it holds

$$\Psi_n(\overline{\lambda}) = O(1)x^{2+\delta}\Big(\frac{B}{\sigma}\Big)^{\delta},$$

which gives the estimation of $\Psi_n(\overline{\lambda})$. This completes the proof of Theorem 3 □

Acknowledgements

We would like to thank the anonymous referee and editor for their helpful comments and suggestions. The work has been partially supported by the National Natural Science Foundation of China (Grant nos. 11601375 and 11626250).

References

1. Bahadur R, Ranga Rao R. On deviations of the sample mean. Ann Math Statist, 1960, 31: 1015-1027
2. Bennett G. Probability inequalities for sum of independent random variables. J Amer Statist Assoc, 1962, 57: 33-45
3. Bentkus V. On Hoeffding's inequality. Ann Probab, 2004, 32: 1650-1673
4. Bentkus V, Kalosha N, van Zuijlen M. On domination of tail probabilities of (super)martingales: explicit bounds. Lithuanian Math J, 2006, 46: 1-43
5. Bercu B, Rouault A. Sharp large deviations for the Ornstein-Uhlenbeck process. Theory Probab Appl, 2006, 46: 1-19
6. Borovkov A A, Mogulskii A A. On large and superlarge deviations of sums of independent random vectors under Cramer's condition. Theory Probab Appl, 2007, 51: 567-594
7. Chen L H Y, Shao Q M. A non-uniform BerryCEsseen bound via Stein's method. Probab Theory Relat Fields, 2001, 120(2): 236-254
8. Cramér H. Sur un nouveau théorème-limite de la théorie des probabilités. Actualite's Sci Indust, 1938, 736: 5-23
9. Efron B. Student's t-test under symmetry conditions. J Amer Statist Assoc, 1969, 64(328): 1278-1302
10. Fan X, Grama I, Liu Q. Hoeffding's inequality for supermartingales. Stochastic Process Appl, 2012, 122: 3545-3559
11. Fan X, Grama I, Liu Q. Cramér large deviation expansions for martingales under Bernstein's condition. Stochastic Process Appl, 2013, 123: 3919-3942
12. Fan X, Grama I, Liu, Q. Sharp large deviation results for sums of independent random variables. Sci China Math, 2015, 58: 1939-1958
13. Grama I, Haeusler E. Large deviations for martingales via Cramer's method. Stochastic Process Appl, 2000, 85: 279-293
14. Hoeffding W. Probability inequalities for sums of bounded random variables. J Amer Statist Assoc, 1963, 58: 13-30
15. Jing B Y, Shao Q M, Wang Q. Self-normalized Cramér-type large deviations for independent random variables. Ann Probab, 2003, 31: 2167-2215
16. Nagaev S V. Lower bounds for the probabilities of large deviations of sums of independent random variables. Theory Probab Appl, 2002, 46: 728-735
17. Petrov V V. Sums of Independent Random Variables. Berlin: Springer-Verlag, 1975
18. Petrov V V. Limit Theorems of Probability Theory. Oxford: Oxford University Press, 1995
19. Petrov V V, Robinson J. Large deviations for sums of independent non identically distributed random variables. Comm Statist Theory Methods, 2008, 37: 2984-2990
20. Pinelis I. On the Bennett-Hoeffding inequality. Ann Inst H Poincaré Probab Statist, 2014, 50: 15-27
21. Saulis L, Statulevičius V A. Limit theorems for large deviations. Netherlands: Springer, 1991
22. Sakhanenko A I. Berry-Esseen type bounds for large deviation probabilities. Siberian Math J, 1991, 32: 647-656
23. Shao Q M. A Cramér type large deviation result for Student's t-statistic. J Theoret Probab, 1999, 12: 385-398
24. Talagrand M. The missing factor in Hoeffding's inequalities. Ann Inst H Poincaré Probab Statist, 1995, 31: 689-702