Self-normalized Cramér type moderate deviations for martingales

XIEQUAN FAN[1], ION GRAMA[2,*], QUANSHENG LIU[2,**] and QI-MAN SHAO[3]

[1]*Center for Applied Mathematics, Tianjin University, Tianjin 300072, China.*
E-mail: fanxiequan@hotmail.com
[2]*Université de Bretagne-Sud, LMBA, UMR CNRS 6205, Campus de Tohannic, 56017 Vannes, France.*
E-mail: *ion.grama@univ-ubs.fr; **quansheng.liu@univ-ubs.fr
[3]*Department of Statistics, The Chinese University of Hong Kong, Shatin, NT, Hong Kong.*
E-mail: qmshao@sta.cuhk.edu.hk

Let $(X_i, \mathcal{F}_i)_{i \geq 1}$ be a sequence of martingale differences. Set $S_n = \sum_{i=1}^n X_i$ and $[S]_n = \sum_{i=1}^n X_i^2$. We prove a Cramér type moderate deviation expansion for $\mathbf{P}(S_n/\sqrt{[S]_n} \geq x)$ as $n \to +\infty$. Our results partly extend the earlier work of Jing, Shao and Wang (*Ann. Probab.* **31** (2003) 2167–2215) for independent random variables.

Keywords: Cramér's moderate deviations; martingales; self-normalized sequences

1. Introduction

Let $(X_i)_{i \geq 1}$ be a sequence of independent random variables with zero means and finite variances: $\mathbf{E}X_i = 0$ and $0 < \mathbf{E}X_i^2 < \infty$ for all $i \geq 1$. Set

$$S_n = \sum_{i=1}^n X_i, \qquad B_n^2 = \sum_{i=1}^n \mathbf{E}X_i^2, \qquad V_n^2 = \sum_{i=1}^n X_i^2.$$

It is well known that under the Lindeberg condition the central limit theorem (CLT) holds

$$\sup_{x \in \mathbf{R}} \left| \mathbf{P}(S_n/B_n \leq x) - \Phi(x) \right| \to 0 \quad \text{as } n \to \infty,$$

where $\Phi(x)$ denotes the standard normal distribution function. Cramér's moderate deviation expansion stated below gives an estimation of the relative error of $\mathbf{P}(S_n/B_n \geq x)$ to $1 - \Phi(x)$. If $(X_i)_{i \geq 1}$ are identically distributed with $\mathbf{E}e^{t_0\sqrt{|X_1|}} < \infty$ for some $t_0 > 0$ (cf. Linnik [22]), then for $0 \leq x = o(n^{1/6})$ as $n \to \infty$,

$$\frac{\mathbf{P}(S_n/B_n \geq x)}{1 - \Phi(x)} = 1 + o(1) \quad \text{and} \quad \frac{\mathbf{P}(S_n/B_n \leq -x)}{\Phi(-x)} = 1 + o(1). \tag{1.1}$$

Expansion is available for $0 \leq x = o(n^{1/2})$ if the moment generating function exists. We refer to Chapter VIII of Petrov [25] for further details on the subject.

However, the limit theorems for self-normalized partial sums of independent random variables have put a new countenance on the classical limit theorems. The study of self-normalized partial sums S_n/V_n originates from Student's t-statistic. Student's t-statistic T_n is defined by

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

where

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

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

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

see Chung [7]. So, if we get an asymptotic bound on the tail probabilities for self-normalized partial sums, then we have an asymptotic bound on the tail probabilities for T_n. Giné, Götze and Mason [14] gave a necessary and sufficient condition for the asymptotic normality. Slavova [30] and Bentkus, Bloznelis and Götze [1] (see also Bentkus and Götze [2]) obtained the Berry–Esseen bounds for self-normalized partial sums. See also Novak [24] and Shao and Wang [28] for Berry–Esseen type inequalities with explicit constants. Shao [26] established a self-normalized Cramér-Chernoff large deviation without any moment assumptions and Shao [27] proved a self-normalized Cramér moderate deviation theorem under $(2+\rho)$th moments: if $(X_i)_{i\geq 1}$ are independent and identically distributed with $\mathbf{E}|X_1|^{2+\rho} < \infty$, $\rho \in (0, 1]$, then for $0 \leq x = o(n^{\rho/(4+2\rho)})$ as $n \to \infty$,

$$\frac{\mathbf{P}(S_n/V_n \geq x)}{1-\Phi(x)} = 1 + o(1). \tag{1.2}$$

The expansion (1.2) was further extended to independent but not necessarily identically distributed random variables by Jing, Shao and Wang [21] under finite $(2+\rho)$th moments, $\rho \in (0, 1]$, showing that

$$\frac{\mathbf{P}(S_n/V_n \geq x)}{1-\Phi(x)} = \exp\{O(1)(1+x)^{2+\rho}\iota_n^\rho\} \tag{1.3}$$

uniformly for $0 \leq x = o(\min\{\iota_n^{-1}, \varsigma_n^{-1}\})$, where $O(1)$ is bounded by an absolute constant and

$$\iota_n^\rho = \sum_{i=1}^{n} \mathbf{E}|X_i|^{2+\rho}/B_n^{2+\rho} \quad \text{and} \quad \varsigma_n^2 = \max_{1\leq i\leq n} \mathbf{E}X_i^2/B_n^2. \tag{1.4}$$

For further self-normalized Cramér type moderate deviation results for independent random variables we refer, for example, to Hu, Shao and Wang [18], Liu, Shao and Wang [23], and Shao and

Zhou [29]. We also refer to de la Peña, Lai and Shao [10] and Shao and Wang [28] for recent developments in this area.

The theory for self-normalized sums of independent random variables has been studied in depth. However, we are not aware of any such results for martingales. For some closely related topic, that is, exponential inequalities for self-normalized martingales, we refer to de la Peña [9], Bercu and Touati [4], Chen, Wang, Xu and Miao [6] and Bercu, Delyon and Rio [3]. The main purpose of this paper is to establish self-normalized Cramér type moderate deviation results for martingales. Let $(\delta_n)_{n\geq 1}$, $(\varepsilon_n)_{n\geq 1}$ and $(\kappa_n)_{n\geq 1}$ be three sequences of nonnegative numbers, such that $\delta_n \to 0$, $\varepsilon_n \to 0$ and $\kappa_n \to 0$ as $n \to \infty$. Let $(X_i, \mathcal{F}_i)_{i\geq 1}$ be a sequence of martingale differences satisfying

$$\left|\sum_{i=1}^n \mathbf{E}\big[X_i^2|\mathcal{F}_{i-1}\big] - B_n^2\right| \leq \delta_n^2 B_n^2,$$

$$\sum_{i=1}^n \mathbf{E}\big[|X_i|^{2+\rho}|\mathcal{F}_{i-1}\big] \leq \varepsilon_n^\rho B_n^{2+\rho},$$

and

$$\max_{1\leq i \leq n} \mathbf{E}\big[X_i^2|\mathcal{F}_{i-1}\big] \leq \kappa_n^2 B_n^2,$$

where $\rho \in (0, \frac{3}{2}]$. Here and hereafter, the inequalities between random variables are understood in the **P**-almost sure sense. From Corollary 2.1 we have

$$\mathbf{P}(S_n/V_n \geq x) = \big(1 - \Phi(x)\big)\big(1 + o(1)\big) \tag{1.5}$$

uniformly for $0 \leq x = o(\min\{\varepsilon_n^{-\rho/(3+\rho)}, \delta_n^{-1}, \kappa_n^{-1}\})$ as $n \to \infty$. A more general Cramér type expansion is obtained in a larger range in our Theorem 2.1, from which we derive a moderate deviation principle for self-normalized martingales. Moreover, when the condition $\sum_{i=1}^n \mathbf{E}[|X_i|^{2+\rho}|\mathcal{F}_{i-1}] \leq \varepsilon_n^\rho B_n^{2+\rho}$ is replaced by a slightly stronger condition

$$\mathbf{E}\big[|X_i|^{2+\rho}|\mathcal{F}_{i-1}\big] \leq (\varepsilon_n B_n)^\rho \mathbf{E}\big[X_i^2|\mathcal{F}_{i-1}\big],$$

equality (1.5) holds for a larger range of $0 \leq x = o(\min\{\varepsilon_n^{-\rho/(4+2\rho)}, \delta_n^{-1}\})$ for $\rho \in (0, 1]$, see Corollary 2.4. Clearly, our results recover (1.2) for i.i.d. random variables.

The rest of the paper is organized as follows. Our main results are stated and discussed in Section 2. Section 3 provides the preliminary lemmas that are used in the proofs of the main results. In Section 4, we prove the main results.

Throughout the paper the symbols c and c_α, probably supplied with some indices, denote respectively a generic positive absolute constant and a generic positive constant depending only on α. Moreover, θ stands for values satisfying $|\theta| \leq 1$.

2. Main results

Let $(X_i, \mathcal{F}_i)_{i=0,\ldots,n}$ be a sequence of martingale differences defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, where $X_0 = 0$ and $\{\varnothing, \Omega\} = \mathcal{F}_0 \subseteq \cdots \subseteq \mathcal{F}_n \subseteq \mathcal{F}$ are increasing σ-fields. Set

$$S_0 = 0, \qquad S_k = \sum_{i=1}^{k} X_i, \quad k = 1, \ldots, n. \tag{2.1}$$

Then $S = (S_k, \mathcal{F}_k)_{k=0,\ldots,n}$ is a martingale. Denote $B_n^2 = \sum_{i=1}^{n} \mathbf{E} X_i^2$. Let $[S]$ and $\langle S \rangle$ be, respectively, the square bracket and the conditional variance of the martingale S, that is

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

and

$$\langle S \rangle_0 = 0, \qquad \langle S \rangle_k = \sum_{i=1}^{k} \mathbf{E}[X_i^2 | \mathcal{F}_{i-1}], \quad k = 1, \ldots, n. \tag{2.2}$$

In the sequel, we use the following conditions:

(A1) There exists $\delta_n \in [0, \frac{1}{4}]$ such that

$$\left| \sum_{i=1}^{n} \mathbf{E}[X_i^2 | \mathcal{F}_{i-1}] - B_n^2 \right| \le \delta_n^2 B_n^2;$$

(A2) There exist $\rho > 0$ and $\varepsilon_n \in (0, \frac{1}{4}]$ such that

$$\sum_{i=1}^{n} \mathbf{E}[|X_i|^{2+\rho} | \mathcal{F}_{i-1}] \le \varepsilon_n^\rho B_n^{2+\rho};$$

(A3) There exists $\kappa_n \in (0, \frac{1}{4}]$ such that

$$\mathbf{E}[X_i^2 | \mathcal{F}_{i-1}] \le \kappa_n^2 B_n^2, \quad 1 \le i \le n;$$

(A4) There exist $\rho \in (0, 1]$ and $\gamma_n \in (0, \frac{1}{4}]$ such that

$$\mathbf{E}[|X_i|^{2+\rho} | \mathcal{F}_{i-1}] \le (\gamma_n B_n)^\rho \mathbf{E}[X_i^2 | \mathcal{F}_{i-1}], \quad 1 \le i \le n.$$

When $\rho \in (0, 1]$ and $\gamma_n \le (16/17)^{1/\rho}/4$, conditions (A1) and (A4) imply condition (A2) with $\varepsilon_n = (17/16)^{1/\rho} \gamma_n$. Thus, we may assume that $\varepsilon_n = O(\gamma_n)$ as $n \to \infty$. It is also easy to see that condition (A4) implies condition (A3) with $\kappa_n = \gamma_n$, see Lemma 3.5.

In practice, we usually have $\max\{\delta_n, \varepsilon_n, \gamma_n, \kappa_n\} \to 0$ as $n \to \infty$. In the case of sums of i.i.d. random variables, conditions (A1), (A2), (A3), and (A4) are satisfied with $\delta_n = 0$, $\varepsilon_n, \gamma_n, \kappa_n = O(\frac{1}{\sqrt{n}})$.

Our first main result is the following Cramér type moderate deviation for the self-normalized martingale

$$W_n = S_n / \sqrt{[S]_n},$$

under conditions (A1), (A2), and (A3).

Theorem 2.1. *Assume that conditions* (A1), (A2), *and* (A3) *are satisfied. Set*

$$\rho_1 = \min\{\rho, 1\}.$$

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

$$\frac{\mathbf{P}(W_n \geq x)}{1 - \Phi(x)} = \exp\left\{\theta c_\rho\left(x^{2+\rho_1}\varepsilon_n^{\rho_1} + x^2\delta_n^2 + (1+x)\left(\varepsilon_n^{\rho/(3+\rho)} + \delta_n\right)\right)\right\}. \tag{2.3}$$

Under condition (A2) the best Berry–Esseen bound for standardized martingales is provided by Haeusler [16]: assuming $\langle S \rangle_n = B_n^2$ a.s., Haeusler proved that

$$\sup_x \left|\mathbf{P}(S_n/B_n \leq x) - \Phi(x)\right| \leq c_\rho \left(\sum_{i=1}^n \mathbf{E}|X_i/B_n|^{2+\rho}\right)^{1/(3+\rho)}.$$

Moreover, it was showed that this bound cannot be improved for martingales with finite $(2 + \rho)$th moments. In fact, there exist a positive constant $c_{0,\rho}$ and a sequence of martingale differences satisfying $\mathbf{P}(S_n \leq 0) - \Phi(0) \geq c_{0,\rho}(\sum_{i=1}^n \mathbf{E}|X_i/B_n|^{2+\rho})^{1/(3+\rho)}$ for all large enough n. In particular, under conditions (A2) and $\langle S \rangle_n = B_n^2$ a.s., Haeusler's result implies that

$$\sup_x \left|\mathbf{P}(S_n/B_n \leq x) - \Phi(x)\right| \leq c_\rho \varepsilon_n^{\rho/(3+\rho)}. \tag{2.4}$$

Notice that Theorem 2.1 implies that, for each absolute constant $c > 0$ there is a positive constant c_ρ depending on ρ such that for n large enough,

$$\sup_{|x| \leq c} \left|\mathbf{P}(W_n \leq x) - \Phi(x)\right| \leq c_\rho\left(\varepsilon_n^{\rho/(3+\rho)} + \delta_n\right). \tag{2.5}$$

Under conditions (A2) and $\langle S \rangle_n = B_n^2$ a.s., the bound in (2.5) for self-normalized martingales is of the same order as the bound in (2.4) for standardized martingales.

