To appear in *Statistics: A Journal of Theoretical and Applied Statistics*
Vol. 00, No. 00, Month 20XX, 1–22

Nonuniform Berry-Esseen bounds for martingales with applications to statistical estimation

Dedicated to Paul Doukhan on his sixtieth birthday

X. Fan[a,b,*], I. Grama[c] and Q. Liu[c,*]

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

(*Submitted 31 Octobre 2015*)

We establish nonuniform Berry-Esseen bounds for martingales under the conditional Bernstein condition. These bounds imply Cramér type large deviations for moderate x's, and are of exponential decay rate as de la Peña's inequality when $x \to \infty$. Statistical applications associated with linear regressions and self-normalized large deviations are also provided.

Keywords: Nonuniform Berry-Esseen bound; Cramér large deviations; Exponential inequality; Linear regressions; Self-normalized large deviations

AMS Subject Classification: 60G42; 60F05; 60F10; 60E15; 62E20; 62J05

1. Introduction

Let $n \geq 1$ be an integer and $(\xi_i)_{i=1,\dots,n}$ be a sequence of independent random variables with zero means and finite variances on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Let $S_n = \sum_{i=1}^n \xi_i$. Without loss of generality, we assume that $\mathbf{E}[S_n^2] = 1$, where $\mathbf{E}$ is the expectation corresponding to $\mathbf{P}$. Nonuniform normal approximation bounds have been first obtained by Esseen (1945) for identically distributed random variables with finite third moments. These were improved to $C\, n\, \mathbf{E}[|\xi_1|^3]/(1+|x|^3)$ by Nagaev (1965), where, throughout this paper, C stands for an absolute constant with possibly different values in different places. Bikelis (1966) generalized Nagaev's result to the sums of non-identically distributed random variables with moments of order larger than 2. Suppose that there exists a constant $\delta \in (0, 1]$ such that $\mathbf{E}[|\xi_i|^{2+\delta}] < \infty$ for all $i \in [1, n]$. Bikelis (1966) (cf. also Petrov (1975), p.132) has established the following nonuniform Berry-Esseen bound: for all $x \in \mathbf{R}$,

$$\Big|\mathbf{P}(S_n \leq x) - \Phi(x)\Big| \leq C\, \frac{\sum_{i=1}^n \mathbf{E}[|\xi_i|^{2+\delta}]}{(1+|x|)^{2+\delta}}, \tag{1}$$

*Fan and Liu are both corresponding authors.
*Email: fanxiequan@hotmail.com (X. Fan), quansheng.liu@univ-ubs.fr (Q. Liu)

where $\Phi(x)$ is the standard normal distribution function. Note that Bikelis' bound implies the Berry-Esseen bound

$$D(S_n) := \sup_{x\in\mathbf{R}} \Big|\mathbf{P}(S_n \leq x) - \Phi(x)\Big| \leq C \sum_{i=1}^{n} \mathbf{E}[|\xi_i|^{2+\delta}], \tag{2}$$

which is known to be optimal. On the other hand the bound (1) decays in the best possible polynomial rate $1/|x|^{2+\delta}$ when $|x| \to \infty$. For random variables without assuming the existence of moments of order lager than 2, Bikelis type bound has been established by Chen and Shao (2001) via Stein's method: they showed that, for all $x \in \mathbf{R}$,

$$\Big|\mathbf{P}(S_n \leq x) - \Phi(x)\Big| \ \leq \ C\sum_{i=1}^{n} \left(\frac{\mathbf{E}[\xi_i^2 \mathbf{1}_{\{|\xi_i|>1+|x|\}}]}{(1+|x|)^2} + \frac{\mathbf{E}[|\xi_i|^3 \mathbf{1}_{\{|\xi_i|\leq 1+|x|\}}]}{(1+|x|)^3} \right). \tag{3}$$

It is easy to see that (3) implies (1), since $\mathbf{E}[\xi_i^2 \mathbf{1}_{\{|\xi_i|>1+|x|\}}] \leq \mathbf{E}[|\xi_i|^{2+\delta}]/(1+|x|)^{\delta}$ and $\mathbf{E}[|\xi_i|^3 \mathbf{1}_{\{|\xi_i|\leq 1+|x|\}}] \leq \mathbf{E}[|\xi_i|^{2+\delta}](1+|x|)^{1-\delta}$ for $\delta \in (0,1]$.

Now we consider the case of martingales. The generalization of Bikelis' bound (1) can be found in Hall and Heyde (1980, 1981) and Haeusler and Joos (1988b). Assume that $(\xi_i, \mathcal{F}_i)_{i=1,\ldots,n}$ is a sequence of martingale differences defined on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Here ξ_i's may depend on n. Denote by $\langle S\rangle_n = \sum_{i=1}^n \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}]$ the quadratic characteristic of S_n. Haeusler and Joos proved that if $\mathbf{E}[|\xi_i|^{2+\delta}] < \infty$ for some $\delta > 0$ and for all $i \in [1,n]$, then there exists a constant C_δ, depending only on δ, such that, for all $x \in \mathbf{R}$,

$$\Big|\mathbf{P}(S_n \leq x) - \Phi(x)\Big| \leq \ C_\delta \left(\sum_{i=1}^{n} \mathbf{E}[|\xi_i|^{2+\delta}] + \mathbf{E}[|\langle S\rangle_n - 1|^{1+\delta/2}] \right)^{1/(3+\delta)} \frac{1}{1+|x|^{2+\delta}}; \tag{4}$$

see also Hall and Heyde (1980, 1981) with the larger factor $\frac{1}{1+|x|^{4(1+\delta/2)^2/(3+\delta)}}$ replacing $\frac{1}{1+|x|^{2+\delta}}$ of (4). Similar to Bikelis' bound, inequality (4) implies the following Berry-Essen bound for martingales:

$$D(S_n) \ \leq \ C_\delta \left(\sum_{i=1}^{n} \mathbf{E}[|\xi_i|^{2+\delta}] + \mathbf{E}[|\langle S\rangle_n - 1|^{1+\delta/2}] \right)^{1/(3+\delta)}. \tag{5}$$

Moreover, Haeusler (1988a) showed that the Berry-Esseen bound (5) is the best possible under the stated conditions. The bound (4) also decays with the best possible polynomial rate $1/|x|^{2+\delta}$ when $|x| \to \infty$.

Apart the nonuniform Berry-Esseen bounds with polynomially decaying rate there are few nonuniform bounds with exponentially decaying rate. The only reference we aware of is Joss (1991) for bounded martingale differences. We refer also to Račkauskas (1990,1995), Grama (1997) and Grama and Haeusler (2000, 2006) where moderate deviations have been obtained.

Assume the following martingale version of Bernstein's conditions, where it is assumed that the martingale differences have moments of all orders:

(A1) There exists some positive number $\epsilon \in (0, 1/2]$ such that

$$\Big|\mathbf{E}[\xi_i^k|\mathcal{F}_{i-1}]\Big| \leq \frac{1}{2}k!\epsilon^{k-2}\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] \quad a.s. \text{ for all } k \geq 2 \text{ and all } 1 \leq i \leq n.$$

(A2) There exists a nonnegative number $\delta \in [0, 1]$ such that

$$|\langle S\rangle_n - 1| \leq \delta^2 \ \ a.s.$$

Here ϵ and δ usually depend on n such that $\epsilon \to 0, \delta \to 0$ as $n \to \infty$. In particular, when $(\xi_i)_{i=1,...,n}$ are independent, condition (A2) is satisfied with $\delta = 0$ for normalized S_n and condition (A1) is known as the Bernstein condition. If $(\xi_i)_{i=1,...,n}$ are also identically distributed, then (A1) holds with $\epsilon = \frac{C}{\sqrt{n}}$ as $n \to \infty$.

Under condition (A1), de la Peña (1999) has obtained the following martingale version of Bennett's inequality (1962) (see also Bernstein (1927)), for all $x, v > 0$,

$$\mathbf{P}(S_n > x, \langle S\rangle_n \leq v^2) \leq \exp\left\{-\frac{x^2}{v^2 + \sqrt{1 + 2x\epsilon/v^2} + x\epsilon}\right\} \tag{6}$$

$$\leq \exp\left\{-\frac{x^2}{2(v^2 + x\epsilon)}\right\}. \tag{7}$$

Recent improvements of (6) are given in Theorem 3.14 of Bercu, Deylon and Rio (2015) and Fan, Grama and Liu (2015b). We refer to Shorack and Wellner (1986) and van de Geer (1995) for inequality (7). Moreover, if, in addition, condition (A2) holds, then de la Peña's inequality (6) implies that, for all $x \geq 0$,

$$\mathbf{P}(S_n > x) \leq \exp\left\{-\frac{\widehat{x}^2}{2}\right\}, \tag{8}$$

where

$$\widehat{x} = \frac{2x/\sqrt{1+\delta^2}}{1 + \sqrt{1 + 2x\epsilon/(1+\delta^2)}}. \tag{9}$$

Since $\widehat{x} \to x$ as $\max\{\epsilon, \delta\} \to 0$, the bound (8) is exponentially decaying with rate $\exp\{-x^2/2\}$ when $\max\{\epsilon, \delta\} \to 0$. Thus, the bound (8) is tight. By considering the martingale differences $(-\xi_i, \mathcal{F}_i)_{i=1,...,n}$, the bound (8) holds equally on tail probabilities $\mathbf{P}(S_n < -x)$, $x \geq 0$. Due to this fact, we expect to establish a nonuniform Berry-Esseen bound with exponentially decaying rate as in (8) when $|x| \to \infty$.

