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

Cramér moderate deviation expansion for martingales with one-sided Sakhanenko's condition and its applications

Xiequan Fan · Ion Grama · Quansheng Liu

Received: date / Accepted: date

Abstract We give a Cramér moderate deviation expansion for martingales with differences having finite conditional moments of order $2+\rho, \rho \in (0,1]$, and finite one-sided conditional exponential moments. The upper bound of the range of validity and the remainder of our expansion are both optimal. Consequently, it leads to a one-sided moderate deviation principle for martingales. Moreover, applications to quantile coupling inequality, β-mixing and ψ-mixing sequences are discussed.

Keywords Martingales · Cramér moderate deviations · quantile coupling inequality · β-mixing sequences · ψ-mixing sequences

Mathematics Subject Classification (2000) 60G42 · 60F10 · 60E15 · 60F05

1 Introduction

Let $(\eta_i)_{i\geq 1}$ be a sequence of independent and identically distributed (i.i.d.) centered real random variables (r.v.s) satisfying Cramér's condition $\mathbf{E}\exp\{c_0|\eta_1|\} < \infty$, for some constant $c_0 > 0$. Without loss of generality, assume that $\mathbf{E}\eta_1^2 = 1$. Cramér [6] established an asymptotic expansion of the probabilities of moderate deviations for the partial sums $\sum_{i=1}^n \eta_i$, based on the powerful technique of conjugate distributions (see also Esscher [10]). The result of Cramér implies that uniformly in $0 \leq x = o(n^{1/2})$,

$$\log \frac{\mathbf{P}(\sum_{i=1}^n \eta_i > x\sqrt{n})}{1-\Phi(x)} = O\left(\frac{1+x^3}{\sqrt{n}}\right) \quad \text{as} \quad n \to \infty, \tag{1.1}$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^x \exp\{-t^2/2\}dt$ is the standard normal distribution function. Cramér type moderate deviations for sums of independent r.v.s have been obtained by many authors. See, for instance, Feller [14], Petrov [19], Sakhanenko [24] and [12]. We refer to the monographs of Petrov [20], Saulis and Statulevičius [25] and the references therein.

X. Fan
Center for Applied Mathematics, Tianjin University, Tianjin 300072, China
E-mail: fanxiequan@hotmail.com

I. Grama · Q. Liu
Univ. Bretagne-Sud, UMR 6205, LMBA, 56000 Vannes, France
E-mail: ion.grama@univ-ubs.fr, quansheng.liu@univ-ubs.fr

In this paper we are concerned with Cramér moderate deviations for martingales. When the martingale differences are bounded, we refer to Bose [3,4], Račkauskas [21–23], Grama and Haeusler [16]. Let $(\eta_i, \mathcal{F}_i)_{i=0,...,n}$ be a sequence of square integrable martingale differences defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, where $\eta_0 = 0$ and $\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq ... \subseteq \mathcal{F}_n \subseteq \mathcal{F}$. Assume that there exist absolute constants $H > 0$ and $N \geq 0$ such that $\max_i |\eta_i| \leq H$ and $\left|\sum_{i=1}^n \mathbf{E}[\eta_i^2|\mathcal{F}_{i-1}] - n\right| \leq N^2$. Here and hereafter, the equalities and inequalities between random variables are understood in the $\mathbf{P}$-almost sure sense. From the results in Grama and Haeusler [16], it follows that

$$\log \frac{\mathbf{P}(\sum_{i=1}^n \eta_i > x\sqrt{n})}{1 - \Phi(x)} = O\bigg(\frac{x^3}{\sqrt{n}}\bigg), \tag{1.2}$$

for all $\sqrt{\log n} \leq x = o(n^{1/4}), n \to \infty$, and that

$$\frac{\mathbf{P}\left(\sum_{i=1}^n \eta_i > x\sqrt{n}\right)}{1 - \Phi\left(x\right)} = 1 + o(1) \tag{1.3}$$

uniformly for $0 \leq x = o\left(n^{1/6}\right), n \to \infty$. In [11] the expansions (1.2) and (1.3) have been extended to the case of martingale differences satisfying the conditional Bernstein condition:

$$\Big|\mathbf{E}[\eta_i^k|\mathcal{F}_{i-1}]\Big| \leq \frac{1}{2}k!H^{k-2}\mathbf{E}[\eta_i^2|\mathcal{F}_{i-1}] \quad \text{for} \;\; k \geq 3 \;\; \text{and} \;\; 1 \leq i \leq n, \tag{1.4}$$

where H is a positive absolute constant. We note that the conditional Bernstein condition implies that the martingale differences have finite two-sided conditional exponential moments.