From Theorem 2.1, we obtain the following result about the equivalence to the normal tail.

Corollary 2.1. *Assume that conditions* (A1), (A2), *and* (A3) *are satisfied with* $\rho \in (0, \frac{3}{2}]$. *Then*

$$\frac{\mathbf{P}(W_n \geq x)}{1 - \Phi(x)} = 1 + o(1)$$

uniformly for $0 \leq x = o(\min\{\varepsilon_n^{-\rho/(3+\rho)}, \kappa_n^{-1}, \delta_n^{-1}\})$ *as* $n \to \infty$.

Theorem 2.1 also implies the following moderate deviation principles (MDP) for self-normalized martingales.

Corollary 2.2. *Assume conditions* (A1), (A2), *and* (A3) *with* $\max\{\delta_n, \varepsilon_n, \kappa_n\} \to 0$ *as* $n \to \infty$. *Let* a_n *be any sequence of real numbers satisfying* $a_n \to \infty$ *and* $a_n \varepsilon_n \to 0$ *as* $n \to \infty$. *Then for each Borel set* B,

$$
-\inf_{x \in B^o} \frac{x^2}{2} \leq \liminf_{n \to \infty} \frac{1}{a_n^2} \ln \mathbf{P}\left(\frac{W_n}{a_n} \in B \right)
$$
$$
\leq \limsup_{n \to \infty} \frac{1}{a_n^2} \ln \mathbf{P}\left(\frac{W_n}{a_n} \in B \right) \leq -\inf_{x \in \overline{B}} \frac{x^2}{2}, \tag{2.6}
$$

where B^o *and* $\overline{B}$ *denote the interior and the closure of* B, *respectively.*

The last corollary shows that the convergence speed of MDP depends only on ε_n and it has nothing to do with the convergence speeds of κ_n and δ_n.

For i.i.d. random variables, the self-normalized MDP was established by Shao [26]. See also Jing, Liang and Zhou [20] for non-identically distributed random variables.

The other main results concern some improvements of Theorem 2.1 when condition (A3) is replaced by the stronger condition (A4). Theorems 2.2 and 2.3 below give respectively lower and upper bounds, while Theorem 2.4 gives a Cramér type moderate deviation expansion sharper than that in Theorem 2.1.

Theorem 2.2. *Assume that conditions* (A1), (A2), *and* (A4) *are satisfied.*

[i] *If* $\rho \in (0, 1)$, *then for* $0 \leq x = o(\gamma_n^{-1})$,

$$
\frac{\mathbf{P}(W_n \geq x)}{1 - \Phi(x)} \geq \exp\left\{ -c_\rho \left(x^{2+\rho} \varepsilon_n^\rho + x^2 \delta_n^2 + (1 + x)\left(x^\rho \gamma_n^\rho + \gamma_n^\rho + \delta_n \right) \right) \right\}. \tag{2.7}
$$

[ii] *If* $\rho = 1$, *then for* $0 \leq x = o(\gamma_n^{-1})$,

$$
\frac{\mathbf{P}(W_n \geq x)}{1 - \Phi(x)} \geq \exp\left\{ -c\left(x^3 \varepsilon_n + x^2 \delta_n^2 + (1 + x)\left(x\gamma_n + \gamma_n |\ln \gamma_n| + \delta_n \right) \right) \right\}. \tag{2.8}
$$

The term $\gamma_n |\ln \gamma_n|$ in (2.8) cannot be replaced by γ_n under the stated conditions. Indeed, Bolthausen [5] (see Example 2 therein) showed that there exists a sequence of martingale differences satisfying $|X_i| \leq 2$ and $\langle S \rangle_n = n$ a.s., such that for all n large enough,

$$
\left| \mathbf{P}(S_n \geq 0) - \Phi(0) \right| \geq \frac{c \log n}{\sqrt{n}}, \tag{2.9}
$$

where c does not depend on n. Inequality (2.9) shows that the term $\gamma_n |\ln \gamma_n|$ in (2.8) cannot be replaced by γ_n even for bounded martingale differences.

For any sequence of positive numbers $(\alpha_n)_{n \geq 1}$ denote

$$\widehat{\alpha}_n(x, \rho) = \frac{\alpha_n^{\rho(2-\rho)/4}}{1 + x^{\rho(2+\rho)/4}}. \tag{2.10}$$

Accordingly, we shall use below the notations $\widehat{\varepsilon}_n(x, \rho)$ and $\widehat{\gamma}_n(x, \rho)$, which mean sequences defined by (2.10) with α_n replaced by ε_n and γ_n.

Theorem 2.3. *Assume that conditions* (A1), (A2), *and* (A4) *are satisfied.*

[i] *If $\rho \in (0, 1)$, then for $0 \leq x = o(\gamma_n^{-1})$,*

$$\frac{\mathbf{P}(W_n \geq x)}{1 - \Phi(x)} \leq \exp\{c_\rho(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2 + (1+x)(x^\rho\gamma_n^\rho + \gamma_n^\rho + \delta_n + \widehat{\varepsilon}_n(x, \rho)))\}.$$

[ii] *If $\rho = 1$, then for $0 \leq x = o(\gamma_n^{-1})$,*

$$\frac{\mathbf{P}(W_n \geq x)}{1 - \Phi(x)} \leq \exp\{c(x^3\varepsilon_n + x^2\delta_n^2 + (1+x)(x\gamma_n + \gamma_n|\ln \gamma_n| + \delta_n + \widehat{\varepsilon}_n(x, 1)))\}.$$

Combining Theorems 2.2 and 2.3, we obtain the following Cramér type moderate deviation expansion for self-normalized martingales under conditions (A1), (A2), and (A4), which is stronger than the expansion in Theorem 2.1 since the term $\varepsilon_n^{\rho/(3+\rho)}$ therein is improved to a smaller one.

Theorem 2.4. *Assume that conditions* (A1), (A2), *and* (A4) *are satisfied.*

[i] *If $\rho \in (0, 1)$, then for $0 \leq x = o(\gamma_n^{-1})$,*

$$\frac{\mathbf{P}(W_n \geq x)}{1 - \Phi(x)} = \exp\{\theta c_\rho(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2 + (1+x)(x^\rho\gamma_n^\rho + \gamma_n^\rho + \delta_n + \widehat{\varepsilon}_n(x, \rho)))\}.$$

[ii] *If $\rho = 1$, then for $0 \leq x = o(\gamma_n^{-1})$,*

$$\frac{\mathbf{P}(W_n \geq x)}{1 - \Phi(x)} = \exp\{\theta c(x^3\varepsilon_n + x^2\delta_n^2 + (1+x)(x\gamma_n + \gamma_n|\ln \gamma_n| + \delta_n + \widehat{\varepsilon}_n(x, 1)))\}.$$

Notice that condition (A4) implies condition (A2) with $\varepsilon_n = \gamma_n$. Therefore, it follows from Theorem 2.4 the following corollary.

Corollary 2.3. *Assume that conditions* (A1) *and* (A4) *are satisfied.*

[i] *If $\rho \in (0, 1)$, then for $0 \leq x = o(\gamma_n^{-1})$,*

$$\frac{\mathbf{P}(W_n \geq x)}{1 - \Phi(x)} = \exp\{\theta c_\rho(x^{2+\rho}\gamma_n^\rho + x^2\delta_n^2 + (1+x)(\delta_n + \widehat{\gamma}_n(x, \rho)))\}.$$

[ii] *If $\rho = 1$, then for $0 \leq x = o(\gamma_n^{-1})$,*

$$\frac{\mathbf{P}(W_n \geq x)}{1 - \Phi(x)} = \exp\left\{\theta c\left(x^3 \gamma_n + x^2 \delta_n^2 + (1 + x)\left(\delta_n + \gamma_n |\ln \gamma_n| + \widehat{\gamma}_n(x, 1)\right)\right)\right\}.$$

From Theorem 2.4, we also obtain the following result about the equivalence to the normal tail.

Corollary 2.4. *Assume conditions* (A1), (A2), *and* (A4) *with $\rho \in (0, 1]$. Then*

$$\frac{\mathbf{P}(W_n \geq x)}{1 - \Phi(x)} = 1 + o(1) \tag{2.11}$$

uniformly for $0 \leq x = o(\min\{\varepsilon_n^{-\rho/(2+\rho)}, \gamma_n^{-\rho/(1+\rho)}, \delta_n^{-1}\})$ as $n \to \infty$.

In the case of i.i.d. random variables, conditions (A1), (A2), and (A4) are satisfied with $\varepsilon_n, \gamma_n = O(1/\sqrt{n})$ and $\delta_n = 0$. Thus, the range $0 \leq x = o(\min\{\varepsilon_n^{-\rho/(2+\rho)}, \delta_n^{-1}, \gamma_n^{-\rho/(1+\rho)}\})$ reduces to $0 \leq x = o(n^{-\rho/(4+2\rho)})$, $n \to \infty$, which is the best possible result such that (2.11) holds (see Shao [27]). Moreover, from Theorem 2.4, we can get the estimation of the rate of convergence in (2.11); for example, when $\rho = 1$ we have:

Corollary 2.5. *Assume conditions* (A1), (A2), *and* (A4) *with $\rho = 1$, $\varepsilon_n, \gamma_n, \delta_n = O(1/\sqrt{n})$. Then with $c_0 > 0$ for $c_0 n^{3/22} \leq x = o(n^{1/2})$ as $n \to \infty$,*

$$\frac{\mathbf{P}(W_n \geq x)}{1 - \Phi(x)} = \exp\left\{\theta c \frac{x^3}{n^{1/2}}\right\}. \tag{2.12}$$

In particular, with $c_0, c_1 > 0$ for $c_0 n^{3/22} \leq x \leq c_1 n^{1/6}$,

$$\left|\frac{\mathbf{P}(W_n \geq x)}{1 - \Phi(x)} - 1\right| \leq c \frac{x^3}{n^{1/2}}. \tag{2.13}$$

Notice that the rate of convergence in (2.12) coincides with that in (1.3) for i.i.d. random variables.

Remark 2.1. Notice that if $(S_k, \mathcal{F}_k)_{k=0,\dots,n}$ satisfies conditions (A1), (A2), (A3), and (A4), then $(-S_k, \mathcal{F}_k)_{k=0,\dots,n}$ also satisfies the same conditions. Thus the assertions in Theorems 2.1–2.4 and Corollaries 2.1–2.5 remain valid when $\frac{\mathbf{P}(W_n \geq x)}{1 - \Phi(x)}$ is replaced by $\frac{\mathbf{P}(W_n \leq -x)}{\Phi(-x)}$.

3. Preliminary lemmas

The proofs of Theorems 2.1–2.4 are based on a conjugate multiplicative martingale technique for changing the probability measure which is similar to that of the transformation of Esscher [12]. Our approach is inspired by the earlier work of Grama and Haeusler [15] on Cramér moderate

deviations for standardized martingales, and by that of Shao [27], Jing, Shao and Wang [21], who developed techniques for moderate deviations of self-normalized sums of independent random variables. We extend these work by introducing a new choice of the density for the change of measure and refining the approaches in Shao [27] and Jing, Shao and Wang [21] to handle self-normalized martingales. A key point of the proof is a new Berry–Esseen bound for martingales under the changed measure, see Proposition 3.1 below.