Our main result is the following nonuniform Berry-Esseen bound for martingales. Assume conditions (A1) and (A2). Then, for all $x \geq 0$,

$$\Big|\mathbf{P}(S_n \leq x) - \Phi(x)\Big| \leq C\left(1 + x^2\right)\left(\epsilon|\log\epsilon| + \frac{\delta}{1+|x|}\right)\exp\left\{-\frac{\widehat{x}^2}{2}\right\}, \tag{10}$$

where $\widehat{x}$ is defined by (9). To prove (10), we need the following strengthened version of de la Peña's inequality (8): for all $x \geq 0$,

$$\mathbf{P}(S_n > x) \leq \Big(1 - \Phi(\widehat{x})\Big)\Big[1 + C\Big(1 + \widehat{x}\Big)\Big(\lambda^2\epsilon + \lambda\delta^2 + \epsilon\,|\log\epsilon| + \delta\Big)\Big], \tag{11}$$

with $\lambda \in [0,\ \epsilon^{-1})$, $\lambda = x + O\big(x^2\epsilon + x\delta\big)$, $\ x^2\epsilon + x\delta \to 0$; see Theorem 2.3 for details. We will show that bound (11) strengthens de la Peña's inequality (8) by adding a factor of type $\frac{1}{1+x}$.

Of course, condition (A2) is very restrictive. Without condition (A2), under solely condition (A1), with a method of Bolthausen (1982), we establish the following nonuniform Berry-Esseen bound under condition (A1): for all $x \in \mathbf{R}$,

$$\Big|\mathbf{P}(S_n \leq x) - \Phi\left(x\right)\Big| \leq C\bigg(\left(1+x^2\right)\epsilon|\log\epsilon|\exp\left\{-\frac{\breve{x}^2}{2}\right\} + \Big(\mathbf{E}|\langle S\rangle_n - 1| + \epsilon^2\Big)^{1/3}\exp\left\{-\frac{x^2}{6}\right\}\bigg), \tag{12}$$

where

$$\breve{x} = \frac{2|x|}{1+\sqrt{1+2|x|\epsilon}}. \tag{13}$$

This result has an exponential decaying rate in x, compared to the polynomial decaying rate in the nonuniform Berry-Esseen bounds of Haeusler and Joos (1988b) and Joos (1991).

The bounds (10) and (12) are closely related to the results of Fan et al. (2013) and Grama and Haeusler (2000). However, we complete on these results in three aspects. First, we establish nonuniform Berry-Esseen bounds, which imply the Cramér large deviations of Fan et al. (2013) and Grama and Haeusler (2000) in the normal range $0 \leq x = o(\epsilon^{-1/3})$. Second, we relax condition (A2) of Fan et al. (2013), replacing it by boundedness of the moment $\mathbf{E}|\langle S\rangle_n - 1|$. Third, our bounds hold for all $x \in \mathbf{R}$, compared with the range $0 \leq x = o(\epsilon^{-1})$ established in Fan et al. (2013).

The paper is organized as follows. Our results are stated and discussed in Section 2. Some applications to linear regressions and self-normalized large deviations are presented in Section 3. The proofs of the results are deferred to Sections 4 and 5.

2. Main Results

Assume that we are given a sequence of martingale differences $(\xi_i, \mathcal{F}_i)_{i=0,...,n}$, defined on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$, where $\xi_0 = 0$ and $\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq ... \subseteq \mathcal{F}_n \subseteq \mathcal{F}$ are increasing σ-fields. Consider the martingale $S = (S_k, \mathcal{F}_k)_{k=0,...,n}$, where

$$S_0 = 0, \qquad S_k = \sum_{i=1}^{k} \xi_i, \quad k = 1, ..., n. \tag{14}$$

Let $\langle S\rangle$ be its predictable quadratic variation:

$$\langle S\rangle_0 = 0, \qquad \langle S\rangle_k = \sum_{i=1}^{k} \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}], \quad k = 1, \ldots, n. \tag{15}$$

Our main result is the following nonuniform Berry-Esseen bound for martingales.

THEOREM 2.1 *Assume conditions (A1) and (A2). Then, for all $x \in \mathbf{R}$,*

$$\Big| \mathbf{P}(S_n \leq x) - \Phi\left(x\right) \Big| \leq C \left(1 + x^2\right) \left(\epsilon |\log \epsilon| + \frac{\delta}{1+|x|} \right) \exp\left\{ - \frac{\widehat{x}^2}{2} \right\}, \tag{16}$$

where

$$\widehat{x} = \frac{2|x|/\sqrt{1+\delta^2}}{1+\sqrt{1+2|x|\epsilon/(1+\delta^2)}}. \tag{17}$$

Let us give some comments on the main result.

(1) When $(\xi_i)_{i=1,...,n}$ is a sequence of independent random variables, it is possible to improve the factor $\epsilon |\log \epsilon| + \frac{\delta}{1+|x|}$ in (16) to ϵ. However, this task is beyond the scope of this paper. We refer to [14] for related bounds.

(2) The exponential factor $\exp\{-\widehat{x}^2/2\}$ in (16) has the same exponential decay rate as de la Peña's bound (8) for all x. For moderate x, this exponential factor has the exponentially decaying rate $\exp\{-x^2/2\}$ as $\max\{\epsilon, \delta\} \to 0$. When the martingale differences are bounded, Joos (1991) also established two nonuniform Berry-Esseen bounds for martingales with exponential decay rates. However, the exponential decay rate in Joos (1991) is much slower than that of (16).

(3) Inequality (16) implies the following Cramér large deviation expansion in the normal range: for all $0 \leq x \leq \min\{\epsilon^{-1/3}, \delta^{-1}\}$,

$$\frac{\mathbf{P}(S_n > x)}{1-\Phi\left(x\right)} = 1 + \theta C \left(1 + x^3\right) \left(\epsilon |\log \epsilon| + \frac{\delta}{1+x} \right), \tag{18}$$

where $|\theta| \leq 1$. Indeed, since $0 \leq x - \widehat{x} = O\Big(x^2\epsilon + |x|\delta\Big)$, $x^2\epsilon + |x|\delta \to 0$, and

$$\frac{1}{\sqrt{2\pi}(1+x)} \leq \Big(1 - \Phi\left(x\right)\Big) \exp\left\{ \frac{x^2}{2} \right\} \leq \frac{1}{\sqrt{\pi}(1+x)} \tag{19}$$

for all $x \geq 0$, we easily obtain (18) from (16). Some earlier results of type (18) have been established by Bose (1986a, 1986b), Račkauskas (1990, 1995) and Grama and Haeusler (2000) for martingales with bounded differences; see also Fan et al. (2013) under conditions (A1) and (A2). Notice that the factor $1 + x^3$ in (18) is the best possible. Thus, the factor $1 + x^2$ in (16) also cannot be improved to a smaller one.

(4) When $|\xi_i| \leq \epsilon$ and condition (A2) holds, Bolthausen (1982) proved that

$$D(S_n) := \sup_{x\in\mathbf{R}} \Big| \mathbf{P}(S_n \leq x) - \Phi(x) \Big| \leq C \left(\epsilon^3 n \log n + \delta \right). \tag{20}$$

As pointed out by Bolthausen (1982), the convergence rate in (20) is sharp in the sense that there exists a sequence of bounded martingale differences $|\xi_i| \leq C/\sqrt{n}$ satisfying $\langle S \rangle_n = 1$ a.s. such that

$$\limsup_{n\to\infty} \sqrt{n}\ (\log n)^{-1} D(S_n) > 0. \tag{21}$$

The factor $\log n$ in the convergence rate is the major difference between the Berry-

Esseen bounds for martingale difference arrays (under suitable conditions) and for i.i.d. sequences, where the Berry-Esseen bounds are of order $1/\sqrt{n}$. It is known that under certain stronger condition (for instance $\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] = 1/n$, $\mathbf{E}[\xi_i^3|\mathcal{F}_{i-1}] = C_1/n^{3/2}$ and $\mathbf{E}[|\xi_i|^{3+\delta}|\mathcal{F}_{i-1}] \leq C_2/n^{(3+\delta)/2}, \delta > 0$, a.s. for all $1 \leq i \leq n$), an uniform Berry-Esseen bound of order $1/\sqrt{n}$ for martingale difference arrays is possible; we refer to Renz (1996) (see also Bolthausen (1982)).