In this paper we extend the expansions (1.2) and (1.3) to the case of martingales with differences having finite $(2+\rho)$th moments, $\rho \in (0, 1]$, and finite one-sided conditional exponential moments. Assume that there exist constants $L, M > 0$ and $N \geq 0$ such that

$$\mathbf{E}[|\eta_i|^{2+\rho} e^{L\,\eta_i^+}|\mathcal{F}_{i-1}] \leq M^\rho\, \mathbf{E}[\eta_i^2|\mathcal{F}_{i-1}] \quad \text{for all } 1 \leq i \leq n \tag{1.5}$$

and

$$\Big|\sum_{i=1}^n \mathbf{E}[\eta_i^2|\mathcal{F}_{i-1}] - n\Big| \leq N^2. \tag{1.6}$$

It is easy to see that the conditional Bernstein condition implies (1.5) with $\rho = 1$, while condition (1.5) generally does not imply the conditional Bernstein condition; see (2.3) for an example. In Theorem 1 of the paper, we prove that if $\rho \in (0, 1)$, then for all $0 \leq x = o\left(n^{1/2}\right)$,

$$\log \frac{\mathbf{P}\left(\sum_{i=1}^n \eta_i > x\sqrt{n}\right)}{1 - \Phi\left(x\right)} = O\bigg(\frac{1 + x^{2+\rho}}{n^{\rho/2}}\bigg) \quad \text{as} \;\; n \to \infty. \tag{1.7}$$

The expansion (1.7) can be regard as an extension of (1.2). We would like to point out that the range of validity of (1.2) has been enlarged to the classical Cramér's one, and therefore is optimal. Moreover, it is worth mentioning that (1.7) is new even for independent r.v.s. The last expansion implies that (1.3) holds uniformly in the range $0 \leq x = o\left(n^{\rho/(4+2\rho)}\right)$. We also show that when $\rho = 1$, equality (1.7) holds for all $\sqrt{\log n} \leq x = o(n^{1/2})$, see Remark 1 for details.

The paper is organized as follows. Our main results for martingales are stated and discussed in Section 2. Applications to quantile coupling inequality, β-mixing and ψ-mixing sequences are discussed in Section 3. Proofs of the theorems and their preliminary lemmas are deferred to Sections 4-10. The proofs of Theorem 1 and Lemma 4 are refinements of Fan et al. [11]. The applications of our results are new, and therefore are of independent interest.

Throughout the paper, c and c_α, probably supplied with some indices, denote respectively a generic positive constant and a generic positive constant depending only on α. Denote by $\xi^+ = \max\{\xi, 0\}$ the positive part of ξ.

2 Main results

Let $n \geq 1$, and let $(\xi_i, \mathcal{F}_i)_{i=0,...,n}$ be a sequence of martingale differences, defined on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$, where $\xi_0 = 0$, $\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq ... \subseteq \mathcal{F}_n \subseteq \mathcal{F}$ are increasing σ-fields and $(\xi_i)_{i=1,...,n}$ are allowed to depend on n. Set

$$X_0 = 0, \qquad X_k = \sum_{i=1}^{k} \xi_i, \qquad k = 1, ..., n. \tag{2.1}$$

Let $\langle X \rangle$ be the conditional variance of the martingale $X = (X_k, \mathcal{F}_k)_{k=0,...,n}$:

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

In the sequel we shall use the following conditions:

(A1) There exist a constant $\rho \in (0, 1]$ and positive numbers $\varepsilon_n \in (0, \frac{1}{2}]$ such that

$$\mathbf{E}[|\xi_i|^{2+\rho} e^{\varepsilon_n^{-1} \xi_i^+} | \mathcal{F}_{i-1}] \leq \varepsilon_n^\rho \, \mathbf{E}[\xi_i^2 | \mathcal{F}_{i-1}] \quad \text{for all } 1 \leq i \leq n.$$

(A2) There exist non-negative numbers $\delta_n \in [0, \frac{1}{2}]$ such that $|\langle X \rangle_n - 1| \leq \delta_n^2$ a.s.