Let

$$\xi_i = \frac{X_i}{B_n}, \quad i = 1, \ldots, n.$$

Then $(\xi_i, \mathcal{F}_i)_{i=0,\ldots,n}$ is also a sequence of martingale differences. Moreover, for simplicity of notations, set

$$M_k = \sum_{i=1}^{k} \xi_i,$$

$$[M]_k = \sum_{i=1}^{k} \xi_i^2 \quad \text{and} \quad \langle M \rangle_k = \sum_{i=1}^{k} \mathbf{E}[\xi_i^2 | \mathcal{F}_{i-1}], \quad k = 1, \ldots, n.$$

Thus,

$$W_n = \frac{S_n}{\sqrt{[S]_n}} = \frac{M_n}{\sqrt{[M]_n}}. \tag{3.1}$$

For any real number λ, consider the *exponential multiplicative martingale* $Z(\lambda) = (Z_k(\lambda), \mathcal{F}_k)_{k=0,\ldots,n}$, where

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

with

$$\zeta_i(\lambda) = \lambda \xi_i - \lambda^2 \xi_i^2 / 2.$$

Thus, for each real number λ and each $k = 1, \ldots, n$, the random variable $Z_k(\lambda)$ is nonnegative and $\mathbf{E}Z_k(\lambda) = 1$. The last observation allows us to introduce the *conjugate probability measure* $\mathbf{P}_\lambda := \mathbf{P}_{\lambda,n}$ on $(\Omega, \mathcal{F})$ defined by

$$d\mathbf{P}_\lambda = Z_n(\lambda)\, d\mathbf{P}. \tag{3.2}$$

Although $(M_k, \mathcal{F}_k)_{k=0,\ldots,n}$ is a martingale under the measure $\mathbf{P}$, it is no longer a martingale under the conjugate probability measure $\mathbf{P}_\lambda$. To obtain a martingale under $\mathbf{P}_\lambda$ we have to center the random variables $\zeta_i(\lambda)$. Denote by $\mathbf{E}_\lambda$ the expectation with respect to $\mathbf{P}_\lambda$. Because $Z(\lambda)$ is a uniformly integrable martingale under $\mathbf{P}$, we have

$$\mathbf{E}_\lambda[\zeta] = \mathbf{E}[\zeta Z_n(\lambda)] \tag{3.3}$$

and

$$\mathbf{E}_\lambda[\zeta|\mathcal{F}_{i-1}] = \frac{\mathbf{E}[\zeta e^{\zeta_i(\lambda)}|\mathcal{F}_{i-1}]}{\mathbf{E}[e^{\zeta_i(\lambda)}|\mathcal{F}_{i-1}]} \tag{3.4}$$

for any $\mathcal{F}_i$-measurable random variable ζ that is integrable with respect to $\mathcal{F}_i$. Set

$$b_i(\lambda) = \mathbf{E}_\lambda\big[\zeta_i(\lambda)|\mathcal{F}_{i-1}\big], \quad i = 1, \ldots, n,$$

$$\eta_i(\lambda) = \zeta_i(\lambda) - b_i(\lambda), \quad i = 1, \ldots, n,$$

and

$$Y_k(\lambda) = \sum_{i=1}^{k} \eta_i(\lambda), \quad k = 1, \ldots, n. \tag{3.5}$$

Then $Y(\lambda) = (Y_k(\lambda), \mathcal{F}_k)_{k=0,\ldots,n}$ is the *conjugate martingale*. The following semimartingale decomposition is well-known:

$$\sum_{i=1}^{k} \zeta_i(\lambda) = B_k(\lambda) + Y_k(\lambda), \quad k = 1, \ldots, n, \tag{3.6}$$

where $B(\lambda) = (B_k(\lambda), \mathcal{F}_k)_{k=0,\ldots,n}$ is the *drift process* defined as

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

By the relation between $\mathbf{E}$ and $\mathbf{E}_\lambda$ on $\mathcal{F}_i$, we have

$$b_i(\lambda) = \frac{\mathbf{E}[\zeta_i(\lambda)e^{\zeta_i(\lambda)}|\mathcal{F}_{i-1}]}{\mathbf{E}[e^{\zeta_i(\lambda)}|\mathcal{F}_{i-1}]}, \quad i = 1, \ldots, n. \tag{3.7}$$

It is easy to compute the conditional variance of the conjugate martingale $Y(\lambda)$ under the measure $\mathbf{P}_\lambda$, for $k = 0, \ldots, n$,

$$\langle Y(\lambda) \rangle_k = \sum_{i=1}^{k} \mathbf{E}_\lambda\big[\eta_i(\lambda)^2|\mathcal{F}_{i-1}\big]$$

$$= \sum_{i=1}^{k} \mathbf{E}_\lambda\big[\big(\zeta_i(\lambda) - b_i(\lambda)\big)^2|\mathcal{F}_{i-1}\big] \tag{3.8}$$

$$= \sum_{i=1}^{k} \left(\frac{\mathbf{E}[\zeta_i^2(\lambda)e^{\zeta_i(\lambda)}|\mathcal{F}_{i-1}]}{\mathbf{E}[e^{\zeta_i(\lambda)}|\mathcal{F}_{i-1}]} - \frac{\mathbf{E}[\zeta_i(\lambda)e^{\zeta_i(\lambda)}|\mathcal{F}_{i-1}]^2}{\mathbf{E}[e^{\zeta_i(\lambda)}|\mathcal{F}_{i-1}]^2} \right).$$

In the sequel, we give the upper and lower bounds for $B_n(\lambda)$. To this end, we need the following three useful lemmas. Their proofs are not given here but they are similar to those of the

corresponding assertions in Shao [27] and Jing, Shao and Wang [21] established for independent random variables. Set

$$\widetilde{\varepsilon}_{i,\lambda} = \lambda^2 \mathbf{E}\big[\xi_i^2 \mathbf{1}_{\{|\lambda\xi_i|>1\}}|\mathcal{F}_{i-1}\big] + \lambda^3 \mathbf{E}\big[|\xi_i|^3 \mathbf{1}_{\{|\lambda\xi_i|\leq 1\}}|\mathcal{F}_{i-1}\big], \quad \lambda \geq 0.$$

If $\mathbf{E}[|\xi_i|^{2+\rho}] < \infty$ for $\rho \in [0, 1]$, then it is obvious that

$$\widetilde{\varepsilon}_{i,\lambda} \leq \lambda^{2+\rho} \mathbf{E}\big[|\xi_i|^{2+\rho}|\mathcal{F}_{i-1}\big], \quad \lambda \geq 0.$$

Lemma 3.1. *For all* $\lambda > 0$ *and* $\tau \in [\frac{1}{8}, 2]$, *we have*

$$\left| \mathbf{E}\big[e^{\lambda\xi_i - \tau\lambda^2\xi_i^2}|\mathcal{F}_{i-1}\big] - 1 - \left(\frac{1}{2} - \tau\right)\lambda^2 \mathbf{E}\big[\xi_i^2|\mathcal{F}_{i-1}\big]\right| \leq c\widetilde{\varepsilon}_{i,\lambda}.$$

Lemma 3.2. *For all* $\lambda > 0$, *we have*

$$\big|\mathbf{E}\big[e^{\zeta_i(\lambda)}|\mathcal{F}_{i-1}\big] - 1\big| \leq c\widetilde{\varepsilon}_{i,\lambda},$$

$$\left|\mathbf{E}\big[\zeta_i(\lambda)e^{\zeta_i(\lambda)}|\mathcal{F}_{i-1}\big] - \frac{1}{2}\lambda^2 \mathbf{E}\big[\xi_i^2|\mathcal{F}_{i-1}\big]\right| \leq c\widetilde{\varepsilon}_{i,\lambda},$$

$$\big|\mathbf{E}\big[\zeta_i^2(\lambda)e^{\zeta_i(\lambda)}|\mathcal{F}_{i-1}\big] - \lambda^2 \mathbf{E}\big[\xi_i^2|\mathcal{F}_{i-1}\big]\big| \leq c\widetilde{\varepsilon}_{i,\lambda},$$

$$\mathbf{E}\big[|\zeta_i(\lambda)|^3 e^{\zeta_i(\lambda)}|\mathcal{F}_{i-1}\big] \leq c\widetilde{\varepsilon}_{i,\lambda},$$

$$\big(\mathbf{E}\big[\zeta_i(\lambda)e^{\zeta_i(\lambda)}|\mathcal{F}_{i-1}\big]\big)^2 \leq c\widetilde{\varepsilon}_{i,\lambda}.$$

Lemma 3.3. *Let* $H_i = \xi_i^2 - \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}]$. *Then for all* $\lambda > 0$,

$$\big|\mathbf{E}\big[H_i e^{\zeta_i(\lambda)}|\mathcal{F}_{i-1}\big]\big| \leq c\frac{1}{\lambda^2}\widetilde{\varepsilon}_{i,\lambda}.$$

Using Lemma 3.2, we obtain the following upper and lower bounds for $B_n(\lambda)$.

Lemma 3.4. *Assume conditions* (A2) *and* (A3) *with* $\rho \in (0, 1]$. *Then for* $0 \leq \lambda = o(\max\{\varepsilon_n^{-1}, \kappa_n^{-1}\})$,

$$\left|B_n(\lambda) - \frac{1}{2}\lambda^2\langle M\rangle_n\right| \leq c\lambda^{2+\rho}\varepsilon_n^\rho. \tag{3.9}$$

Proof. According to the definition of $b_i(\lambda)$, we have

$$b_i(\lambda) = \frac{\mathbf{E}[\zeta_i(\lambda)e^{\zeta_i(\lambda)}|\mathcal{F}_{i-1}]}{\mathbf{E}[e^{\zeta_i(\lambda)}|\mathcal{F}_{i-1}]}.$$

By Lemma 3.2, it follows that

$$\left|\mathbf{E}\big[\zeta_i(\lambda)e^{\zeta_i(\lambda)}|\mathcal{F}_{i-1}\big] - \frac{1}{2}\lambda^2\mathbf{E}\big[\xi_i^2|\mathcal{F}_{i-1}\big]\right| \leq c\widetilde{\varepsilon}_{i,\lambda}$$

and

$$\left|\mathbf{E}\left[e^{\xi_i(\lambda)}|\mathcal{F}_{i-1}\right] - 1\right| \le c\widetilde{\varepsilon}_{i,\lambda}. \tag{3.10}$$

Therefore, conditions (A2) and (A3) imply that for $0 \le \lambda = o(\max\{\varepsilon_n^{-1}, \kappa_n^{-1}\})$,

$$\left|b_i(\lambda) - \frac{1}{2}\lambda^2 \mathbf{E}\left[\xi_i^2|\mathcal{F}_{i-1}\right]\right| \le c\widetilde{\varepsilon}_{i,\lambda}$$

and

$$\left|B_n(\lambda) - \frac{1}{2}\lambda^2\langle M\rangle_n\right| \le c\lambda^{2+\rho}\varepsilon_n^\rho$$

as desired. $\square$

The following lemma shows that condition (A4) implies condition (A3) with $\kappa_n = \gamma_n$.

Lemma 3.5. *Assume condition* (A4). *Then* $\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] \le \gamma_n^2$.

Proof. By Jensen's inequality and condition (A4), it holds that

$$\mathbf{E}\left[\xi_i^2|\mathcal{F}_{i-1}\right]^{(2+\rho)/2} \le \mathbf{E}\left[|\xi_i|^{2+\rho}|\mathcal{F}_{i-1}\right] \le \gamma_n^\rho \mathbf{E}\left[\xi_i^2|\mathcal{F}_{i-1}\right],$$

from which we get $\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] \le \gamma_n^2$. $\square$

Lemma 3.6. *Assume condition* (A4). *Then for any* $t \in [0, \rho)$,

$$\mathbf{E}\left[|\xi_i|^{2+t}|\mathcal{F}_{i-1}\right] \le \gamma_n^t \mathbf{E}\left[\xi_i^2|\mathcal{F}_{i-1}\right]. \tag{3.11}$$

Proof. Let l, p, q be defined by the following equations

$$lp = 2, \qquad (2+t-l)q = 2+\rho, \qquad p^{-1} + q^{-1} = 1, \qquad l > 0, \quad \text{and} \quad p, q \ge 1.$$

Solving the last equations, we get

$$l = \frac{2(\rho - t)}{\rho}, \qquad p = \frac{\rho}{\rho - t}, \qquad q = \frac{\rho}{t}.$$

By Hölder's inequality and condition (A4), it is easy to see that

$$\begin{aligned}
\mathbf{E}\left[|\xi_i|^{2+t}|\mathcal{F}_{i-1}\right] &= \mathbf{E}\left[|\xi_i|^l|\xi_i|^{2+t-l}|\mathcal{F}_{i-1}\right] \\
&\le \left(\mathbf{E}\left[|\xi_i|^{lp}|\mathcal{F}_{i-1}\right]\right)^{1/p}\left(\mathbf{E}\left[|\xi_i|^{(2+t-l)q}|\mathcal{F}_{i-1}\right]\right)^{1/q} \\
&\le \left(\mathbf{E}\left[\xi_i^2|\mathcal{F}_{i-1}\right]\right)^{1/p}\left(\mathbf{E}\left[|\xi_i|^{2+\rho}|\mathcal{F}_{i-1}\right]\right)^{1/q} \\
&\le \left(\mathbf{E}\left[\xi_i^2|\mathcal{F}_{i-1}\right]\right)^{1/p}\left(\gamma_n^\rho \mathbf{E}\left[\xi_i^2|\mathcal{F}_{i-1}\right]\right)^{1/q}
\end{aligned}$$

$$\leq \gamma_n^{\rho/q} \mathbf{E}[\xi_i^2 | \mathcal{F}_{i-1}]$$
$$= \gamma_n^t \mathbf{E}[\xi_i^2 | \mathcal{F}_{i-1}].$$

This completes the proof of the lemma. □

Lemma 3.7. *Assume conditions* (A1) *and* (A2). *Then for any* $t \in [0, \rho)$,

$$\sum_{i=1}^n \mathbf{E}[|\xi_i|^{2+t} | \mathcal{F}_{i-1}] \leq 2\varepsilon_n^t. \tag{3.12}$$

Proof. Recall the notations in the proof of Lemma 3.6. It is easy to see that

$$\sum_{i=1}^n \mathbf{E}[|\xi_i|^{2+t} | \mathcal{F}_{i-1}] \leq \sum_{i=1}^n \left(\mathbf{E}[\xi_i^2 | \mathcal{F}_{i-1}]\right)^{1/p} \left(\mathbf{E}[|\xi_i|^{2+\rho} | \mathcal{F}_{i-1}]\right)^{1/q}.$$

Using Hölder's inequality and conditions (A1) and (A2), we have

$$\sum_{i=1}^n \mathbf{E}[|\xi_i|^{2+t} | \mathcal{F}_{i-1}] \leq \left(\sum_{i=1}^n \mathbf{E}[\xi_i^2 | \mathcal{F}_{i-1}]\right)^{1/p} \left(\sum_{i=1}^n \mathbf{E}[|\xi_i|^{2+\rho} | \mathcal{F}_{i-1}]\right)^{1/q}$$
$$\leq 2\varepsilon_n^t,$$

which gives the desired inequality. □

We will also need the following two lemmas.