It is easy to see that inequality (16) implies the following Berry-Esseen bound

$$D(S_n) \leq C\Big(\epsilon\,|\log \epsilon| + \delta\Big). \tag{22}$$

For martingales with bounded differences (22) has been established earlier in Grama (1987a,b, 1988). Note that (22) implies Bolthausen's inequality (20) under the less restrictive condition (A1). Indeed, by condition (A2), we have $3/4 \leq \langle X\rangle_n \leq n\epsilon^2$ a.s. and then $\epsilon \geq \sqrt{3/(4n)}$. For $\epsilon \leq 1/2$, it is easy to see that $\epsilon^3 n \log n \geq 3\,\epsilon|\log \epsilon|/4$. Thus (22) implies (20). Note that the bound in (20) may converge to infinity while that in (22) converges to 0 as $\epsilon, \delta \to 0$ and $n \to \infty$. For instance, if ϵ is of the order $n^{-1/3}$ and $\delta = o(1)$ as $n \to \infty$, then it is obvious that $\epsilon\,|\log \epsilon| = O(n^{-1/3}\log n)$ while $\epsilon^3 n \log n \geq \log n$. Thus inequality (22) strengthens Bolthausen's inequality (20).

(5) When condition (A2) fails, the optimal Berry-Esseen bounds for martingales have been obtained by several authors; we refer to Bolthausen (1982), Haeusler (1988a), Grama (1988, 1993) and Mourrat (2013). In these papers, the authors assume that the random variable $\langle S\rangle_n - 1$ has finite moments.

Of course, condition (A2) in our theorem may be very restrictive. Using the method from Bolthausen (1982), we deduce from (2.3) the following nonuniform Berry-Esseen bound where the condition (A2) is relaxed.

Corollary 2.2 *Assume condition (A1). Then, for all $x \in \mathbf{R}$,*

$$\Big|\mathbf{P}(S_n \leq x) - \Phi\,(x)\Big| \leq C\Bigg(\Big(1 + x^2\Big)\,\epsilon\,|\log \epsilon|\exp\left\{-\frac{\breve{x}^2}{2}\right\} + \Big(\mathbf{E}|\langle S\rangle_n - 1| + \epsilon^2\Big)^{1/3}\exp\left\{-\frac{x^2}{6}\right\}\Bigg), \tag{23}$$

where $\breve{x}$ is defined by (13). In particular,

$$\sup_{x\in\mathbf{R}}\Big|\mathbf{P}(S_n \leq x) - \Phi(x)\Big| \leq C\left(\Big(\mathbf{E}|\langle S\rangle_n - 1|\Big)^{1/3} + \epsilon^{2/3}\right). \tag{24}$$

For earlier results for martingales with bounded differences, we refer to Joos (1991), where the slower rate $\exp\left\{-\frac{|x|}{5\log(1+|x|)}\right\}$ has been obtained. If in the hypothesis (A1) the conditional expectation is replaced by the non-conditional expectation, then the convergence rate is less sharp than that in (23), as shown by Theorem 3.2 of Lesigne and Volný (2001).

Inspecting the proof of Theorem 1.5 of Mourrat (2013), we see that using the Burkholder inequality instead of Chebyshev inequality (67), in the proof of Corollary 2.2, the bound

(24) can be generalized to

$$\sup_{x\in\mathbf{R}}\Big|\mathbf{P}(S_n\leq x)-\Phi(x)\Big| \leq C_p\left(\Big(\mathbf{E}[|\left\langle S\right\rangle_n-1|^p]\Big)^{1/(2p+1)}+\epsilon^{2p/(2p+1)}\right) \tag{25}$$

for any $p\geq 1$, where C_p depends only on p. Moreover, Mourrat (2013) (see also Grama (1988) and Haeusler (1988a)) showed that when $|\xi_i|\leq C/\sqrt{n}$, a.s., the term $\big(\mathbf{E}[|\left\langle S\right\rangle_n-1|^p]\big)^{1/(2p+1)}$ in the bound (25) is sharp in the sense of that (21) holds true. With $p=1$ this implies that the term $\big(\mathbf{E}|\left\langle S\right\rangle_n-1|\big)^{1/3}$ in (24) is also sharp.

To prove Theorem 2.1, we establish the following strengthened version of de la Peña's inequality (8).

Theorem 2.3 *Assume conditions (A1) and (A2). Then, for all* $x\geq 0$,

$$\mathbf{P}(S_n>x)\leq\Big(1-\Phi\left(\widehat{x}\right)\Big)\Big[1+C\Big(1+\widehat{x}\Big)\Big(\lambda^2\epsilon+\lambda\delta^2+\epsilon\left|\log\epsilon\right|+\delta\Big)\Big], \tag{26}$$

where $\widehat{x}$ *is defined by (17) and*

$$\lambda=\frac{2x/(1+\delta^2)}{1+2x\epsilon/(1+\delta^2)+\sqrt{1+2x\epsilon/(1+\delta^2)}}\in[0,\epsilon^{-1}). \tag{27}$$

In particular, for all $0\leq x=o\Big((\sqrt[3]{\epsilon}+\delta)^{-1}\Big)$

$$\mathbf{P}(S_n>x)\ \leq\ \Big(1-\Phi\left(\widehat{x}\right)\Big)\left[1+o(1)\right],\quad \max\{\epsilon,\delta\}\rightarrow 0. \tag{28}$$

Since the martingale differences $(-\xi_i,\mathcal{F}_i)_{i=1,\dots,n}$ also satisfy conditions (A1) and (A2), the bound (26) can be also applied to obtain upper bounds of $\mathbf{P}(S_n<-x), x>0$. Our bound (26) decays exponentially to zero as $x\rightarrow\infty$ and also recovers closely the shape of the standard normal tail $1-\Phi(x)$ as $\max\{\epsilon,\delta\}\rightarrow 0$.

Using the two sides bound (19), it is easy to see that (26) implies the following inequality: for all $x\geq 0$,

$$\mathbf{P}(S_n>x)\leq F(x)\exp\left\{-\frac{\widehat{x}^2}{2}\right\}, \tag{29}$$

where $F(x)=C\left(\frac{1}{1+\widehat{x}}+\lambda^2\epsilon+\lambda\delta^2+\epsilon\left|\log\epsilon\right|+\delta\right)$. Note that when $\max\{\epsilon,\delta\}\rightarrow 0$, we have $F(x)\rightarrow\frac{C}{1+x}$. Inequality (28) strengthens de la Peña's inequality (8) by adding a factor of order $\frac{1}{1+x}$ as $\max\{\epsilon,\delta\}\rightarrow 0$. Thus (26) is a strengthened version of de la Peña's inequality (8).

Our result (26) can be compared with the classical Cramér large deviation results in the i.i.d. case (see Cramér (1938)). With respect to Cramér's results, the advantage of (26) is that it is valid for all $x\geq 0$ instead of only for all $0\leq x=(\sqrt{n}), n\rightarrow\infty$.

3. Applications in Statistics

3.1. *Linear regression*

The linear regression model is given by

$$X_k = \theta \phi_k + \varepsilon_k, \quad 1 \leq k \leq n, \tag{30}$$

where $(X_k), (\phi_k)$ and (ε_k) are, respectively, the responce variable, the covariate and the noise. We assume that (ϕ_k) is a sequence of independent random variables. We also assume that (ε_k) is a sequence of martingale differences with respect to its natural filtration $\mathcal{F}_k = \sigma\{\varepsilon_i, 1 \leq i \leq k\}$ with conditional variances satisfying $\mathbf{E}[\varepsilon_k^2|\mathcal{F}_{k-1}] = \sigma^2 > 0$ a.s. Moreover, we suppose that the sequences (ϕ_k) and (ε_k) are independent. Our interest is to estimate the unknown parameter θ, based on the random variables (X_k) and (ϕ_k). The well-known least squares estimator θ_n is given by

$$\theta_n = \frac{\sum_{k=1}^n \phi_k X_k}{\sum_{k=1}^n \phi_k^2}. \tag{31}$$

To measure the accuracy of the convergence $\theta_n \to \theta$ as $n \to \infty$, many tight exponential inequalities on tail probability of $\theta_n - \theta$ have been established in Bercu and Touati (2008). Here, we give an estimation on the rate of convergence of the distribution of $(\theta_n - \theta)\sqrt{\sum_{k=1}^n \phi_k^2}$ to the normal one.