Condition (A1) can be seen as a one-sided version of Sakhanenko's condition [24]. In the case of normalized sums of i.i.d. random variables, conditions (A1) and (A2) are satisfied with $\varepsilon_n = O(\frac{1}{\sqrt{n}})$ and $\delta_n = 0$. In the case of martingales, ε_n and δ_n usually are satisfying $\varepsilon_n, \delta_n \to 0$ as $n \to \infty$.

Notice that condition (A1) implies that $\mathbf{E}[e^{\varepsilon_n^{-1} \xi_i^+} | \mathcal{F}_{i-1}]$ must be finite, which means that the positive part of the conditional distribution of ξ_i / ε_n has an exponential moment, and therefore has conditional moments of any order. However, such an assumption is not required for the negative part of the conditional distribution. For the negative part of ξ_i, we assume a finite conditional moment of order $2+\rho$. Thus, condition (A1) does not imply the conditional Cramér condition, because $\mathbf{E}[e^{\varepsilon_n^{-1} |\xi_i|} | \mathcal{F}_{i-1}]$ may not exist.

Let us remark that if ξ_i is bounded, say $|\xi_i| \leq \gamma_n$, then condition (A1) is satisfied with $\varepsilon_n = e^{1/\rho} \gamma_n$. On the other hand, if ξ_i satisfies

$$\xi_i \leq \gamma_n \quad \text{and} \quad \mathbf{E}[|\xi_i|^{2+\rho} | \mathcal{F}_{i-1}] \leq \tau_n^\rho \, \mathbf{E}[\xi_i^2 | \mathcal{F}_{i-1}] \quad \text{for all } 1 \leq i \leq n, \tag{2.3}$$

then condition (A1) is also satisfied with $\varepsilon_n = \max\{\gamma_n, e^{1/\rho} \tau_n\}$. Here we assume that $0 < \gamma_n, \tau_n \leq \frac{1}{2} e^{-1/\rho}$.

The following theorem gives a Cramér moderate deviation expansion for martingales.

Theorem 1 *Assume conditions (A1) and (A2).*

[i] *If* $\rho \in (0, 1)$*, then there is a constant* $\alpha > 0$*, such that for all* $0 \leq x \leq \alpha \varepsilon_n^{-1}$,

$$\left| \ln \frac{\mathbf{P}(X_n > x)}{1 - \Phi(x)} \right| \leq c_{\alpha,\rho} \left(x^{2+\rho} \varepsilon_n^\rho + x^2 \delta_n^2 + (1 + x) \left(\varepsilon_n^\rho + \delta_n \right) \right). \tag{2.4}$$

[ii] *If* $\rho = 1$, *then there is a constant* $\alpha > 0$, *such that for all* $0 \le x \le \alpha \varepsilon_n^{-1}$,

$$\left| \ln \frac{\mathbf{P}(X_n > x)}{1 - \Phi(x)} \right| \le c_\alpha \bigg(x^3 \varepsilon_n + x^2 \delta_n^2 + (1+x) \left(\varepsilon_n |\ln \varepsilon_n| + \delta_n \right) \bigg). \tag{2.5}$$

The term $\varepsilon_n |\ln \varepsilon_n|$ in (2.5) cannot be replaced by ε_n under the stated conditions. Indeed, Bolthausen [2] showed that there exists a sequence of martingale differences satisfying $|\xi_i| \le 2/\sqrt{n}$ and $\langle X \rangle_n = 1$ a.s., such that for all n large enough,

$$\sup_{x \in \mathbf{R}} \Big| \mathbf{P}(X_n \le x) - \Phi(x) \Big| \frac{\sqrt{n}}{\log n} \ \ge \ c, \tag{2.6}$$

where c is a positive constant and does not depend on n. See also [13] for general ε_n. If $\varepsilon_n |\ln \varepsilon_n|$ in (2.5) could be improved to ε_n, then we can deduce the following Berry-Esseen bound

$$\sup_{x \in \mathbf{R}} |\mathbf{P}(X_n \le x) - \Phi(x)| \le c\,(\varepsilon_n + \delta_n), \tag{2.7}$$

which would violate Bolthausen's result (2.6). Thus $\varepsilon_n |\ln \varepsilon_n|$ in (2.5) cannot be improved to ε_n even for bounded martingale differences.