Lemma 3.8. *Assume condition* (A1). *Then for all* $x > 0$,

$$\mathbf{P}\left(M_n \geq x\sqrt{[M]_n}, [M]_n \geq 16\right) \leq \frac{2}{3} x^{-2/3} \exp\left\{-\frac{3}{4}x^2\right\}.$$

Proof. By inequality (11) of Delyon [8], we have for all $\lambda \in \mathbf{R}$,

$$\mathbf{E}\exp\left\{\lambda M_n - \frac{\lambda^2}{2}\left(\frac{1}{3}[M]_n + \frac{2}{3}\langle M \rangle_n\right)\right\} \leq 1.$$

Applying the last inequality to the exponential inequality of de la Peña and Pang [11] with $p = q = 2$, we deduce that for all $x > 0$,

$$\mathbf{P}\left(\frac{|M_n|}{\sqrt{\frac{3}{2}(\frac{1}{3}[M]_n + \frac{2}{3}\langle M \rangle_n + \mathbf{E}M_n^2)}} \geq x\right) \leq \left(\frac{2}{3}\right)^{2/3} x^{-2/3} \exp\left\{-\frac{1}{2}x^2\right\}. \tag{3.13}$$

By condition (A1) and the fact that $\mathbf{E}\langle M\rangle_n = \mathbf{E}M_n^2 = 1$, it is easy to see that

$$\frac{3}{2}\langle M\rangle_n + \frac{9}{4}\mathbf{E}M_n^2 \leq \frac{3}{2}(1+\delta_n^2) + \frac{9}{4} \leq \frac{3}{2}\left(1+\frac{1}{16}\right) + \frac{9}{4} < 4.$$

Therefore, for all $x > 0$,

$$\mathbf{P}\left(M_n \geq x\sqrt{[M]_n}, [M]_n \geq 16\right) \leq \mathbf{P}\left(M_n \geq x\sqrt{\frac{3}{4}[M]_n + 4}, [M]_n \geq 16\right)$$

$$\leq \mathbf{P}\left(M_n \geq x\sqrt{\frac{3}{4}[M]_n + \frac{3}{2}\langle M\rangle_n + \frac{9}{4}\mathbf{E}M_n^2}, [M]_n \geq 16\right)$$

$$\leq \mathbf{P}\left(M_n \geq x\sqrt{\frac{3}{4}[M]_n + \frac{3}{2}\langle M\rangle_n + \frac{9}{4}\mathbf{E}M_n^2}\right)$$

$$= \mathbf{P}\left(M_n \geq \sqrt{\frac{3}{2}}x\sqrt{\frac{1}{2}[M]_n + \langle M\rangle_n + \frac{3}{2}\mathbf{E}M_n^2}\right)$$

$$\leq \frac{2}{3}x^{-2/3}\exp\left\{-\frac{3}{4}x^2\right\}$$

as desired.   □

Lemma 3.9. *Assume conditions* (A1) *and* (A2). *Then for all* $\rho > 0$,

$$\mathbf{P}\left(\left|[M]_n - \langle M\rangle_n\right| \geq 1\right) \leq c_\rho\left(\varepsilon_n^{(2+\rho)/2} + \varepsilon_n^\rho\right).$$

Proof. Notice that $[M]_n - \langle M\rangle_n = \sum_{i=1}^n (\xi_i^2 - \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}])$ is a martingale. For ρ, we distinguish two cases as follows.

When $\rho \in (0, 2]$, by the inequality of von Bahr and Esseen [32], it follows that

$$\mathbf{E}\left[\left|[M]_n - \langle M\rangle_n\right|^{(2+\rho)/2}\right] \leq \sum_{i=1}^n \mathbf{E}\left[\left|\xi_i^2 - \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}]\right|^{(2+\rho)/2}\right]$$

$$\leq c_1 \sum_{i=1}^n \mathbf{E}\left[|\xi_i|^{2+\rho}\right]$$

$$\leq c_2\varepsilon_n^\rho,$$

where the last line follows by conditions (A1) and (A2). Hence, by Markov's inequality,

$$\mathbf{P}\left(\left|[M]_n - \langle M\rangle_n\right| \geq 1\right) \leq \mathbf{E}\left[\left|[M]_n - \langle M\rangle_n\right|^{(2+\rho)/2}\right]$$

$$\leq c_2\varepsilon_n^\rho.$$

When $\rho > 2$, by Rosenthal's inequality (cf., Theorem 2.12 of Hall and Heyde [17]), Lemma 3.7, and condition (A2), it follows that

$$
\mathbf{E}\big[\big|[M]_n - \langle M \rangle_n\big|^{(2+\rho)/2}\big]
$$

$$
\leq c_{\rho,1}\Bigg(\mathbf{E}\bigg(\sum_{i=1}^n \mathbf{E}\big[(\xi_i^2 - \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}])^2|\mathcal{F}_{i-1}\big]\bigg)^{(2+\rho)/4} + \sum_{i=1}^n \mathbf{E}\big|\xi_i^2 - \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}]\big|^{(2+\rho)/2}\Bigg)
$$

$$
\leq c_{\rho,2}\Bigg(\mathbf{E}\bigg(\sum_{i=1}^n \mathbf{E}[\xi_i^4|\mathcal{F}_{i-1}]\bigg)^{(2+\rho)/4} + \sum_{i=1}^n \mathbf{E}|\xi_i|^{2+\rho}\Bigg)
$$

$$
\leq c_{\rho,3}\big(\varepsilon_n^{(2+\rho)/2} + \varepsilon_n^\rho\big).
$$

(3.14)

This completes the proof of the lemma. □

Consider the predictable process $\Psi(\lambda) = (\Psi_k(\lambda), \mathcal{F}_k)_{k=0,\dots,n}$, which is related to the martingale M as follows:

$$
\Psi_k(\lambda) = \sum_{i=1}^k \ln \mathbf{E}\big[e^{\zeta_i(\lambda)}|\mathcal{F}_{i-1}\big].
$$

(3.15)

By equality (3.10), we easily obtain the following elementary bound for the process $\Psi(\lambda)$.

Lemma 3.10. *Assume conditions* (A2) *and* (A3) *with* $\rho \in (0, 1]$. *Then for* $0 \leq \lambda = o(\min\{\varepsilon_n^{-1}, \kappa_n^{-1}\})$,

$$
\big|\Psi_n(\lambda)\big| \leq c\lambda^{2+\rho}\varepsilon_n^\rho.
$$

In the proofs of Theorems 2.2 and 2.3, we make use of the following assertion, which gives us a rate of convergence in the CLT for the conjugate martingale $Y(\lambda)$ under the probability measure $\mathbf{P}_\lambda$.

Proposition 3.1. *Assume conditions* (A1) *and* (A4). *With the convention that* $Y_n(0)/0 = M_n$, *we have*:

[i] *If* $\rho \in (0, 1)$, *then for* $0 \leq \lambda = o(\gamma_n^{-1})$,

$$
\sup_x \big|\mathbf{P}_\lambda\big(Y_n(\lambda)/\lambda \leq x\big) - \Phi(x)\big| \leq c_\rho\big(\lambda^\rho \gamma_n^\rho + \gamma_n^\rho + \delta_n\big).
$$

[ii] *If* $\rho = 1$, *then for* $0 \leq \lambda = o(\gamma_n^{-1})$,

$$
\sup_x \big|\mathbf{P}_\lambda\big(Y_n(\lambda)/\lambda \leq x\big) - \Phi(x)\big| \leq c\big(\lambda \gamma_n + \gamma_n|\ln \gamma_n| + \delta_n\big).
$$

Similarly, we have the following Berry–Esseen bound.

Proposition 3.2. *Assume conditions* (A1), (A2) *and* (A3). *Then for* $0 \le \lambda = o(\max\{\varepsilon_n^{-1}, \kappa_n^{-1}\})$,

$$\sup_x \big| \mathbf{P}_\lambda \big(Y_n(\lambda)/\lambda \le x \big) - \Phi(x) \big| \le c_\rho \big(\lambda^{\rho/2} \gamma_n^{\rho/2} + \varepsilon_n^{\rho/(3+\rho)} + \delta_n \big),$$

with the convention that $Y_n(0)/0 = M_n$.

The proofs of Propositions 3.1 and 3.2 are much more complicated and we give details in the supplemental article Fan, Grama, Liu and Shao [13].

4. Proof of the main results

We start with the proofs of Theorems 2.2 and 2.3, and conclude with the proof of Theorem 2.1. Theorem 2.4 is an easy consequence of Theorems 2.2 and 2.3

4.1. Proof of Theorem 2.2

Recall that

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

By (3.1), it is easy to see that

$$\big\{ S_n \ge x \sqrt{[S]_n} \big\} = \big\{ M_n \ge x \sqrt{[M]_n} \big\} \supseteq \left\{ M_n \ge \frac{x^2 + \lambda^2 [M]_n}{2\lambda} \right\} = \left\{ \sum_{i=1}^n \zeta_i(\lambda) \ge \frac{x^2}{2} \right\}.$$

For $0 \le \lambda = o(\gamma_n^{-1})$, according to (3.2), (3.6) and (3.15), we have the following representation:

$$\mathbf{P}(W_n \ge x) = \mathbf{E}_\lambda \big[Z_n(\lambda)^{-1} \mathbf{1}_{\{S_n \ge x \sqrt{[S]_n}\}} \big]$$

$$= \mathbf{E}_\lambda \left[\exp\left\{ -\sum_{i=1}^n \zeta_i(\lambda) + \Psi_n(\lambda) \right\} \mathbf{1}_{\{M_n \ge x \sqrt{[M]_n}\}} \right]$$

$$\ge \mathbf{E}_\lambda \big[\exp\{ -Y_n(\lambda) - B_n(\lambda) + \Psi_n(\lambda) \} \mathbf{1}_{\{\sum_{i=1}^n \zeta_i(\lambda) \ge \frac{x^2}{2}\}} \big]$$

$$= \mathbf{E}_\lambda \big[\exp\{ -Y_n(\lambda) - B_n(\lambda) + \Psi_n(\lambda) \} \mathbf{1}_{\{Y_n(\lambda) \ge \frac{x^2}{2} - B_n(\lambda)\}} \big].$$

Using Lemmas 3.5, 3.4 and 3.10, we get

$$\mathbf{P}(W_n \ge x) \ge \mathbf{E}_\lambda \left[\exp\left\{ -Y_n(\lambda) - \left(\frac{1}{2} \lambda^2 \langle M \rangle_n + c_1 \lambda^{2+\rho} \varepsilon_n^\rho \right) \right\} \right.$$

$$\left. \times \mathbf{1}_{\{Y_n(\lambda) \ge \frac{x^2}{2} - (\frac{1}{2} \lambda^2 \langle M \rangle_n - c_1 \lambda^{2+\rho} \varepsilon_n^\rho)\}} \right].$$

Condition (A1) implies that

$$\left|\langle M\rangle_n - 1\right| \leq \delta_n^2,$$

and thus

$$\mathbf{P}(W_n \geq x) \geq \mathbf{E}_\lambda\left[\exp\left\{-Y_n(\lambda) - \left(\frac{1}{2}\lambda^2 + c_1\lambda^{2+\rho}\varepsilon_n^\rho\right)(1 + \delta_n^2)\right\}\right.$$

$$\left. \times \mathbf{1}_{\{Y_n(\lambda) \geq \frac{x^2}{2} - (\frac{1}{2}\lambda^2(1-\delta_n^2) - c_1\lambda^{2+\rho}\varepsilon_n^\rho)\}}\right]. \tag{4.1}$$

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

$$\frac{1}{2}\lambda^2(1 - \delta_n^2) - c_1\lambda^{2+\rho}\varepsilon_n^\rho = \frac{x^2}{2}.$$

The definition of $\overline{\lambda}$ implies that for $0 \leq x = o(\gamma_n^{-1})$,

$$x \leq \overline{\lambda} \leq c_2 \frac{x}{\sqrt{1 - \delta_n^2}} \tag{4.2}$$

and

$$\overline{\lambda} = x + c_3\theta_0\left(x^{1+\rho}\varepsilon_n^\rho + x\delta_n^2\right), \tag{4.3}$$

where $0 \leq \theta_0 \leq 1$. From (4.1), we obtain

$$\mathbf{P}(W_n \geq x) \geq \exp\left\{-\left(\frac{1}{2}\overline{\lambda}^2 + c_1\overline{\lambda}^{2+\rho}\varepsilon_n^\rho\right)(1 + \delta_n^2)\right\}\mathbf{E}_{\overline{\lambda}}\left[e^{-Y_n(\overline{\lambda})}\mathbf{1}_{\{Y_n(\overline{\lambda})\geq 0\}}\right]. \tag{4.4}$$

Setting $F_n(y) = \mathbf{P}_{\overline{\lambda}}(Y_n(\overline{\lambda}) \leq y)$, we get

$$\mathbf{P}(W_n \geq x) \geq \exp\left\{-c_4\left(\overline{\lambda}^2\delta_n^2 + \overline{\lambda}^{2+\rho}\varepsilon_n^\rho\right) - \frac{\overline{\lambda}^2}{2}\right\}\int_0^\infty e^{-y}\,dF_n(y). \tag{4.5}$$

By integration by parts, we have the following bound:

$$\int_0^\infty e^{-y}\,dF_n(y) \geq \int_0^\infty e^{-y}\,d\Phi(y/\overline{\lambda}) - 2\sup_y\left|F_n(y) - \Phi(y/\overline{\lambda})\right|. \tag{4.6}$$

We distinguish two cases according to the values of ρ.