THEOREM 3.1 *Assume that there exist two positive numbers ϵ_1 and ϵ_2 satifying $\epsilon := \epsilon_1 \epsilon_2 / \sigma \leq \frac{1}{2}$ and such that*

$$\frac{|\phi_k|}{\sqrt{\sum_{j=1}^n \phi_j^2}} \leq \epsilon_1 \qquad a.s.\ for\ all\ 1 \leq k \leq n \tag{32}$$

and

$$\Big|\mathbf{E}[\varepsilon_k^l|\mathcal{F}_{k-1}]\Big| \leq \frac{1}{2}\, l!\, \epsilon_2^{l-2}\sigma^2 \qquad a.s.\ for\ all\ l \geq 2\ and\ all\ 1 \leq k \leq n. \tag{33}$$

Then, for all $x \in \mathbf{R}$,

$$\Big|\mathbf{P}\Big((\theta_n - \theta)\sqrt{\Sigma_{k=1}^n \phi_k^2} \leq x\sigma\Big) - \Phi(x)\Big| \leq C\,\Big(1 + x^2\Big)\,\epsilon\,|\log \epsilon| \exp\left\{-\frac{\breve{x}^2}{2}\right\}, \tag{34}$$

where $\breve{x}$ is defined by (13). Moreover, the following Berry-Esseen bound holds

$$\Big|\mathbf{P}\Big((\theta_n - \theta)\sqrt{\Sigma_{k=1}^n \phi_k^2} \leq x\sigma\Big) - \Phi(x)\Big| \leq C\,\epsilon\,|\log \epsilon|\,. \tag{35}$$

For all $0 \leq x = O\big(\epsilon^{-1/3}\big)$ as $\epsilon \to 0$, the following Cramér large deviation result holds

$$\frac{\mathbf{P}\Big((\theta_n - \theta)\sqrt{\Sigma_{k=1}^n \phi_k^2} \geq x\sigma\Big)}{1 - \Phi(x)} = 1 + \vartheta\, C\,\Big(1 + x^3\Big)\,\epsilon\,|\log \epsilon|\,, \tag{36}$$

where $|\vartheta| \leq 1$.

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 $a \leq \phi_k \leq b$ a.s. for two positive constants a and b and that ϵ_2 is a constant. In particular, if $(\varepsilon_k)_{k=1,...,n}$ is a sequence of independent random variables satisfying the Bernstein condition

$$\Big|\mathbf{E}[\varepsilon_k^l]\Big| \leq \frac{1}{2} l! C^{l-2} \mathbf{E}[\varepsilon_k^2] \ \text{ for all } l \geq 2 \text{ and all } 1 \leq k \leq n,$$

then the conditions (32) and (33) of Theorem 3.1 are satisfied with $\epsilon_1 = \frac{b}{a\sqrt{n}}$ and $\epsilon = O(\frac{1}{\sqrt{n}})$ as $n \to \infty$.

3.2. *Self-normalized deviations*

The self-normalized deviations have attracted lots of attentions due to the useful application to Student's t-statistic; we refer to Shao (1997, 1999) and Jing, Shao and Wang (2003). The first exponential nonuniform Berry-Esseen bound for the self-normalized mean for symmetric random variables $(\xi_i)_{i=1,...,n}$ has been established by Wang and Jing (1999) (cf. Theorem 2.1 therein): if $\mathbf{E}[|\xi_i|^3] < \infty$ for all $i \in [1, n]$, then

$$\begin{aligned} &\left| \mathbf{P}\Big(\frac{S_n}{\sqrt{[S]_n}} \leq x \Big) - \Phi(x) \right| \\ &\leq \begin{cases} C\Big(L_{3n}\Big(1+x^2\Big) + \sum_{i=1}^n \mathbf{P}\Big(|\xi_i| \geq B_n (6|x|)^{-1}\Big)\Big) \exp\Big\{-\frac{x^2}{2}\Big\}, & \\ & \text{if } |x| \leq \Big(5L_{3n}^{1/3}\Big)^{-1}, \\ \Big(1+\frac{1}{\sqrt{2\pi}|x|}\Big) \exp\Big\{-\frac{x^2}{2}\Big\}, & \text{if } |x| > \Big(5L_{3n}^{1/3}\Big)^{-1}, \end{cases} \end{aligned} \tag{37}$$

where $B_n^2 = \mathbf{E}[S_n^2]$, $L_{3n} = B_n^{-3} \sum_{i=1}^n \mathbf{E}[|\xi_i|^3]$ and $[S]_n = \sum_{i=1}^n \xi_i^2$ is the square bracket of S_n. Here, by convention, we assume $\frac{0}{0} = 0$. In the following theorem, we give a result similar to inequality (37) of Wang and Jing (1999) via a new method based on martingale theory. In particular, our result does not depend on the moments of random variables.

THEOREM 3.2 *Let $(\xi_i)_{i=1,...,n}$ be a sequence of non-degenerate, independent and symmetric random variables. If there exists a number $\epsilon \in (0, \frac{1}{2}]$, possibly depending on n, such that*

$$\frac{|\xi_i|}{\sqrt{[S]_n}} \leq \epsilon \qquad \text{a.s. for all } \ i \in [1, n], \tag{38}$$

then, for all $x \in \mathbf{R}$,

$$\Big|\mathbf{P}\Big(\frac{S_n}{\sqrt{[S]_n}} \leq x \Big) - \Phi(x) \Big| \leq C \, \epsilon \, |\log \epsilon| \, \Big(1 + x^2\Big) \exp\Big\{ -\frac{x^2}{2} \Big\}, \tag{39}$$

and, for all $x \geq 0$,

$$\frac{\mathbf{P}\Big(\frac{S_n}{\sqrt{[S]_n}} > x\Big)}{1-\Phi(x)} \;=\; 1+\vartheta\, C\,\Big(1+x^3\Big)\,\epsilon\,|\log \epsilon|, \tag{40}$$

where $|\vartheta| \leq 1$.

We continue with some comments and remarks on the obtained results.

(1) Condition (38) in Theorem 3.2 is satisfied when ξ_i are all bounded. For instance, if $a \leq |\xi_i| \leq b$ a.s. for two positive constants a and b, then it is obvious that the condition of Theorem 3.2 is satisfied with $\epsilon = \frac{b}{a\sqrt{n}} = O(\frac{1}{\sqrt{n}})$ as $n \to \infty$.

(2) We conjecture that when $(\xi_i)_{1\leq i \leq n}$ are symmetric and i.i.d. non-degenerate random variables with moments of order 3, the following results (similar to that of Fan, Grama and Liu (2015a)) hold: for all $0 \leq x = o(\sqrt{n})$,

$$\mathbf{P}\Big(\frac{S_n}{\sqrt{[S]_n}} > x\Big) \;=\; \inf_{\lambda \geq 0} \mathbf{E}\Big[e^{\lambda(\frac{S_n}{\sqrt{[S]_n}}-x)}\Big]\Big(1+O\Big(\frac{1+x}{\sqrt{n}}\Big)\Big), \tag{41}$$

and, for all $0 \leq x = o(n^{1/4})$, with $|\vartheta| \leq 1$,

$$\frac{\mathbf{P}\Big(\frac{S_n}{\sqrt{[S]_n}} > x\Big)}{1-\Phi(x)} \;=\; 1+\vartheta\, C\,\frac{1+x^3}{\sqrt{n}}. \tag{42}$$

(3) For more Cramér-type large deviation results on self-normalized sequences, we refer to Shao (1997) and Jing, Shao and Wang (2003). In particular, without assuming that $(\xi_i)_{i=1,...,n}$ are symmetric, equalities of type (42) in the range $0 \leq x = o(n^{1/6})$ have been established therein. This range of x is the best possible for non-symmetric random variables $(\xi_i)_{i=1,...,n}$.

4. Proofs of Theorems 2.1 - 2.3

In this section we prove the main results of the paper. We start with some auxiliary statements, then we prove Theorem 2.3 which in turn will be used in the proof of Theorem 2.1.

In the sequel, for simplicity, the equalities and inequalities involving random variables will be understood in the a.s. sense without mentioning this.

4.1. *Auxiliary results*

Let $(\xi_i, \mathcal{F}_i)_{i=0,...,n}$ be a sequence of martingale differences satisfying condition (A1) and $S = (S_k, \mathcal{F}_k)_{k=0,...,n}$ be the corresponding martingale defined by (14). For any real number λ with $|\lambda| < \epsilon^{-1}$, define the *exponential multiplicative martingale* $Z(\lambda) = (Z_k(\lambda), \mathcal{F}_k)_{k=0,...,n}$, where

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

For each $k = 1, ..., n$, the random variable $Z_k(\lambda)$ defines a probability density on $(\Omega, \mathcal{F}, \mathbf{P})$. This allows us to introduce, for $|\lambda| < \epsilon^{-1}$, the *conjugate probability measure* $\mathbf{P}_\lambda$ on $(\Omega, \mathcal{F})$ defined by

$$d\mathbf{P}_\lambda = Z_n(\lambda) d\mathbf{P}. \tag{43}$$

Denote by $\mathbf{E}_\lambda$ the expectation with respect to $\mathbf{P}_\lambda$. For all $i = 1, \ldots, n$, let

$$\eta_i(\lambda) = \xi_i - b_i(\lambda) \qquad \text{and} \qquad b_i(\lambda) = \mathbf{E}_\lambda[\xi_i|\mathcal{F}_{i-1}].$$

We thus obtain the following decomposition:

$$X_k = Y_k(\lambda) + B_k(\lambda), \qquad k = 1, ..., n, \tag{44}$$

where $Y(\lambda) = (Y_k(\lambda), \mathcal{F}_k)_{k=1,...,n}$ is the *conjugate martingale* defined as

$$Y_k(\lambda) = \sum_{i=1}^{k} \eta_i(\lambda), \qquad k = 1, ..., n, \tag{45}$$

and $B(\lambda) = (B_k(\lambda), \mathcal{F}_k)_{k=1,...,n}$ is the *drift process* defined as

$$B_k(\lambda) = \sum_{i=1}^{k} b_i(\lambda), \qquad k = 1, ..., n.$$

In the proofs of theorems, we shall make use of the following bounds of $B_n(\lambda)$.