If the martingale differences are bounded $|\xi_i| \le \varepsilon_n$ and satisfy condition (A2), Grama and Haeusler [16] proved the asymptotic expansion (2.5) for all $x \in [0, \alpha \min\{\varepsilon_n^{-1/2}, \delta_n^{-1}\}]$. Now Theorem 1 holds for a larger range $x \in [0, \alpha \varepsilon_n^{-1}]$ and a much more general class of martingales.

The following corollary states that under conditions (A1) and (A2), the tail probabilities $\mathbf{P}(X_n > x)$ can be uniformly approximated by the tail probabilities of the standard normal random variable, when x is in a certain reduced range.

Corollary 1 *Assume conditions (A1) and (A2).*

[i] *If* $\rho \in (0, 1)$, *then for all* $0 \le x = o(\min\{\varepsilon_n^{-\rho/(2+\rho)}, \delta_n^{-1}\})$,

$$\left| \frac{\mathbf{P}(X_n > x)}{1 - \Phi(x)} - 1 \right| \le c_\rho \bigg(x^{2+\rho} \varepsilon_n^\rho + (1+x) \left(\varepsilon_n^\rho + \delta_n \right) \bigg). \tag{2.8}$$

[ii] *If* $\rho = 1$, *then for all* $0 \le x = o(\min\{\varepsilon_n^{-1/3}, \delta_n^{-1}\})$,

$$\left| \frac{\mathbf{P}(X_n > x)}{1 - \Phi(x)} - 1 \right| \le c \bigg(x^3 \varepsilon_n + (1+x) \left(\varepsilon_n |\ln \varepsilon_n| + \delta_n \right) \bigg). \tag{2.9}$$

In particular, this implies that

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

holds uniformly for $0 \le x = o(\min\{\varepsilon_n^{-\rho/(2+\rho)}, \delta_n^{-1}\})$ *as* $\max\{\varepsilon_n, \delta_n\} \to 0$.

The inequalities (2.4) and (2.5) together implies that there is a constant $\alpha > 0$ such that for $\rho \in (0, 1]$ and all $0 \le x \le \alpha \varepsilon_n^{-1}$,

$$\left| \log \frac{\mathbf{P}(X_n > x)}{1 - \Phi(x)} \right| \le c_\alpha \bigg(x^{2+\rho} \varepsilon_n^\rho + x^2 \delta_n^2 + (1+x) \left(\varepsilon_n^\rho |\ln \varepsilon_n| + \delta_n \right) \bigg). \tag{2.10}$$

By (2.10), we obtain the following moderate deviation principle (MDP) result.

Corollary 2 *Assume that conditions (A1) and (A2) are satisfied with* $\max\{\delta_n, \varepsilon_n\} \to 0$ *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* $B \subset [0, \infty)$,

$$
\begin{aligned}
-\inf_{x \in B^o} \frac{x^2}{2} &\leq \liminf_{n \to \infty} \frac{1}{a_n^2} \log \mathbf{P}\left(\frac{1}{a_n} X_n \in B\right) \\
&\leq \limsup_{n \to \infty} \frac{1}{a_n^2} \log \mathbf{P}\left(\frac{1}{a_n} X_n \in B\right) \leq -\inf_{x \in \overline{B}} \frac{x^2}{2}\,, \qquad (2.11)
\end{aligned}
$$

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

Since (2.11) may not hold for all Borel set $B \subset (-\infty, 0]$, inequality (2.11) does not imply the usual MDP, but it can be seen as a one-sided MDP.

Similar MDP results for martingales can be found in Dembo [8], Gao [15] and Djellout [7]. For the most recent work on MDP for martingales with the conditional Cramér condition and the assumption that $\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] = 1/n$ a.s. for all i., we refer to Eichelsbacher and Löwe [9] where the authors established a MDP result via Lindeberg's method.

Remark 1 The sequence of martingale differences $(\xi_i, \mathcal{F}_i)_{i=0,\ldots,n}$ discussed so far is standardized. For a general sequence of martingale differences $(\eta_i, \mathcal{F}_i)_{i \geq 1}$, one can restate the conditions (A1) and (A2) as below.