Case 1: $\rho \in (0, 1)$. Combining (4.5) and (4.6), by Proposition 3.1, we have for $0 \leq x = o(\gamma_n^{-1})$,

$$\mathbf{P}(W_n \geq x) \geq \exp\left\{-c_4\left(\overline{\lambda}^2\delta_n^2 + \overline{\lambda}^{2+\rho}\varepsilon_n^\rho\right) - \frac{\overline{\lambda}^2}{2}\right\}$$

$$\times \left(\int_0^\infty e^{-\overline{\lambda}y}\,d\Phi(y) - c_{1,\rho}\left(\overline{\lambda}^\rho\gamma_n^\rho + \gamma_n^\rho + \delta_n\right)\right). \tag{4.7}$$

Because

$$e^{-\lambda^2/2} \int_0^\infty e^{-\lambda y} \, d\Phi(y) = 1 - \Phi(\lambda) \tag{4.8}$$

and

$$\frac{1}{1+\lambda} e^{-\lambda^2/2} \leq \sqrt{2\pi}\left(1 - \Phi(\lambda)\right), \quad \lambda \geq 0, \tag{4.9}$$

we obtain the following lower bound

$$\frac{\mathbf{P}(W_n \geq x)}{1 - \Phi(\bar{\lambda})} \geq \exp\left\{-c_4\left(\bar{\lambda}^2 \delta_n^2 + \bar{\lambda}^{2+\rho} \varepsilon_n^\rho\right)\right\}\left(1 - c_{2,\rho}(1+\bar{\lambda})\left(\bar{\lambda}^\rho \gamma_n^\rho + \gamma_n^\rho + \delta_n\right)\right)$$

$$\geq \exp\left\{-c_{3,\rho}\left(\bar{\lambda}^2 \delta_n^2 + \bar{\lambda}^{2+\rho} \varepsilon_n^\rho + (1+\bar{\lambda})\left(\bar{\lambda}^\rho \gamma_n^\rho + \gamma_n^\rho + \delta_n\right)\right)\right\}, \tag{4.10}$$

for $0 \leq \bar{\lambda} \leq \frac{1}{2c_{2,\rho}} \min\{\gamma_n^{-\rho/(1+\rho)}, \delta_n^{-1}\}$.

Next, we consider the case of $\frac{1}{2c_{2,\rho}} \min\{\gamma_n^{-\rho/(1+\rho)}, \delta_n^{-1}\} \leq \bar{\lambda} = o(\gamma_n^{-1})$. Let $K \geq 1$ be an absolute constant, whose exact value is chosen later. It is easy to see that

$$\mathbf{E}_{\bar{\lambda}}\left[e^{-Y_n(\bar{\lambda})} \mathbf{1}_{\{Y_n(\bar{\lambda}) \geq 0\}}\right] \geq \mathbf{E}_{\bar{\lambda}}\left[e^{-Y_n(\bar{\lambda})} \mathbf{1}_{\{0 \leq Y_n(\bar{\lambda}) \leq \bar{\lambda} K \tau\}}\right]$$

$$\geq e^{-\bar{\lambda} K \tau} \mathbf{P}_{\bar{\lambda}}\left(0 \leq Y_n(\bar{\lambda}) \leq \bar{\lambda} K \tau\right), \tag{4.11}$$

where $\tau = \bar{\lambda}^\rho \gamma_n^\rho + \delta_n$. By Proposition 3.1, we have

$$\mathbf{P}_{\bar{\lambda}}\left(0 \leq Y_n(\bar{\lambda}) \leq \bar{\lambda} K \tau\right) \geq \mathbf{P}\left(0 \leq \mathcal{N}(0, 1) \leq K \tau\right) - c_{4,\rho} \tau$$

$$\geq \frac{1}{\sqrt{2\pi}} K \tau e^{-K^2 \tau^2/2} - c_{4,\rho} \tau$$

$$\geq \left(\frac{1}{3} K - c_{4,\rho}\right) \tau.$$

Letting $K \geq 12 c_{4,\rho}$, it follows that

$$\mathbf{P}_{\bar{\lambda}}\left(0 \leq Y_n(\bar{\lambda}) \leq \bar{\lambda} K \tau\right) \geq \frac{1}{4} K \tau = \frac{1}{4} K \frac{\bar{\lambda}^{1+\rho} \gamma_n^\rho + \bar{\lambda} \delta_n}{\bar{\lambda}}.$$

Choosing

$$K = \max\left\{12 c_{4,\rho}, \frac{4}{\sqrt{\pi}}(2 c_{2,\rho})^{1+\rho}\right\}$$

and taking into account that $\frac{1}{2c_{2,\rho}} \min\{\gamma_n^{-\rho/(1+\rho)}, \delta^{-1}\} \leq \bar{\lambda} = o(\gamma_n^{-1})$, we conclude that

$$\mathbf{P}_{\bar{\lambda}}\left(0 \leq Y_n(\bar{\lambda}) \leq \bar{\lambda} K \tau\right) \geq \frac{1}{\sqrt{\pi} \bar{\lambda}}.$$

Because the inequality $\frac{1}{\sqrt{\pi}\lambda}e^{-\lambda^2/2} \geq 1 - \Phi(\lambda)$ is valid for all $\lambda \geq 1$, it follows that for $\frac{1}{2c_{2,\rho}}\min\{\gamma_n^{-\rho/(1+\rho)}, \delta^{-1}\} \leq \bar{\lambda} = o(\gamma_n^{-1})$,

$$\mathbf{P}_{\bar{\lambda}}(0 \leq Y_n(\bar{\lambda}) \leq K\tau) \geq (1 - \Phi(\bar{\lambda}))e^{\bar{\lambda}^2/2}. \tag{4.12}$$

Combining (4.4), (4.11), and (4.12), we obtain

$$\frac{\mathbf{P}(W_n \geq x)}{1 - \Phi(\bar{\lambda})} \geq \exp\{-c_{5,\rho}(\bar{\lambda}^2\delta_n^2 + \bar{\lambda}^{2+\rho}\varepsilon_n^\rho + (1 + \bar{\lambda})(\bar{\lambda}^\rho\gamma_n^\rho + \gamma_n^\rho + \delta_n))\}, \tag{4.13}$$

which is valid for $\frac{1}{2c_{2,\rho}}\min\{\gamma_n^{-\rho/(1+\rho)}, \delta^{-1}\} \leq \bar{\lambda} = o(\gamma_n^{-1})$.

From (4.10) and (4.13), we get for $0 \leq \bar{\lambda} = o(\gamma_n^{-1})$,

$$\frac{\mathbf{P}(W_n \geq x)}{1 - \Phi(\bar{\lambda})} \geq \exp\{-c_{6,\rho}(\bar{\lambda}^2\delta_n^2 + \bar{\lambda}^{2+\rho}\varepsilon_n^\rho + (1 + \bar{\lambda})(\bar{\lambda}^\rho\gamma_n^\rho + \gamma_n^\rho + \delta_n))\}. \tag{4.14}$$

Next, we substitute x for $\bar{\lambda}$ in the tail of the normal law $1 - \Phi(\bar{\lambda})$. By (4.2), (4.3), and (4.9), we get

$$\begin{aligned}
1 \leq \frac{\int_{\bar{\lambda}}^\infty \exp\{-t^2/2\}\,dt}{\int_x^\infty \exp\{-t^2/2\}\,dt} &\leq 1 + \frac{\int_{\bar{\lambda}}^x \exp\{-t^2/2\}\,dt}{\int_x^\infty \exp\{-t^2/2\}\,dt} \\
&\leq 1 + c_1 x(x - \bar{\lambda})\exp\{(x^2 - \bar{\lambda}^2)/2\} \\
&\leq \exp\{c_2(x^2\delta_n^2 + x^{2+\rho}\varepsilon_n^\rho)\}
\end{aligned} \tag{4.15}$$

and hence

$$1 - \Phi(\bar{\lambda}) = (1 - \Phi(x))\exp\{\theta_1 c(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2)\}. \tag{4.16}$$

Implementing (4.16) in (4.14) and using (4.2), we obtain for $0 \leq x = o(\gamma_n^{-1})$,

$$\frac{\mathbf{P}(W_n \geq x)}{1 - \Phi(x)} \geq \exp\{-c_{7,\rho}(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2 + (1 + x)(x^\rho\gamma_n^\rho + \gamma_n^\rho + \delta_n))\},$$

which gives the desired lower bound (2.7).

Case 2: $\rho = 1$. Using Proposition 3.1 with $\rho = 1$, we have for $0 \leq x = o(\gamma_n^{-1})$,

$$\begin{aligned}
\mathbf{P}(W_n \geq x) \geq \exp\Big\{&-c_1(\bar{\lambda}^2\delta_n^2 + \bar{\lambda}^3\varepsilon_n) - \frac{\bar{\lambda}^2}{2}\Big\} \\
&\times \Big(\int_0^\infty e^{-\bar{\lambda}y}\,d\Phi(y) - c_2(\bar{\lambda}\gamma_n + \gamma_n|\ln\gamma_n| + \delta_n)\Big),
\end{aligned}$$

that is, the term γ_n^ρ in inequality (4.7) has been replaced by $\gamma_n|\ln\gamma_n|$. By an argument similar to that of *Case 1*, we obtain the desired lower bound (2.8).

4.2. Proof of Theorem 2.3

We first prove Theorem 2.3 for $1 \le x = o(\gamma_n^{-1})$. Observe that

$$
\begin{aligned}
\mathbf{P}(W_n \ge x) &= \mathbf{P}\big(W_n \ge x, \big|[M]_n - \langle M \rangle_n\big| \le \delta_n + 1/(2x)\big) \\
&+ \mathbf{P}\big(W_n \ge x, \big|[M]_n - \langle M \rangle_n\big| > \delta_n + 1/(2x)\big).
\end{aligned}
\tag{4.17}
$$

For the the first term on the right-hand side of (4.17), by (3.2) and (3.5) with $\lambda = x$, we have the following representation:

$$
\begin{aligned}
&\mathbf{P}\big(W_n \ge x, \big|[M]_n - \langle M \rangle_n\big| \le \delta_n + 1/(2x)\big) \\
&= \mathbf{E}_x\big[Z_n(x)^{-1}\mathbf{1}_{\{M_n \ge x\sqrt{[M]_n},\,|[M]_n - \langle M \rangle_n| \le \delta_n + 1/(2x)\}}\big] \\
&= \mathbf{E}_x\big[e^{-Y_n(x) - B_n(x) + \Psi_n(x)}\mathbf{1}_{\{xM_n \ge x^2\sqrt{1 + [M]_n - 1},\,|[M]_n - \langle M \rangle_n| \le \delta_n + 1/(2x)\}}\big].
\end{aligned}
$$

By the inequality

$$
\sqrt{1 + y} \ge 1 + y/2 - y^2/2, \quad y \ge -1,
$$

condition (A1) and Lemma 3.4, we have for $1 \le x = o(\gamma_n^{-1})$,

$$
\begin{aligned}
&\mathbf{P}\big(W_n \ge x, \big|[M]_n - \langle M \rangle_n\big| \le \delta_n + 1/(2x)\big) \\
&\le \mathbf{E}_x\big[\exp\{-Y_n(x) - B_n(x) + \Psi_n(x)\} \\
&\quad \times \mathbf{1}_{\{xM_n - \frac{1}{2}x^2[M]_n + \frac{1}{2}x^2([M]_n - 1)^2 \ge \frac{1}{2}x^2,\,|[M]_n - \langle M \rangle_n| \le \delta_n + 1/(2x)\}}\big] \\
&\le \mathbf{E}_x\big[\exp\{-Y_n(x) - B_n(x) + \Psi_n(x)\} \\
&\quad \times \mathbf{1}_{\{xM_n - \frac{1}{2}x^2[M]_n + x^2([M]_n - \langle M \rangle_n)^2 + x^2(1 - \langle M \rangle_n)^2 \ge \frac{1}{2}x^2,\,|[M]_n - \langle M \rangle_n| \le \delta_n + 1/(2x)\}}\big] \\
&\le \mathbf{E}_x\big[\exp\{-Y_n(x) - B_n(x) + \Psi_n(x)\} \\
&\quad \times \mathbf{1}_{\{Y_n(x) \ge -x^2([M]_n - \langle M \rangle_n)^2 - x^2\delta_n^4 + \frac{1}{2}x^2 - B_n(x),\,|[M]_n - \langle M \rangle_n| \le \delta_n + 1/(2x)\}}\big].
\end{aligned}
$$