LEMMA 4.1 *Assume conditions (A1) and (A2). Then, for all* $0 \leq \lambda < \epsilon^{-1}$,

$$\lambda - \lambda\delta^2 - C\lambda^2\epsilon \ \leq \ B_n(\lambda) \ \leq \ \frac{\lambda - 0.5\lambda^2\epsilon}{(1-\lambda\epsilon)^2}\left(1+\delta^2\right).$$

Proof. Since $\mathbf{E}[\xi_i|\mathcal{F}_{i-1}] = 0$ and $\lambda \geq 0$, it follows that

$$\mathbf{E}[\xi_i e^{\lambda\xi_i}|\mathcal{F}_{i-1}] = \mathbf{E}[\xi_i(e^{\lambda\xi_i} - 1)|\mathcal{F}_{i-1}] \geq 0$$

and, by Jensen's inequality, $\mathbf{E}[e^{\lambda\xi_i}|\mathcal{F}_{i-1}] \geq 1$. Using Taylor's expansion of e^x, we get

$$\begin{aligned} B_n(\lambda) &= \sum_{i=1}^{n} \frac{\mathbf{E}[\xi_i e^{\lambda\xi_i}|\mathcal{F}_{i-1}]}{\mathbf{E}[e^{\lambda\xi_i}|\mathcal{F}_{i-1}]} \\ &\leq \sum_{i=1}^{n} \mathbf{E}[\xi_i e^{\lambda\xi_i}|\mathcal{F}_{i-1}] \\ &= \lambda\langle S\rangle_n + \sum_{i=1}^{n}\sum_{k=2}^{+\infty} \mathbf{E}\left[\frac{\xi_i(\lambda\xi_i)^k}{k!}\middle|\mathcal{F}_{i-1}\right]. \end{aligned} \tag{46}$$

By condition (A1), for all $0 \le \lambda < \epsilon^{-1}$, we deduce

$$\begin{aligned} \sum_{i=1}^{n} \sum_{k=2}^{+\infty} \left| \mathbf{E}\left[\frac{\xi_i(\lambda\xi_i)^k}{k!} \middle| \mathcal{F}_{i-1} \right] \right| &= \sum_{i=1}^{n} \sum_{k=2}^{+\infty} |\mathbf{E}[\xi_i^{k+1}|\mathcal{F}_{i-1}]| \frac{\lambda^k}{k!} \\ &\le \frac{1}{2}\, \lambda^2 \epsilon \, \langle S \rangle_n \sum_{k=2}^{+\infty} (k+1)(\lambda\epsilon)^{k-2} \\ &= \frac{(3-2\lambda\epsilon)}{2(1-\lambda\epsilon)^2}\, \lambda^2 \epsilon \, \langle S \rangle_n. \end{aligned} \tag{47}$$

Inequalities (47) and (46) imply that, for all $0 \le \lambda < \epsilon^{-1}$,

$$B_n(\lambda) \le \lambda \langle S \rangle_n + \frac{(3-2\lambda\epsilon)}{2(1-\lambda\epsilon)^2}\, \lambda^2 \epsilon \langle S \rangle_n = \frac{\lambda - 0.5\lambda^2\epsilon}{(1-\lambda\epsilon)^2} \langle S \rangle_n.$$

This bound together with condition (A2) gives the upper bound of $B_n(\lambda)$. The lower bound of $B_n(\lambda)$ can be found in Lemma 4.2 of [13]. ■

Consider the predictable process $\Psi(\lambda) = (\Psi_k(\lambda), \mathcal{F}_k)_{k=0,...,n}$, where

$$\Psi_k(\lambda) = \sum_{i=1}^{k} \log \mathbf{E}[e^{\lambda \xi_i} | \mathcal{F}_{i-1}]. \tag{48}$$

We have the following elementary bound.

LEMMA 4.2 *Assume conditions (A1) and (A2). Then, for all* $0 \le \lambda < \epsilon^{-1}$,

$$\Psi_n(\lambda) \le \frac{\lambda^2(1+\delta^2)}{2(1-\lambda\epsilon)}. \tag{49}$$

Proof. Since $\log(1+t) \le t$ for all $t \ge 0$, we have, for all $0 \le \lambda < \epsilon^{-1}$,

$$\Psi_n(\lambda) = \sum_{i=1}^{n} \log \Big(1 + \mathbf{E}[e^{\lambda\xi_i}|\mathcal{F}_{i-1}] - 1 \Big) \le \sum_{i=1}^{n} \Big(\mathbf{E}[e^{\lambda\xi_i}|\mathcal{F}_{i-1}] - 1 \Big).$$

By condition (A1), it is easy to see that, for all $0 \le \lambda < \epsilon^{-1}$,

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

Thus, we obtain, for all $0 \leq \lambda < \epsilon^{-1}$,

$$\Psi_n(\lambda) \leq \frac{\lambda^2 \langle S \rangle_n}{2(1-\lambda\epsilon)}.$$

This inequality together with condition (A2) gives inequality (49). ■

4.2. *Proof of Theorem 2.3*

For all $0 \leq \lambda < \epsilon^{-1}$ and $x \geq 0$, by the definition of the conjugate measure (43), we have the following representation:

$$\begin{aligned} \mathbf{P}(S_n > x) &= \mathbf{E}_\lambda\left[Z_n(\lambda)^{-1}\mathbf{1}_{\{S_n>x\}}\right] \\ &= \mathbf{E}_\lambda\left[\exp\left\{-\lambda S_n + \Psi_n(\lambda)\right\}\mathbf{1}_{\{S_n>x\}}\right] \\ &= \mathbf{E}_\lambda\Big[\exp\left\{-\lambda x + \Psi_n(\lambda) - \lambda(Y_n(\lambda)+B_n(\lambda)-x)\right\} \\ &\qquad \times \mathbf{1}_{\{Y_n(\lambda)+B_n(\lambda)-x>0\}}\Big]. \end{aligned}$$

Setting $U_n(\lambda) = \lambda(Y_n(\lambda)+B_n(\lambda)-x)$, we deduce, for all $0 \leq \lambda < \epsilon^{-1}$,

$$\begin{aligned} \mathbf{P}(S_n > x) &\leq \exp\left\{-\lambda x + \frac{\lambda^2(1+\delta^2)}{2(1-\lambda\epsilon)}\right\}\mathbf{E}_\lambda\left[e^{-U_n(\lambda)}\mathbf{1}_{\{U_n(\lambda)>0\}}\right] \\ &= \exp\left\{-\lambda x + \frac{\lambda^2(1+\delta^2)}{2(1-\lambda\epsilon)}\right\}\int_0^\infty e^{-t}\mathbf{P}_\lambda(0 < U_n(\lambda) \leq t)dt. \end{aligned} \tag{50}$$

To optimize the term in the last exponent, let $\overline{\lambda} = \overline{\lambda}(x) \in [0, \epsilon^{-1})$ be the unique solution of the equation

$$-x + \frac{d}{d\lambda}\Big(\frac{\lambda^2(1+\delta^2)}{2(1-\lambda\epsilon)}\Big) = 0, \quad \text{or equivalently} \quad \frac{\lambda - 0.5\lambda^2\epsilon}{(1-\lambda\epsilon)^2} = \frac{x}{1+\delta^2}. \tag{51}$$

The definition of $\overline{\lambda}$ and Lemma 4.1 imply that, for all $x \geq 0$,

$$\overline{\lambda} = \frac{2x/(1+\delta^2)}{1 + 2x\epsilon/(1+\delta^2) + \sqrt{1+2x\epsilon/(1+\delta^2)}}. \tag{52}$$

Taking $\lambda = \overline{\lambda}$ in (50), we get

$$\mathbf{P}(S_n > x) \leq \exp\left\{-\frac{\hat{x}^2}{2}\right\}\int_0^\infty e^{-t}\mathbf{P}_{\overline{\lambda}}(0 < U_n(\overline{\lambda}) \leq t)dt, \tag{53}$$

where

$$\hat{x} = \frac{\overline{\lambda}\sqrt{1+\delta^2}}{1-\overline{\lambda}\epsilon} = \frac{2x/\sqrt{1+\delta^2}}{1+\sqrt{1+2x\epsilon/(1+\delta^2)}}. \tag{54}$$

To bound the integral term of (53), we make use of the following lemma, which gives a rate of convergence in the central limit theorem for the process $U(\overline{\lambda})$ under the conjugate probability measure $\mathbf{P}_{\overline{\lambda}}$. The proof of this lemma is given in Appendix A.

LEMMA 4.3 *Assume conditions (A1) and (A2). Then, for all* $0 \leq \overline{\lambda} < \epsilon^{-1}$,

$$\sup_{u \in \mathbf{R}} \Big| \mathbf{P}_{\overline{\lambda}}(U_n(\overline{\lambda}) \leq \widehat{x} u) - \Phi(u) \Big| \leq C \Big(\overline{\lambda}^2 \epsilon + \overline{\lambda} \delta^2 + \epsilon \left| \log \epsilon \right| + \delta \Big).$$

Now we return to the proof of Theorem 2.3. From (53), using Lemma 4.3, it follows that, with $\mathcal{N}(0,1)$ the standard normal random variable,

$$\begin{aligned} \int_0^\infty e^{-t} \mathbf{P}_{\overline{\lambda}}(0 < U_n(\overline{\lambda}) \leq t) dt &= \int_0^\infty \exp\{-\widehat{x} t\} \, \mathbf{P}_{\overline{\lambda}}(0 < U_n(\overline{\lambda}) \leq \widehat{x} t) \widehat{x} dt \\ &\leq \int_0^\infty \exp\{-\widehat{x} t\} \, \mathbf{P}_{\overline{\lambda}}(0 < \mathcal{N}(0,1) \leq t) \widehat{x} dt \\ &\quad + C \left(\overline{\lambda}^2 \epsilon + \overline{\lambda} \delta^2 + \epsilon \left| \log \epsilon \right| + \delta \right) \\ &\leq \exp\left\{ \frac{\widehat{x}^2}{2} \right\} \Big(1 - \Phi(\widehat{x}) \Big) \\ &\quad + C \left(\overline{\lambda}^2 \epsilon + \overline{\lambda} \delta^2 + \epsilon \left| \log \epsilon \right| + \delta \right). \end{aligned} \tag{55}$$

Combining (53) and (55) together, we have, for all $x \geq 0$,

$$\mathbf{P}(S_n > x) \leq 1 - \Phi(\widehat{x}) + C \exp\left\{ -\frac{\widehat{x}^2}{2} \right\} \Big(\overline{\lambda}^2 \epsilon + \overline{\lambda} \delta^2 + \epsilon \left| \log \epsilon \right| + \delta \Big).$$

Using the two sided bound (19), we obtain, for all $x \geq 0$,

$$\mathbf{P}(S_n > x) \leq \Big(1 - \Phi(\widehat{x}) \Big) \Big[1 + C \Big(1 + \widehat{x} \Big) \Big(\overline{\lambda}^2 \epsilon + \overline{\lambda} \delta^2 + \epsilon \left| \log \epsilon \right| + \delta \Big) \Big].$$

This completes the proof of Theorem 2.3. ■

4.3. *Proof of Theorem 2.1*

We firstly prove that (39) holds for all $x \geq 0$. From (56) and the fact $\mathbf{P}(S_n > x) = 1 - \mathbf{P}(S_n \leq x)$, it follows that, for all $x \geq 0$,

$$\begin{aligned}
\Phi(x) - \mathbf{P}(S_n \leq x) &= \mathbf{P}(S_n > x) - \Big(1 - \Phi(x)\Big) \\
&= \Phi(x) - \Phi(\widehat{x}) + \mathbf{P}(S_n > x) - \Big(1 - \Phi(\widehat{x})\Big) \\
&\leq \Phi(x) - \Phi(\widehat{x}) \\
&\quad + C_1 \Big(1 - \Phi(\widehat{x})\Big)\Big(1 + \widehat{x}\Big)\Big(\overline{\lambda}^2 \epsilon + \overline{\lambda}\delta^2 + \epsilon|\log \epsilon| + \delta\Big) \\
&\leq \frac{1}{\sqrt{2\pi}} |x - \widehat{x}| \exp\left\{-\frac{\widehat{x}^2}{2}\right\} \qquad (56) \\
&\quad + C_2 \Big(x^2 \epsilon + x\delta^2 + \epsilon|\log \epsilon| + \delta\Big) \exp\left\{-\frac{\widehat{x}^2}{2}\right\},
\end{aligned}$$

where the last line follows from $\overline{\lambda} \leq x$. From the definitions of $\overline{\lambda}$ and $\widehat{x}$ (cf. (51) and (54)), we deduce that

$$\begin{aligned}
|x - \widehat{x}| &= \frac{\overline{\lambda}(1 - 0.5\overline{\lambda}\epsilon)(1 + \delta^2)}{(1 - \overline{\lambda}\epsilon)^2} - \frac{\overline{\lambda}\sqrt{1 + \delta^2}}{1 - \overline{\lambda}\epsilon} \\
&= \frac{\overline{\lambda}(1 + \delta^2 - \sqrt{1 + \delta^2}\,)}{(1 - \overline{\lambda}\epsilon)^2} + \frac{\overline{\lambda}^2 \epsilon \left(\sqrt{1 + \delta^2} - 0.5(1 + \delta^2)\right)}{(1 - \overline{\lambda}\epsilon)^2} \\
&\leq \frac{\overline{\lambda}\delta^2}{(1 - \overline{\lambda}\epsilon)^2} + \frac{\overline{\lambda}^2 \epsilon}{(1 - \overline{\lambda}\epsilon)^2} \\
&\leq C\, x \Big(\delta^2 + x\epsilon\Big). \qquad (57)
\end{aligned}$$

Combining (56) and (57) together, we have, for all $x \geq 0$,

$$\Phi(x) - \mathbf{P}(S_n \leq x) \leq C\Big(1 + x^2\Big)\left(\epsilon|\log \epsilon| + \frac{\delta}{1 + x}\right) \exp\left\{-\frac{\widehat{x}^2}{2}\right\}. \qquad (58)$$

To prove the lower bound of $\Phi(x) - \mathbf{P}(S_n \leq x)$, we shall use the following Cramér large deviation expansion: for all $0 \leq x \leq \min\{\epsilon^{-1/3}, \delta^{-1}\}$,

$$\frac{\mathbf{P}(S_n > x)}{1 - \Phi(x)} = 1 + \theta C \left(\Big(1 + x\Big)\Big(\epsilon\,|\log \epsilon| + \delta\Big) + x^3 \epsilon\right), \qquad (59)$$

where $|\theta| \leq 1$. This Cramér large deviation expansion is a simple consequence of Corollary 2.1 of [13]. Using the equality $\mathbf{P}(S_n > x) = 1 - \mathbf{P}(S_n \leq x)$ and the two-sides bound (19), we get, for all $0 \leq x \leq \min\{\epsilon^{-1/3}, \delta^{-1}\}$,

$$\begin{aligned}\Phi(x) - \mathbf{P}(S_n \leq x) &\geq -C_1 \Big(1 - \Phi(x)\Big)\Big(\Big(1 + x\Big)\Big(\epsilon\,|\log\epsilon| + \delta\Big) + x^3\epsilon\Big) \\ &\geq -C_2 \Big(\epsilon\,|\log\epsilon| + \delta + x^2\epsilon\Big) \exp\Big\{-\frac{x^2}{2}\Big\} \\ &\geq -C_2 \Big(1 + x^2\Big)\Big(\epsilon\,|\log\epsilon| + \frac{\delta}{1+x}\Big) \exp\Big\{-\frac{x^2}{2}\Big\}. \end{aligned} \tag{60}$$

For all $x \geq \min\{\epsilon^{-1/3}, \delta^{-1}\}$, we have $\frac{1}{1+\widehat{x}} \leq C\,(x^2\epsilon + x\delta)$. Then, from (29), it is easy to see that, for all $x \geq \min\{\epsilon^{-1/3}, \delta^{-1}\}$,

$$\mathbf{P}(S_n > x) \leq C\Big(1 + x^2\Big)\Big(\epsilon\,|\log\epsilon| + \frac{\delta}{1+x}\Big) \exp\Big\{-\frac{\widehat{x}^2}{2}\Big\}. \tag{61}$$

Using (19) and the bound $x \geq \min\{\epsilon^{-1/3}, \delta^{-1}\}$, the same upper bound is obtained for $1 - \Phi(x)$. Therefore, we have, for all $x \geq \min\{\epsilon^{-1/3}, \delta^{-1}\}$,

$$\Phi(x) - \mathbf{P}(S_n \leq x) \geq -C\Big(1 + x^2\Big)\Big(\epsilon|\log\epsilon| + \frac{\delta}{1+x}\Big) \exp\Big\{-\frac{\widehat{x}^2}{2}\Big\}. \tag{62}$$