Thus, for $1 \le x = o(\gamma_n^{-1})$,

$$
\begin{aligned}
&\mathbf{P}\big(W_n \ge x, \big|[M]_n - \langle M \rangle_n\big| \le \delta_n + 1/(2x)\big) \\
&\le \mathbf{E}_x\big[\exp\{-Y_n(x) - B_n(x) + \Psi_n(x)\} \\
&\quad \times \mathbf{1}_{\{Y_n(x) \ge -x^{2+\rho}\varepsilon_n^\rho - x^2\delta_n^4 + \frac{1}{2}x^2 - B_n(x),\,|[M]_n - \langle M \rangle_n| \le (x\varepsilon_n)^{\rho/2}\}}\big] \\
&+ \mathbf{E}_x\big[\exp\{-Y_n(x) - B_n(x) + \Psi_n(x)\} \\
&\quad \times \mathbf{1}_{\{0 > Y_n(x) \ge -x^2([M]_n - \langle M \rangle_n)^2 - x^2\delta_n^4 + \frac{1}{2}x^2 - B_n(x),\,(x\varepsilon_n)^{\rho/2} < |[M]_n - \langle M \rangle_n| \le \delta_n + 1/(2x)\}}\big]. \\
&\le I_1(x) + I_2(x),
\end{aligned}
\tag{4.18}
$$

where

$$I_1(x) = \mathbf{E}_x\Big[\exp\big\{-Y_n(x) - B_n(x) + \Psi_n(x)\big\}\mathbf{1}_{\{Y_n(x)\geq -c_1(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2)\}}\Big]$$

and

$$I_2(x) = \mathbf{E}_x\Big[\exp\big\{-Y_n(x) - B_n(x) + \Psi_n(x)\big\}$$
$$\times \mathbf{1}_{\{0 > Y_n(x)\geq -1 - c_2(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2),\, (x\varepsilon_n)^{\rho/2} < |[M]_n - \langle M\rangle_n| \leq \delta_n + 1/(2x)\}}\Big].$$

For $I_1(x)$, by an argument similar to the proof of Theorem 2.2, we get for $1 \leq x = o(\gamma_n^{-1})$,

$$\frac{I_1(x)}{1 - \Phi(x)} \leq \begin{cases} \exp\big\{c_\rho\big(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2 + (1+x)\big(x^\rho\gamma_n^\rho + \gamma_n^\rho + \delta_n\big)\big)\big\} & \text{if } \rho \in (0,1), \\ \exp\big\{c\big(x^3\varepsilon_n + x^2\delta_n^2 + (1+x)\big(x\gamma_n + \gamma_n|\ln\gamma_n| + \delta_n\big)\big)\big\} & \text{if } \rho = 1. \end{cases} \tag{4.19}$$

Next, consider the item $I_2(x)$. By condition (A1), Lemmas 3.4 and 3.10, it is obvious that for $1 \leq x = o(\gamma_n^{-1})$,

$$I_2(x) \leq \exp\Big\{-\frac{1}{2}x^2 + c_1\big(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2\big)\Big\}$$
$$\times \mathbf{E}_x\Big[e^{-Y_n(x)}\mathbf{1}_{\{0 > Y_n(x)\geq -1 - c_2(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2),\, (x\varepsilon_n)^{\rho/2} < |[M]_n - \langle M\rangle_n|\}}\Big]$$
$$\leq \exp\Big\{-\frac{1}{2}x^2 + c_1\big(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2\big)\Big\} \tag{4.20}$$
$$\times \mathbf{E}_x\Big[e^{1+c_2(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2)}\mathbf{1}_{\{(x\varepsilon_n)^{\rho/2} < |[M]_n - \langle M\rangle_n|\}}\Big]$$
$$\leq \exp\Big\{1 - \frac{1}{2}x^2 + c_3\big(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2\big)\Big\}\mathbf{E}_x\big[\mathbf{1}_{\{(x\varepsilon_n)^{\rho/2} < |[M]_n - \langle M\rangle_n|\}}\big].$$

Denote by $\langle M(x)\rangle_n = \sum_{i=1}^n \mathbf{E}_x[\xi_i^2|\mathcal{F}_{i-1}]$. Notice that $\varepsilon_n = O(\gamma_n)$. From (3.4), using (3.10), Lemmas 3.3, 3.5 and condition (A2), we obtain for $1 \leq x = o(\gamma_n^{-1})$,

$$\big|\langle M(x)\rangle_n - \langle M\rangle_n\big|$$
$$\leq \sum_{i=1}^n \left|\frac{\mathbf{E}[\xi_i^2 e^{x\xi_i - x^2\xi_i^2/2}|\mathcal{F}_{i-1}]}{\mathbf{E}[e^{x\xi_i - x^2\xi_i^2/2}|\mathcal{F}_{i-1}]} - \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}]\right| + \sum_{i=1}^n\left(\frac{\mathbf{E}[\xi_i e^{x\xi_i - x^2\xi_i^2/2}|\mathcal{F}_{i-1}]^2}{\mathbf{E}[e^{x\xi_i - x^2\xi_i^2/2}|\mathcal{F}_{i-1}]^2}\right)$$
$$\leq c_4 \sum_{i=1}^n \big(\mathbf{E}[x^\rho|\xi_i|^{2+\rho}|\mathcal{F}_{i-1}] + \big(\mathbf{E}[x\xi_i^2|\mathcal{F}_{i-1}]\big)^2\big) \tag{4.21}$$
$$\leq c_4 \sum_{i=1}^n \big(\mathbf{E}[x^\rho|\xi_i|^{2+\rho}|\mathcal{F}_{i-1}] + x^2\mathbf{E}[|\xi_i|^{2+\rho}|\mathcal{F}_{i-1}]\big(\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}]\big)^{(2-\rho)/2}\big)$$
$$\leq c_5 x^\rho \varepsilon_n^\rho.$$

Thus, for $1 \leq x = o(\gamma_n^{-1})$,

$$
I_2(x) \leq \exp\left\{1 - \frac{1}{2}x^2 + c_3\left(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2\right)\right\}\mathbf{E}_x[\mathbf{1}_{\{\frac{1}{2}(x\varepsilon_n)^{\rho/2} < |[M]_n - \langle M(x)\rangle_n|\}}]
$$

$$
\leq \frac{4e}{(x\varepsilon_n)^{\rho(2+\rho)/4}}\exp\left\{-\frac{1}{2}x^2 + c_3\left(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2\right)\right\}\mathbf{E}_x\left[\left|[M]_n - \langle M(x)\rangle_n\right|^{(2+\rho)/2}\right].
$$

It is obvious that

$$
[M]_n - \langle M(x)\rangle_n = \sum_{i=1}^n(\xi_i^2 - \mathbf{E}_x[\xi_i^2|\mathcal{F}_{i-1}]).
$$

Thus, $([M]_i - \langle M(x)\rangle_i, \mathcal{F}_i)_{i=0,\ldots,n}$ is a martingale with respect to the probability measure $\mathbf{P}_x$. By the inequality of von Bahr and Esseen [32], it follows that for $1 \leq x = o(\gamma_n^{-1})$,

$$
\begin{aligned}
\mathbf{E}_x\left[\left|[M]_n - \langle M(x)\rangle_n\right|^{(2+\rho)/2}\right] &\leq c_6\sum_{i=1}^n\mathbf{E}_x\left[\left|\xi_i^2 - \mathbf{E}_x[\xi_i^2|\mathcal{F}_{i-1}]\right|^{(2+\rho)/2}\right]\\
&\leq c_7\sum_{i=1}^n\mathbf{E}_x\left[|\xi_i|^{2+\rho}\right]\\
&= c_7\sum_{i=1}^n\frac{\mathbf{E}[|\xi_i|^{2+\rho}e^{\zeta_i(x)}|\mathcal{F}_{i-1}]}{\mathbf{E}[e^{\zeta_i(x)}|\mathcal{F}_{i-1}]}\\
&\leq c_8\varepsilon_n^\rho.
\end{aligned}
\tag{4.22}
$$

Hence, for $1 \leq x = o(\gamma_n^{-1})$,

$$
I_2(x) \leq c\frac{\varepsilon_n^{\rho(2-\rho)/4}}{x^{\rho(2+\rho)/4}}\exp\left\{-\frac{1}{2}x^2 + c_3\left(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2\right)\right\}.
\tag{4.23}
$$

Next, we give an estimation for $\mathbf{P}(W_n \geq x, |[M]_n - \langle M\rangle_n| > \delta_n + 1/(2x))$. Since $|1 - \langle M\rangle_n| \leq \delta_n^2 \leq \delta_n/2$, it is obvious that

$$
\begin{aligned}
&\mathbf{P}\left(W_n \geq x, \left|[M]_n - \langle M\rangle_n\right| > \delta_n + 1/(2x)\right)\\
&\quad\leq \mathbf{P}\left(W_n \geq x, \left|[M]_n - 1\right| + \left|1 - \langle M\rangle_n\right| > \delta_n + 1/(2x)\right)\\
&\quad\leq \mathbf{P}\left(W_n \geq x, \left|[M]_n - 1\right| > \delta_n/2 + 1/(2x)\right).
\end{aligned}
$$

To estimate the tail probability in the last line, we follow the argument of Shao and Zhou [29]. We have the following decomposition:

$$
\begin{aligned}
&\mathbf{P}\left(W_n \geq x, \left|[M]_n - 1\right| > \delta_n/2 + 1/(2x)\right)\\
&\quad\leq \mathbf{P}\left(M_n/\sqrt{[M]_n} \geq x, 1 + \delta_n/2 + 1/(2x) < [M]_n \leq 16\right)
\end{aligned}
$$

$$+ \mathbf{P}\big(M_n/\sqrt{[M]_n} \geq x, [M]_n < 1 - \delta_n/2 - 1/(2x)\big) \qquad (4.24)$$

$$+ \mathbf{P}\big(M_n/\sqrt{[M]_n} \geq x, [M]_n > 16\big)$$

$$:= \sum_{v=1}^{3} \mathbf{P}\big((M_n, \sqrt{[M]_n}) \in \mathcal{E}_v\big),$$

where $\mathcal{E}_v \subset \mathbf{R} \times \mathbf{R}^+$, $1 \leq v \leq 3$, are given by

$$\mathcal{E}_1 = \big\{(u, v) \in \mathbf{R} \times \mathbf{R}^+ : u/v \geq x, \sqrt{1 + \delta_n/2 + 1/(2x)} < v \leq 4\big\},$$

$$\mathcal{E}_2 = \big\{(u, v) \in \mathbf{R} \times \mathbf{R}^+ : u/v \geq x, v < \sqrt{1 - \delta_n/2 - 1/(2x)}\big\},$$

$$\mathcal{E}_3 = \big\{(u, v) \in \mathbf{R} \times \mathbf{R}^+ : u/v \geq x, v > 4\big\}.$$

To estimate the probability $\mathbf{P}((M_n, \sqrt{[M]_n}) \in \mathcal{E}_1)$, we introduce the following new conjugate probability measure $\widetilde{\mathbf{P}}_x$ defined by

$$d\widetilde{\mathbf{P}}_x = \widetilde{Z}_n(x)\, d\mathbf{P},$$

where

$$\widetilde{Z}_n(x) = \prod_{i=1}^{k} \frac{e^{\widetilde{\zeta}_i(x)}}{\mathbf{E}[e^{\widetilde{\zeta}_i(x)}|\mathcal{F}_{i-1}]} \quad \text{and} \quad \widetilde{\zeta}_i(x) = x\xi_i - x^2\xi_i^2/8.$$

Denote by $\widetilde{\mathbf{E}}_x$ the expectation with respect to $\widetilde{\mathbf{P}}_x$ and $\langle \widetilde{M}(x)\rangle_n = \sum_{i=1}^{n} \widetilde{\mathbf{E}}_x[\xi_i^2|\mathcal{F}_{i-1}]$. By an argument similar to (4.21), it follows that for $1 \leq x = o(\gamma_n^{-1})$,

$$\big|\langle \widetilde{M}(x)\rangle_n - \langle M\rangle_n\big| \leq cx^\rho \varepsilon_n^\rho.$$

By Markov's inequality, we deduce that

$$\mathbf{P}\big((M_n, \sqrt{[M]_n}) \in \mathcal{E}_1\big)$$

$$\leq \big(\delta_n/2 + 1/(2x)\big)^{-2} e^{-\inf_{(u,v)\in\mathcal{E}_1}(xu-(vx)^2/8)} \mathbf{E}\big[([M]_n - 1)^2 e^{xM_n - [M]_n x^2/8}\big]$$

$$\leq 16x^2 e^{-\inf_{(u,v)\in\mathcal{E}_1}(xu-(vx)^2/8)} \mathbf{E}\big[\big([M]_n - \langle \widetilde{M}(x)\rangle_n\big)^2 e^{xM_n - [M]_n x^2/8}\big]$$

$$\quad + 16x^2 e^{-\inf_{(u,v)\in\mathcal{E}_1}(xu-(vx)^2/8)} \mathbf{E}\big[\big(\langle \widetilde{M}(x)\rangle_n - \langle M\rangle_n\big)^2 e^{xM_n - [M]_n x^2/8}\big]$$

$$\quad + 16\delta_n^{-2} e^{-\inf_{(u,v)\in\mathcal{E}_1}(xu-(vx)^2/8)} \mathbf{E}\big[\big(\langle M\rangle_n - 1\big)^2 e^{xM_n - [M]_n x^2/8}\big] \qquad (4.25)$$

$$\leq 16x^2 e^{-\inf_{(u,v)\in\mathcal{E}_1}(xu-(vx)^2/8)} \mathbf{E}\big[\big([M]_n - \langle \widetilde{M}(x)\rangle_n\big)^2 e^{xM_n - [M]_n x^2/8}\big]$$

$$\quad + cx^{2+2\rho}\varepsilon_n^{2\rho} e^{-\inf_{(u,v)\in\mathcal{E}_1}(xu-(vx)^2/8)} \mathbf{E}\big[e^{xM_n - [M]_n x^2/8}\big]$$

$$\quad + 16\delta_n^2 e^{-\inf_{(u,v)\in\mathcal{E}_1}(xu-(vx)^2/8)} \mathbf{E}\big[e^{xM_n - [M]_n x^2/8}\big],$$

where it is easy to verify that

$$\inf_{(u,v)\in\mathcal{E}_1}\left(xu - \frac{1}{8}(vx)^2\right) \geq \frac{7}{8}x^2 + \frac{1}{4}x - cx^2\delta_n^2. \tag{4.26}$$