Combining (58), (60) and (62) together, we obtain, for all $x \geq 0$,

$$\Big|\mathbf{P}(S_n \leq x) - \Phi(x)\Big| \leq C\Big(1 + x^2\Big)\Big(\epsilon|\log\epsilon| + \frac{\delta}{1+x}\Big) \exp\Big\{-\frac{\widehat{x}^2}{2}\Big\}.$$

Notice that the same argument is applied to $-S_n$. Thus, for all $x < 0$,

$$\begin{aligned}\Big|\mathbf{P}(S_n \leq x) - \Phi(x)\Big| &= \Big|\mathbf{P}(-S_n \geq -x) - \Phi(x)\Big| \\ &= \Big|\mathbf{P}(-S_n \geq -x) - 1 - \Big(\Phi(x) - 1\Big)\Big| \\ &= \Big|\mathbf{P}(-S_n < -x) + \Phi(x) - 1\Big| \\ &= \Big|\mathbf{P}(-S_n < -x) - \Phi(-x)\Big| \\ &\leq C\Big(1 + x^2\Big)\Big(\epsilon|\log\epsilon| + \frac{\delta}{1+|x|}\Big) \exp\Big\{-\frac{\widehat{x}^2}{2}\Big\}. \end{aligned} \tag{63}$$

This completes the proof of Theorem 2.1. ∎

Proof of Corollary 2.2. Following Bolthausen (1982) we consider the stopping time $\tau = \sup\{0 \leq k \leq n : \langle S\rangle_k \leq 1\}$. Let $r = \lfloor(1 - \langle S\rangle_\tau)/\epsilon^2\rfloor$, where $\lfloor x\rfloor$ is the largest integer less than x. Then $r \leq \lfloor 1/\epsilon^2\rfloor$. Let $N = n + \lfloor 1/\epsilon^2\rfloor + 1$. Consider a sequence of independent Rademacher random variables (η_i) (taking values $+1$ and -1 with equal probabilities) which is also independent of the martingale differences (ξ_i). For each $i = 1, \ldots, N$ define $\xi_i' = \xi_i$ if $i \leq \tau$, $\xi_i' = \varepsilon\eta_i$ if $\tau < i \leq \tau + r$, $\xi_i' = (1 - \langle S\rangle_\tau - r\epsilon^2)^{1/2}$ if $i = \tau + r + 1$, and $\xi_i' = 0$ if $\tau + r + 1 < i \leq N$. Clearly, $S_k' = \sum_{i=1}^k \xi_i'$, $k = 0, \ldots, N$ (with $S_0' = 0$) is a martingale sequence w.r.t. the enlarged probability space and the enlarged filtration.

Moreover $\langle S' \rangle_N = 1$ and condition (A1) is satisfied for $(S'_k)_{k=1,\dots,N}$. By Theorem 2.1, it holds, for all $x \in \mathbf{R}$,

$$\Big| \mathbf{P}(S'_N \leq x) - \Phi(x) \Big| \leq C \, \Big(1 + x^2\Big) \, \epsilon |\log \epsilon| \exp\left\{ -\frac{\widehat{x}^2}{2} \right\}, \tag{64}$$

where $\widehat{x} = \frac{2|x|}{1+\sqrt{1+2|x|\epsilon}}$. Since $\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] \leq 12\epsilon^2$ for all i (cf. the proof of Lemma 4.1 in [13]), it holds

$$1 - 12\epsilon^2 \leq 1 - \mathbf{E}[\xi_\tau^2|\mathcal{F}_{\tau-1}] \leq \langle S \rangle_\tau \leq 1.$$

Then it is easy to see that

$$\mathbf{E}[(S'_N - S_n)^2] \leq C \left(\mathbf{E}|\langle S \rangle_n - 1| + \epsilon^2 \right); \tag{65}$$

cf. Mourrat [33] for more details. For all $x \geq 0$ and any $t > 0$, it holds

$$\begin{aligned}
\mathbf{P}(S_n \leq x) &\leq \mathbf{P}(S_n \leq x, \ |S_n - S'_N| \leq t) + \mathbf{P}(|S_n - S'_N| > t) & (66)\\
&\leq \mathbf{P}(S'_N \leq x + t) + \frac{1}{t^2}\mathbf{E}[|S_n - S'_N|^2] & (67)\\
&\leq \Phi(x+t) + C_1 \Big(1 + x^2\Big) \epsilon |\log \epsilon| \exp\left\{ -\frac{\widehat{x}^2}{2} \right\} \\
&\quad + \frac{C_2}{t^2} \Big(\mathbf{E}|\langle S \rangle_n - 1| + \epsilon^2 \Big) \\
&\leq \Phi(x) + \frac{t}{\sqrt{2\pi}} e^{-x^2/2} \\
&\quad + C_1 \Big(1 + x^2\Big) \epsilon |\log \epsilon| \exp\left\{ -\frac{\widehat{x}^2}{2} \right\} + \frac{C_2}{t^2} \Big(\mathbf{E}|\langle S \rangle_n - 1| + \epsilon^2 \Big),
\end{aligned}$$

where $\widehat{x}$ is defined by (13). Now putting $t = (\mathbf{E}|\langle S \rangle_n - 1| + \epsilon^2) e^{x^2/6}$, one has, for all $x \geq 0$,

$$\begin{aligned}
\mathbf{P}(S_n \leq x) - \Phi(x) \ \leq \ & C \Bigg(\Big(1 + x^2\Big) \epsilon |\log \epsilon| \exp\left\{ -\frac{\widehat{x}^2}{2} \right\} \\
& + \Big(\mathbf{E}|\langle S \rangle_n - 1| + \epsilon^2 \Big)^{1/3} \exp\left\{ -\frac{x^2}{6} \right\} \Bigg).
\end{aligned}$$

It is easy to see that the same bound holds for $\Phi(x) - \mathbf{P}(S_n \leq x)$. Therefore we obtain (23) for all $x \geq 0$. If $x \leq 0$, we consider $-S_n$ instead of S_n and we use the fact that $\mathbf{P}(-S_n \leq -x) = \mathbf{P}(S_n \geq x) = 1 - \mathbf{P}(S < x)$. This implies (23) for all $x \leq 0$, which completes the proof of Corollary 2.2. ■

5. Proofs of Theorems 3.1 and 3.2

The proofs of Theorems 3.1 and 3.2 are based on Theorem 2.1.

Proof of Theorem 3.1. From (30) and (31), it is easy to see that

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

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

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

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

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

Notice that

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

and

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

Applying Theorem 2.1 to $(\xi_i, \mathcal{F}_i')_{i=1,...,n}$, we obtain the required inequalities. ∎

Proof of Theorem 3.2. Let $\mathcal{F}_0 = \sigma\Big(\xi_j^2, 1 \leq j \leq n\Big)$, and set, for all $i = 1, ..., n$,

$$\eta_i = \frac{\xi_i}{\sqrt{[S]_n}} \quad \text{and} \quad \mathcal{F}_i = \sigma\Big(\xi_k, 1 \leq k \leq i,\ \xi_j^2, 1 \leq j \leq n\Big). \tag{68}$$

Since $(\xi_i)_{i=1,...,n}$ is a sequence of independent and symmetric random variables, we deduce that

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

Thus $\big(\eta_i, \mathcal{F}_i\big)_{i=1,...,n}$ is a sequence of conditionally symmetric martingale differences, i.e. $\mathbf{E}[\eta_i > y |\, \mathcal{F}_{i-1}] = \mathbf{E}[-\eta_i > y |\, \mathcal{F}_{i-1}]$, and satisfies $\mathbf{P}(|\eta_i| \leq \epsilon | \mathcal{F}_0) = 1$ by assumption. Note that $S_n/\sqrt{[S]_n} = \sum_{i=1}^n \eta_i$ and $\sum_{i=1}^n \mathbf{E}[\eta_i^2|\mathcal{F}_{i-1}] = \sum_{i=1}^n \eta_i^2 = 1$. By the fact that η_i is conditionally symmetric w.r.t. $\mathcal{F}_{i-1}$, we have, for all $1 \leq i \leq n$ and all $\lambda \geq 0$,

$$\mathbf{E}\Big[\exp\{\lambda \eta_i\}\Big|\mathcal{F}_{i-1}\Big] = \frac{1}{2}\,\mathbf{E}\Big[\exp\{\lambda \eta_i\} + \exp\{-\lambda \eta_i\}\Big|\mathcal{F}_{i-1}\Big].$$

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

$$\mathbf{E}\Big[\exp\{\lambda \eta_i\}\Big|\mathcal{F}_{i-1}\Big] \leq \exp\left\{\frac{\lambda^2 \eta_i^2}{2}\right\}.$$

Thus it holds, for all $1 \leq k \leq n$,

$$\Psi_k(\lambda) \ \leq \ \frac{\lambda^2}{2}\sum_{i=1}^k \eta_i^2 \ \leq \ \frac{\lambda^2}{2}, \tag{69}$$

which improves Lemma 4.2. With this improvement and a proof similar to that of Theorem 2.1, we obtain the required inequality. ■

Appendix A. Proof of Lemma 4.3

To complete the proof of Theorems 2.3, we need to prove Lemma 4.3. We will make use of the following assertion which gives a rate of convergence in the central limit theorem for the martingale $Y(\lambda)$ under the conjugate probability measure $\mathbf{P}_\lambda$.

LEMMA A.1 *Assume conditions (A1) and (A2). Then, for all* $0 \leq \lambda < \epsilon^{-1}$,

$$\sup_{u \in \mathbf{R}} \Big|\mathbf{P}_\lambda(Y_n(\lambda) \leq u) - \Phi(u)\Big| \ \leq \ C\Big(\lambda\epsilon + \epsilon\,|\log \epsilon| + \delta\Big).$$

This assertion is proved in [13], Lemma 3.1 (for an earlier result for bounded martingale differences see Lemma 3.3 of [22]). Next we use Lemma A.1 to prove Lemma 4.3.

Proof of Lemma 4.3. Notice that $|B_n(\overline{\lambda}) - x| \leq C\overline{\lambda}^2\epsilon + \overline{\lambda}\delta^2$. Thus

$$\sup_{u \in \mathbf{R}} \Big|\Phi\Big(\frac{\widehat{x}u - \overline{\lambda}(B_n(\overline{\lambda}) - x)}{\overline{\lambda}}\Big) - \Phi(u)\Big| \ \leq \ C\Big(\overline{\lambda}^2\epsilon + \overline{\lambda}\delta^2\Big).$$

By Lemma A.1, it is easy to see that

$$\begin{aligned}
& \sup_{u\in\mathbf{R}} \Big|\mathbf{P}_{\overline{\lambda}}(U_n(\overline{\lambda}) \leq \widehat{x}u) - \Phi(u)\Big| \qquad \text{(A1)} \\
\leq\ & \sup_{u\in\mathbf{R}} \Big|\mathbf{P}_{\overline{\lambda}}\Big(Y_n(\overline{\lambda}) \leq \frac{\widehat{x}u - \overline{\lambda}(B_n(\overline{\lambda}) - x)}{\overline{\lambda}}\Big) - \Phi\Big(\frac{\widehat{x}u - \overline{\lambda}(B_n(\overline{\lambda}) - x)}{\overline{\lambda}}\Big)\Big| \\
& + \sup_{u\in\mathbf{R}} \Big|\Phi\Big(\frac{\widehat{x}u - \overline{\lambda}(B_n(\overline{\lambda}) - x)}{\overline{\lambda}}\Big) - \Phi(u)\Big| \\
\leq\ & C_1\Big(\overline{\lambda}\epsilon + \epsilon\,|\log \epsilon| + \delta\Big) + C_2\Big(\overline{\lambda}^2\epsilon + \overline{\lambda}\delta^2\Big) \\
\leq\ & C\Big(\overline{\lambda}^2\epsilon + \overline{\lambda}\delta^2 + \epsilon\,|\log \epsilon| + \delta\Big).
\end{aligned}$$

This completes the proof. ∎

Acknowledgements

The authors are grateful to the reviewers for their comments and remarks which helped to improve the manuscript. Fan and Liu have been partially supported by the National Natural Science Foundation of China (Grants no. 11601375, no. 11501146, no. 11571052 and no. 11401590).

References

[1] Bercu, B., Delyon, B. and Rio, E. (2015). *Concentration inequalities for sums and martingales. Springer Briefs in Mathematics. Springer, New York.*

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

[3] Bennett, G. (1962). Probability inequalities for the sum of independent random variables, *J. Amer. Statist. Assoc.* **57**, No. 297, 33–45.

[4] Bernstein, S. (1927). *Theory of Probability.* Moscow (in Russian).

[5] Bikelis, A. (1966). Estimates of the remainder in the central limit theorem, *Litovsk. Mat. Sb.* **6**(3), 323–46.

[6] Bolthausen, E. (1982). Exact convergence rates in some martingale central limit theorems. *Ann. Probab.* **10**, 672–688.

[7] Bose, A. (1986a). Certain non-uniform rates of convergence to normality for martingale differences. *J. Statist. Plann. Inference* **14**, 155–167.

[8] Bose, A. (1986b). Certain non-uniform rates of convergence to normality for a restricted class of martingales. *Stochastics* **16**, 279–294.

[9] Chen, L. H. Y. and Shao, Q. M. (2001). A non-uniform Berry-Esseen bound via Stein's method. *Probab. Theory Relat. Fields* **120**, 236–254.

[10] Cramér, H. (1938). Sur un nouveau théorème-limite de la théorie des probabilités, *Actualite's Sci. Indust.* **736**, 5–23.

[11] de la Peña, V. H. (1999). A general class of exponential inequalities for martingales and ratios, *Ann. Probab.* **27**, No. 1, 537–564.

[12] Esseen, C. G. (1945). Fourier analysis of distribution functions: a mathematical study of the Laplace-Gaussian law, *Acta Math.* **77**, 1–125.

[13] Fan, X., Grama, I. and Liu, Q. (2013). Cramér large deviation expansions for martingales under Bernstein's condition, *Stochastic Process. Appl.* **123**, 3919–3942.

[14] Fan, X., Grama, I. and Liu, Q. (2015a). Sharp large deviation results for sums of independent random variables. *Sci. China Math.* **58**, 1939–1958.

[15] Fan, X., Grama, I. and Liu, Q. (2015b). Exponential inequalities for martingales with applications. *Electron. J. Probab.* **20**, no. 1, 1–22.

[16] Grama I. (1987a) On the improved rate of convergence in the CLT for semimartingales. Stokhasticeskii Analiz. Matematiceskie Issledovania., No 97, pp. 34–40, Kishinev: Stiinza (in Russian).

[17] Grama I. (1987b) Normal approximation for semimartingales. Uspehi Matematicheskih Nauk, Vol. 42, N 6, pp. 169–170 (in Russian).

[18] Grama I. (1988) Rates of normal approximation for semimartingales. Ph. D. Thesis, V.A.Steklov Mathematical Institute (MIAN), Moscow, 138 pp. (in Russian).

[19] Grama, I. (1993). On the rate of convergence in the central limit theorem for d-dimentional semimartinagles. *Stochastic Stochastic Rep.* **44**(3-4), 131–152.

[20] Grama, I. (1995). The probabilities of large deviations for semimartingales. *Stochastic Stochastic Rep.* **54**(1-2), 1–19.

[21] Grama, I. (1997). On moderate deviations for martingales. *Ann. Probab.* **25**, 152–184.

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

[23] Grama, I. and Haeusler, E. (2006). An asymptotic expansion for probabilities of moderate deviations for multivariate martingales. *J. Theoret. Probab.* **19**, 1–44.

[24] Haeusler, E., (1988a). On the rate of convergence in the central limit theorem for martingales with discrete and continuous time. *Ann. Probab.* **16**, No. 1, 275–299.

[25] Haeusler, E., Joos, K., (1988b). A nonuniform bound on the rate of convergence in the martingale central limit theorem. *Ann. Probab.* **16**, No. 4, 1699–1720.

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

[27] Hall, P. and Heyde, C. C. (1981). Rates of convergence in the martingale central limit theorem. *Ann. Probab.* **9**(3), 395–404.

[28] Heyde, C. C. and Brown, B. M. (1970). On the departure from nomality of a certain class of martingales. *Ann. Math. Statist.* **41**, No. 6, 2161–2165.

[29] Hitczenko, P. (1990). Upper bounds for the L_p-norms of martingales. *Probab. Theory Related Fields* **86**, 225–238.

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

[31] Joos, K. (1991). Nonuniform bounds on the rate of convergence in the central limit theorem for martingales. *J. Multivariate Anal.* **36**, No. 2, 297–313.

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

[33] Mourrat, J. C. (2013). On the rate of convergence in the martingale central limit theorem. *Bernoulli* **19**(2), 633–645.

[34] Petrov, V. V. (1975). *Sums of independent random variables.* Springer-Verlag, Berlin, Heidelberg, New York.

[35] Račkauskas, A. (1990). On probabilities of large deviations for martingales. *Liet. Mat. Rink.* **30**, No. 4, 784–795.

[36] Račkauskas, A. (1995). Large deviations for martingales with some applications. *Acta Appl. Math.* **38**, 109-129.

[37] Renz, J. (1996). A note on exact convergence rates in some martingale central limit theorems. *Ann. Probab.* **24**(3), 1616–1637.

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

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

[40] Shorack, G. R., Wellner, J., (1986). *Emprical processes with application to statistics.* Wiley, New York.

[41] van de Geer, S. (1995). Exponential inequalities for martingales, with application to maxi-

mum likelihood estimation for counting process. *Ann. Statist.* **23**, 1779–1801.
[42] Wang, Q. Y., Jing, B. Y. (1999). An exponential nonuniform Berry-Esseen bound for self-normalized sums, *Ann. Probab.* **27**, No. 4, 2068–2088.