By Lemma 3.1, conditions (A1) and (A2), it follows that

$$\prod_{i=1}^{n}\mathbf{E}\big[e^{\widetilde{\zeta}_i(x)}|\mathcal{F}_{i-1}\big] \leq \prod_{i=1}^{n}\left(1 + \frac{3}{8}x^2\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] + cx^{2+\rho}\mathbf{E}\big[|\xi_i|^{2+\rho}|\mathcal{F}_{i-1}\big]\right)$$

$$\leq \prod_{i=1}^{n}\exp\left\{\frac{3}{8}x^2\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] + cx^{2+\rho}\mathbf{E}\big[|\xi_i|^{2+\rho}|\mathcal{F}_{i-1}\big]\right\}$$

$$= \exp\left\{\frac{3}{8}x^2\langle M\rangle_n + cx^{2+\rho}\sum_{i=1}^{n}\mathbf{E}\big[|\xi_i|^{2+\rho}|\mathcal{F}_{i-1}\big]\right\}$$

$$\leq \exp\left\{\frac{3}{8}x^2 + c\big(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2\big)\right\}.$$

Therefore, for $1 \leq x = o(\gamma_n^{-1})$,

$$\mathbf{E}\big[\big([M]_n - \langle\widetilde{M}(x)\rangle_n\big)^2 e^{xM_n - [M]_n x^2/8}\big]$$

$$= \mathbf{E}\left[\left(\prod_{i=1}^{n}\mathbf{E}\big[e^{\widetilde{\zeta}_i(x)}|\mathcal{F}_{i-1}\big]\right)\big([M]_n - \langle\widetilde{M}(x)\rangle_n\big)^2\widetilde{Z}_n(x)\right]$$

$$\leq \mathbf{E}\big[\big([M]_n - \langle\widetilde{M}(x)\rangle_n\big)^2\widetilde{Z}_n(x)\big]\exp\left\{\frac{3}{8}x^2 + c\big(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2\big)\right\}$$

$$= \widetilde{\mathbf{E}}_x\big[\big([M]_n - \langle\widetilde{M}(x)\rangle_n\big)^2\big]\exp\left\{\frac{3}{8}x^2 + c\big(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2\big)\right\}$$

$$= \sum_{i=1}^{n}\widetilde{\mathbf{E}}_x\big[\big(\xi_i^2 - \widetilde{\mathbf{E}}_x[\xi_i^2|\mathcal{F}_{i-1}]\big)^2\big]\exp\left\{\frac{3}{8}x^2 + c\big(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2\big)\right\},$$

where the last line follows because $([M]_i - \langle\widetilde{M}(x)\rangle_i, \mathcal{F}_i)_{i=0,\dots,n}$ is a martingale with respect to the probability measure $\widetilde{\mathbf{P}}_x$. Therefore, by Lemma 3.1, conditions (A1) and (A2) again, we have for $1 \leq x = o(\gamma_n^{-1})$,

$$\mathbf{E}\big[\big([M]_n - \langle\widetilde{M}(x)\rangle_n\big)^2 e^{xM_n - [M]_n x^2/8}\big]$$

$$\leq \sum_{i=1}^{n}\widetilde{\mathbf{E}}_x\big[\widetilde{\mathbf{E}}_x[\xi_i^4|\mathcal{F}_{i-1}]\big]\exp\left\{\frac{3}{8}x^2 + c\big(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2\big)\right\}$$

$$= \sum_{i=1}^{n} \widetilde{\mathbf{E}}_x \left[\mathbf{E}\big[\xi_i^4 e^{\widetilde{\zeta}_i(x)}|\mathcal{F}_{i-1}\big]/\mathbf{E}\big[e^{\widetilde{\zeta}_i(x)}|\mathcal{F}_{i-1}\big]\right] \exp\left\{\frac{3}{8}x^2 + c\big(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2\big)\right\}$$

$$\leq c_0 \sum_{i=1}^{n} \widetilde{\mathbf{E}}_x \left[\frac{1}{x^{2-\rho}} \sum_{i=1}^{n} \mathbf{E}\big[|\xi_i|^{2+\rho}|\mathcal{F}_{i-1}\big]\right] \exp\left\{\frac{3}{8}x^2 + c\big(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2\big)\right\}$$

$$\leq c_1 \varepsilon_n^\rho \exp\left\{\frac{3}{8}x^2 + c\big(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2\big)\right\}.$$

Lemma 3.1 implies that for $1 \leq x = o(\gamma_n^{-1})$,

$$\mathbf{E}\left[\exp\left\{xM_n - \frac{1}{8}x^2[M]_n - \frac{3}{8}x^2\langle M\rangle_n - cx^{2+\rho}\sum_{i=1}^{n}\mathbf{E}\big[|\xi_i|^{2+\rho}|\mathcal{F}_{i-1}\big]\right\}\right]$$

$$\leq \mathbf{E}\left[\exp\left\{xM_{n-1} - \frac{1}{8}x^2[M]_{n-1} - \frac{3}{8}x^2\langle M\rangle_{n-1} - cx^{2+\rho}\sum_{i=1}^{n-1}\mathbf{E}\big[|\xi_i|^{2+\rho}|\mathcal{F}_{i-1}\big]\right\}\right]$$

$$\leq 1.$$

By conditions (A1), (A2) and the last inequality, we obtain for $1 \leq x = o(\gamma_n^{-1})$,

$$\mathbf{E}\big[e^{xM_n - [M]_n x^2/8}\big] \leq \exp\left\{\frac{3}{8}x^2 + c\big(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2\big)\right\}.$$

Thus, from (4.25), we deduce that for $1 \leq x = o(\gamma_n^{-1})$,

$$\mathbf{P}\big((M_n, \sqrt{[M]_n}) \in \mathcal{E}_1\big)$$
$$\leq c_2\big(\varepsilon_n^\rho + x^{2+2\rho}\varepsilon_n^{2\rho} + \delta_n^2\big)\exp\left\{-\frac{1}{2}x^2 - \frac{1}{4}x + c\big(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2\big)\right\} \qquad (4.27)$$
$$\leq c_3\big(\varepsilon_n^\rho + \delta_n^2\big)\exp\left\{-\frac{1}{2}x^2 + c\big(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2\big)\right\}.$$

Similarly, we have

$$\mathbf{P}\big((M_n, \sqrt{[M]_n}) \in \mathcal{E}_2\big)$$
$$\leq \big(\delta_n/2 + 1/(2x)\big)^{-2} e^{-\inf_{(u,v)\in\mathcal{E}_2}(xu - 2(vx)^2)}\mathbf{E}\big[([M]_n - 1)^2 e^{xM_n - 2[M]_n x^2}\big] \qquad (4.28)$$
$$\leq c_4\big(\varepsilon_n^\rho + \delta_n^2\big)\exp\left\{-\frac{1}{2}x^2 + c\big(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2\big)\right\}.$$

For the last term $\mathbf{P}((M_n, \sqrt{[M]_n}) \in \mathcal{E}_3)$, we obtain the following estimation

$$
\begin{aligned}
\mathbf{P}\big((M_n, \sqrt{[M]_n}) \in \mathcal{E}_3\big) &= \mathbf{P}\big(M_n \geq x\sqrt{[M]_n}, [M]_n > 16\big) \\
&\leq \frac{2}{3}x^{-2/3}\exp\left\{-\frac{3}{4}x^2\right\},
\end{aligned}
\tag{4.29}
$$

where the last line follows by Lemma 3.8. Moreover, by Lemma 3.9, it holds that for $\rho \in (0, 1]$,

$$
\begin{aligned}
\mathbf{P}\big((M_n, \sqrt{[M]_n}) \in \mathcal{E}_3\big) &\leq \mathbf{P}\big(\big|[M]_n - \langle M\rangle_n\big| \geq 1\big) \\
&\leq c\varepsilon_n^\rho.
\end{aligned}
$$

By the last inequality and (4.29), we get for $1 \leq x = o(\gamma_n^{-1})$,

$$
\begin{aligned}
\mathbf{P}\big((M_n, \sqrt{[M]_n}) \in \mathcal{E}_3\big) &\leq \min\left\{c\varepsilon_n^\rho, \frac{2}{3}x^{-2/3}e^{-3x^2/4}\right\} \\
&\leq c\frac{\varepsilon_n^{\rho(2-\rho)/4}}{x^{\rho(2+\rho)/4}}\exp\left\{-\frac{1}{2}x^2\right\}.
\end{aligned}
\tag{4.30}
$$

Thus, combining the inequalities (4.24), (4.27), (4.28) and (4.30) together, we deduce that for $1 \leq x = o(\gamma_n^{-1})$,

$$
\begin{aligned}
&\mathbf{P}\big(W_n \geq x, \big|[M]_n - \langle M\rangle_n\big| > \delta_n + 1/(2x)\big) \\
&\leq c\left(\frac{\varepsilon_n^{\rho(2-\rho)/4}}{x^{\rho(2+\rho)/4}} + \delta_n^2\right)\exp\left\{-\frac{1}{2}x^2 + c\big(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2\big)\right\}.
\end{aligned}
\tag{4.31}
$$

Combining (4.18), (4.19), (4.23), and (4.31), we obtain for $1 \leq x = o(\gamma_n^{-1})$,

$$
\begin{aligned}
\frac{\mathbf{P}(W_n \geq x)}{1 - \Phi(x)} &\leq \left(1 + c_1(1 + x)\left(\frac{\varepsilon_n^{\rho(2-\rho)/4}}{x^{\rho(2+\rho)/4}} + \delta_n^2\right)\right) \\
&\quad \times \begin{cases} \exp\{c_{1,\rho}\big(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2 + (1 + x)\big(x^\rho\gamma_n^\rho + \gamma_n^\rho + \delta_n\big)\big)\} & \text{if } \rho \in (0, 1), \\ \exp\{c_2\big(x^3\varepsilon_n + x^2\delta_n^2 + (1 + x)\big(x\gamma_n + \gamma_n|\ln\gamma_n| + \delta_n\big)\big)\} & \text{if } \rho = 1 \end{cases} \\
&\leq \begin{cases} \exp\left\{c_\rho\left(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2 + (1 + x)\left(x^\rho\gamma_n^\rho + \gamma_n^\rho + \delta_n + \frac{\varepsilon_n^{\rho(2-\rho)/4}}{x^{\rho(2+\rho)/4}}\right)\right)\right\} \\ \quad \text{if } \rho \in (0, 1), \\ \exp\left\{c\left(x^3\varepsilon_n + x^2\delta_n^2 + (1 + x)\left(x\gamma_n + \gamma_n|\ln\gamma_n| + \delta_n + \frac{\varepsilon_n^{\rho(2-\rho)/4}}{x^{\rho(2+\rho)/4}}\right)\right)\right\} \\ \quad \text{if } \rho = 1, \end{cases}
\end{aligned}
$$

which gives the desired inequalities.

For the case $0 \leq x < 1$, the assertion of Theorem 2.3 follows by a similar argument, but with $1/(2x)$ replaced by $1/2$ in (4.17) and $(x\varepsilon_n)^{\rho/2}$ replaced by $\varepsilon_n^{\rho/2}$ in (4.18), and accordingly in the subsequent statements. This completes the proof of Theorem 2.3.

4.3. Proof of Theorem 2.1

Using Proposition 3.2, by an argument similar to the proofs of Theorems 2.2 and 2.3, we obtain the following result. If $\rho \in (0, 1)$, then for $0 \leq x = o(\min\{\varepsilon_n^{-1}, \kappa_n^{-1}\})$,

$$\frac{\mathbf{P}(W_n \geq x)}{1 - \Phi(x)}$$

$$= \exp\left\{\theta c_\rho\left(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2 + (1+x)\left(x^{\rho/2}\varepsilon_n^{\rho/2} + \varepsilon_n^{\rho/(3+\rho)} + \delta_n + \frac{\varepsilon_n^{\rho(2-\rho)/4}}{1 + x^{\rho(2+\rho)/4}}\right)\right)\right\}.$$

Notice that the following three inequalities hold:

$$x^{1+\rho/2}\varepsilon_n^{\rho/2} \leq x^{2+\rho}\varepsilon_n^\rho, \quad x \geq \varepsilon_n^{-\rho/(2+\rho)},$$

$$x^{\rho/2}\varepsilon_n^{\rho/2} \leq \varepsilon_n^{\rho/(3+\rho)}, \quad 0 \leq x \leq \varepsilon_n^{-\rho/(2+\rho)},$$

$$\varepsilon_n^{\rho(2-\rho)/4} \leq \varepsilon_n^{\rho/(3+\rho)}, \quad \rho \in (0, 1].$$

Therefore, for $\rho \in (0, 1)$ and $0 \leq x = o(\min\{\varepsilon_n^{-1}, \kappa_n^{-1}\})$,

$$\frac{\mathbf{P}(W_n \geq x)}{1 - \Phi(x)} = \exp\left\{\theta c_\rho\left(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2 + (1+x)\left(\varepsilon_n^{\rho/(3+\rho)} + \delta_n\right)\right)\right\},$$

which gives the desired equality for $\rho \in (0, 1)$.

Assume that condition (A2) holds for $\rho \geq 1$. When $\rho \in [1, 2]$, by Markov's inequality and (4.22), we have for $x \geq 1$,

$$\mathbf{E}_x[\mathbf{1}_{\{(x\varepsilon_n)^{1/2} < |[M]_n - \langle M \rangle_n|\}}] \leq \frac{1}{(x\varepsilon_n)^{(2+\rho)/4}}\mathbf{E}_x\left[\left|[M]_n - \langle M(x) \rangle_n\right|^{(2+\rho)/2}\right]$$

$$\leq \frac{1}{x^{(2+\rho)/4}}\varepsilon_n^{(3\rho-2)/4} \tag{4.32}$$

$$\leq \varepsilon_n^{(3\rho-2)/4}.$$

When $\rho > 2$, Lemma 3.7 implies that condition (A2) also holds for $\rho = 2$, with the term ε_n in condition (A2) replaced by $2\varepsilon_n$. Then (4.32) with $\rho = 2$ shows that

$$\mathbf{E}_x[\mathbf{1}_{\{(x\varepsilon_n)^{1/2} < |[M]_n - \langle M \rangle_n|\}}] \leq 2\varepsilon_n.$$

Thus, for $\rho \geq 1$, it holds that

$$\mathbf{E}_x[\mathbf{1}_{\{(x\varepsilon_n)^{1/2} < |[M]_n - \langle M \rangle_n|\}}] \leq \max\{\varepsilon_n^{(3\rho-2)/4}, 2\varepsilon_n\} \leq 2\varepsilon_n^{\rho/(3+\rho)}.$$

Notice that Lemma 3.7 also implies that condition (A2) holds for $\rho = 1$. Therefore, by (4.20), (4.23) can be improved to

$$I_2(x) \leq \exp\left\{1 - \frac{1}{2}x^2 + c_3\left(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2\right)\right\}\mathbf{E}_x[\mathbf{1}_{\{(x\varepsilon_n)^{1/2} < |[M]_n - \langle M\rangle_n|\}}]$$

$$\leq c\varepsilon_n^{\rho/(3+\rho)}\exp\left\{-\frac{1}{2}x^2 + c_3\left(x^3\varepsilon_n + x^2\delta_n^2\right)\right\}.$$

Notice also that for $\rho \geq 1$,

$$\mathbf{P}\left((M_n, \sqrt{[M]_n}) \in \mathcal{E}_3\right) \leq \min\left\{c_1\varepsilon_n^\rho, \frac{2}{3}x^{-2/3}e^{-3x^2/4}\right\}$$

$$\leq c_2\varepsilon_n^{\rho/(3+\rho)}\exp\left\{-\frac{1}{2}x^2\right\}.$$

By an argument similar to the proof for case $\rho \in (0, 1)$ but with the term $(x\varepsilon_n)^{\rho/2}$ in (4.18) replaced by $(x\varepsilon_n)^{1/2}$, we have for $0 \leq x = o(\min\{\varepsilon_n^{-1}, \kappa_n^{-1}\})$,

$$\frac{\mathbf{P}(W_n \geq x)}{1 - \Phi(x)} = \exp\left\{\theta c_3\left(x^3\varepsilon_n + x^2\delta_n^2 + (1+x)\left(x^{\rho/2}\varepsilon_n^{\rho/2} + \varepsilon_n^{\rho/(3+\rho)} + \delta_n\right)\right)\right\}$$

$$= \exp\left\{\theta c_4\left(x^3\varepsilon_n + x^2\delta_n^2 + (1+x)\left(\varepsilon_n^{\rho/(3+\rho)} + \delta_n\right)\right)\right\},$$

which gives the desired equality for $\rho \geq 1$.

4.4. Proof of Corollary 2.2

To prove Corollary 2.2, we need the following two sides bound on the tail probabilities of the standard normal random variable:

$$\frac{1}{\sqrt{2\pi}(1+x)}e^{-x^2/2} \leq 1 - \Phi(x) \leq \frac{1}{\sqrt{\pi}(1+x)}e^{-x^2/2}, \quad x \geq 0. \tag{4.33}$$

See page 17 in Itô and McKean [19] or Talagrand [31]. First, we prove that

$$\limsup_{n\to\infty} \frac{1}{a_n^2}\ln\mathbf{P}\left(\frac{W_n}{a_n} \in B\right) \leq -\inf_{x\in\overline{B}}\frac{x^2}{2}. \tag{4.34}$$

For any given Borel set $B \subset \mathbf{R}$, let $x_0 = \inf_{x\in B}|x|$. Then, it is obvious that $x_0 \geq \inf_{x\in\overline{B}}|x|$. Therefore, by Theorem 2.1 and Remark 2.1,

$$\mathbf{P}\left(\frac{W_n}{a_n} \in B\right) \leq \mathbf{P}\left(|W_n| \geq a_n x_0\right)$$

$$\leq 2\left(1 - \Phi(a_n x_0)\right)$$

$$\times \exp\left\{c_\rho\left((a_n x_0)^{2+\rho}\varepsilon_n^\rho + (a_n x_0)^2\delta_n^2 + (a_n x_0)\left(\varepsilon_n^{\rho/(3+\rho)} + \delta_n\right)\right)\right\}.$$

Using (4.33), we deduce that

$$\limsup_{n\to\infty} \frac{1}{a_n^2} \ln \mathbf{P}\left(\frac{W_n}{a_n} \in B\right) \le -\frac{x_0^2}{2} \le -\inf_{x\in\overline{B}} \frac{x^2}{2},$$

which gives (4.34).

Next, we prove that

$$\liminf_{n\to\infty} \frac{1}{a_n^2} \ln \mathbf{P}\left(\frac{W_n}{a_n} \in B\right) \ge -\inf_{x\in B^o} \frac{x^2}{2}. \tag{4.35}$$

We may assume that $B^o \ne \varnothing$. For any $\varepsilon_1 > 0$, there exists an $x_0 \in B^o$, such that

$$0 < \frac{x_0^2}{2} \le \inf_{x\in B^o} \frac{x^2}{2} + \varepsilon_1. \tag{4.36}$$

For $x_0 \in B^o$, there exists small $\varepsilon_2 \in (0, x_0)$, such that $(x_0 - \varepsilon_2, x_0 + \varepsilon_2] \subset B$. Then it is obvious that $x_0 \ge \inf_{x\in\overline{B}} x$. Without loss of generality, we may assume that $x_0 > 0$. By Theorem 2.1, we deduce that

$$\mathbf{P}\left(\frac{W_n}{a_n} \in B\right) \ge \mathbf{P}(W_n \in (a_n(x_0 - \varepsilon_2), a_n(x_0 + \varepsilon_2)])$$

$$\ge \mathbf{P}\big(W_n > a_n(x_0 - \varepsilon_2)\big) - \mathbf{P}\big(W_n > a_n(x_0 + \varepsilon_2)\big).$$

Using Theorem 2.1 and (4.33), it follows that

$$\liminf_{n\to\infty} \frac{1}{a_n^2} \ln \mathbf{P}\left(\frac{W_n}{a_n} \in B\right) \ge -\frac{1}{2}(x_0 - \varepsilon_2)^2.$$

Letting $\varepsilon_2 \to 0$, we get

$$\liminf_{n\to\infty} \frac{1}{a_n^2} \ln \mathbf{P}\left(\frac{W_n}{a_n} \in B\right) \ge -\frac{x_0^2}{2} \ge -\inf_{x\in B^o} \frac{x^2}{2} - \varepsilon_1.$$

Because ε_1 can be arbitrarily small, we obtain (4.35). This completes the proof of Corollary 2.2.

Acknowledgements

Supplementary Material

Supplement to "Self-normalized Cramér type moderate deviations for martingales" (DOI: [10.3150/18-BEJ1071SUPP](); .pdf). The supplement gives the detailed proofs of Propositions 3.1 and 3.2.

References

[1] Bentkus, V., Bloznelis, M. and Götze, F. (1996). A Berry–Esséen bound for Student's statistic in the non-i.i.d. case. *J. Theoret. Probab.* **9** 765–796. MR1400598

[2] Bentkus, V. and Götze, F. (1996). The Berry–Esseen bound for Student's statistic. *Ann. Probab.* **24** 491–503. MR1387647

[3] Bercu, B., Delyon, B. and Rio, E. (2015). *Concentration Inequalities for Sums and Martingales*. *SpringerBriefs in Mathematics*. Cham: Springer. MR3363542

[4] Bercu, B. and Touati, A. (2008). Exponential inequalities for self-normalized martingales with applications. *Ann. Appl. Probab.* **18** 1848–1869. MR2462551

[5] Bolthausen, E. (1982). Exact convergence rates in some martingale central limit theorems. *Ann. Probab.* **10** 672–688. MR0659537

[6] Chen, S., Wang, Z., Xu, W. and Miao, Y. (2014). Exponential inequalities for self-normalized martingales. *J. Inequal. Appl.* **2014** 289. MR3317493

[7] Chung, K.-L. (1946). The approximate distribution of Student's statistic. *Ann. Math. Stat.* **17** 447–465. MR0018390

[8] Delyon, B. (2009). Exponential inequalities for sums of weakly dependent variables. *Electron. J. Probab.* **14** 752–779. MR2495559

[9] de la Peña, V.H. (1999). A general class of exponential inequalities for martingales and ratios. *Ann. Probab.* **27** 537–564. MR1681153

[10] de la Peña, V.H., Lai, T.L. and Shao, Q.-M. (2009). *Self-Normalized Processes*: *Limit Theory and Statistical Applications*. *Probability and Its Applications* (*New York*). Berlin: Springer. MR2488094

[11] de la Peña, V.H. and Pang, G. (2009). Exponential inequalities for self-normalized processes with applications. *Electron. Commun. Probab.* **14** 372–381. MR2545288

[12] Esscher, F. (1924). On a method of determining correlation from the ranks of the variates. *Scand. Actuar. J.* **1** 201–219.

[13] Fan, X., Grama, I., Liu, Q. and Shao, Q.-M. (2018). Supplement to "Self-normalized Cramér type moderate deviations for martingales." DOI:10.3150/18-BEJ1071SUPP.

[14] Giné, E., Götze, F. and Mason, D.M. (1997). When is the Student t-statistic asymptotically standard normal? *Ann. Probab.* **25** 1514–1531. MR1457629

[15] Grama, I. and Haeusler, E. (2000). Large deviations for martingales via Cramér's method. *Stochastic Process. Appl.* **85** 279–293. MR1731027

[16] Haeusler, E. (1988). On the rate of convergence in the central limit theorem for martingales with discrete and continuous time. *Ann. Probab.* **16** 275–299. MR0920271

[17] Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and Its Application*: *Probability and Mathematical Statistics*. New York: Academic Press [Harcourt Brace Jovanovich, Publishers]. MR0624435

[18] Hu, Z., Shao, Q.-M. and Wang, Q. (2009). Cramér type moderate deviations for the maximum of self-normalized sums. *Electron. J. Probab.* **14** 1181–1197. MR2511281

[19] Itô, K. and McKean, H.P. Jr. (1996). *Difussion Processes and Their Sample Paths*. Berlin: Springer.

[20] Jing, B., Liang, H. and Zhou, W. (2012). Self-normalized moderate deviations for independent random variables. *Sci. China Math.* **55** 2297–2315. MR2994121

[21] Jing, B.-Y., Shao, Q.-M. and Wang, Q. (2003). Self-normalized Cramér-type large deviations for independent random variables. *Ann. Probab.* **31** 2167–2215. MR2016616

[22] Linnik, Ju.V. (1961). On the probability of large deviations for the sums of independent variables. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. II* 289–306. Berkeley, CA: Univ. California Press. MR0137142

[23] Liu, W., Shao, Q.-M. and Wang, Q. (2013). Self-normalized Cramér type moderate deviations for the maximum of sums. *Bernoulli* **19** 1006–1027. MR3079304

[24] Novak, S.Y. (2011). *Extreme Value Methods with Applications to Finance. Monographs on Statistics and Applied Probability* **122**. Boca Raton, FL: CRC Press. MR2933280

[25] Petrov, V.V. (1975). *Sums of Independent Random Variables*. New York: Springer. Translated from the Russian by A. A. Brown, *Ergebnisse der Mathematik und ihrer Grenzgebiete*, Band 82. MR0388499

[26] Shao, Q.-M. (1997). Self-normalized large deviations. *Ann. Probab.* **25** 285–328. MR1428510

[27] Shao, Q.-M. (1999). A Cramér type large deviation result for Student's *t*-statistic. *J. Theoret. Probab.* **12** 385–398. MR1684750

[28] Shao, Q.-M. and Wang, Q. (2013). Self-normalized limit theorems: A survey. *Probab. Surv.* **10** 69–93. MR3161676

[29] Shao, Q.-M. and Zhou, W.-X. (2016). Cramér type moderate deviation theorems for self-normalized processes. *Bernoulli* **22** 2029–2079. MR3498022

[30] Slavova, V.V. (1985). On the Berry–Esseen bound for Student's statistic. In *Stability Problems for Stochastic Models (Uzhgorod*, 1984). *Lecture Notes in Math.* **1155** 355–390. Berlin: Springer. MR0825335

[31] Talagrand, M. (1995). The missing factor in Hoeffding's inequalities. *Ann. Inst. Henri Poincaré Probab. Stat.* **31** 689–702. MR1355613

[32] von Bahr, B. and Esseen, C.-G. (1965). Inequalities for the *r*th absolute moment of a sum of random variables, $1 \leq r \leq 2$. *Ann. Math. Stat.* **36** 299–303. MR0170407