(A1′) There exist three positive constants $\rho \in (0, 1], K$ and L such that

$$
\mathbf{E}[|\eta_i|^{2+\rho} e^{K\eta_i^+}|\mathcal{F}_{i-1}] \leq L^\rho\, \mathbf{E}[\eta_i^2|\mathcal{F}_{i-1}] \quad \text{for all } 1 \leq i \leq n;
$$

(A2′) There exists a constant $N \geq 0$ such that

$$
\Big|\sum_{i=1}^n \mathbf{E}[\eta_i^2|\mathcal{F}_{i-1}] - n\Big| \leq N^2 \qquad \text{a.s.}
$$

Under conditions (A1′) and (A2′), the inequalities (2.4)-(2.11) remain valid for

$$
W_n = \sum_{i=1}^n \frac{\eta_i}{\sqrt{n}} \qquad (2.12)
$$

instead of X_n, with $\varepsilon_n = n^{-1/2} \max\{K, L\}$ and $\delta_n = n^{-1/2} L$.

3 Applications

3.1 Quantile coupling inequality

Thanks to the work of Mason and Zhou [18], it is known that the Cramér moderate deviation expansion can be applied to establishing quantile coupling inequalities. When the martingale differences are bounded, a quantile coupling inequality has been established by Mason and Zhou, see Corollary 2 of [18]. Here, we give a generalization of the inequality of Mason and Zhou [18].

Let $(W_n)_{n \geq 1}$ be a sequence of random variables and for each integer $n \geq 1$, and let

$$
F_n(x) = \mathbf{P}(W_n \leq x), \quad x \in \mathbf{R},
$$

denote the cumulative distribution function of W_n. Its *quantile function* is defined by

$$H_n(s) = \inf\{x : F_n(x) \geq s\}, \quad s \in (0, 1).$$

Let Z denote a standard normal random variable. Since $\Phi(Z) =_d U$ the uniformly distribution random variable, then it is obvious that for each integer $n \geq 1$,

$$H_n(\Phi(Z)) =_d W_n,$$

where $=_d$ stands for equivalent in distribution. For this reason, we define

$$W_n = H_n(\Phi(Z)). \tag{3.1}$$

By Theorem 1, we prove the following quantile inequality.

Theorem 2 *Let $(\eta_i, \mathcal{F}_i)_{i\geq 1}$ be a sequence of martingale differences satisfying the following conditional Sakhanenko condition*

$$\mathbf{E}[|\eta_i|^3 e^{K|\eta_i|}|\mathcal{F}_{i-1}] \leq L\, \mathbf{E}[\eta_i^2|\mathcal{F}_{i-1}], \qquad i \geq 1,$$

and

$$\Big|\sum_{i=1}^{n} \mathbf{E}[\eta_i^2|\mathcal{F}_{i-1}] - n\Big| \leq M \quad a.s.,$$

where $\rho \in (0, 1]$, K, L and M are positive constants. Assume that $W_n =_d \sum_{i=1}^n \eta_i/\sqrt{n}$ and W_n is defined as in (3.1). There there exist constants $\alpha > 0$ and $D > 0$ and an integer n_0 such that whenever $n \geq n_0$ and

$$|W_n| \leq \alpha\sqrt{n}, \tag{3.2}$$

we have

$$\sqrt{n}|W_n - Z|/\ln n \leq 2D(W_n^2 + 1) \quad a.s. \tag{3.3}$$

Furthermore, there exist two positive constants C and λ such that whenever $n \geq n_0$, we have for all $x \geq 0$,

$$\mathbf{P}\Big(\sqrt{n}|W_n - Z|/\ln n > x\Big) \leq C \exp\Big\{ -\lambda\, x\Big\}. \tag{3.4}$$

When the martingale differences are *bounded*, Mason and Zhou [18] proved that (3.3) holds whenever $|W_n| \leq \alpha\sqrt[4]{n}$. Notice that the bounded martingale differences satisfy the conditional Sakhanenko condition. Moreover, the range $|W_n| \leq \alpha\sqrt[4]{n}$ has been extended to a much larger one $|W_n| \leq \alpha\sqrt{n}$ in our theorem.

3.2 β-mixing sequences

Let $(\eta_i)_{i\geq 1}$ be a random process that may be non-stationary. Write $S_{k,m} = \sum_{i=k+1}^{k+m} \eta_i$. Assume that there exists a constant $\rho \in (0,1)$ such that

$$\mathbf{E}\eta_i = 0 \quad \text{for all } i, \tag{3.5}$$

$$\mathbf{E}|S_{k,m}|^{2+\rho} \leq m^{1+\rho/2} c_1^{2+\rho}, \tag{3.6}$$

and

$$\mathbf{E}S_{k,m}^2 \geq c_2^2 m \qquad \text{for all } k \geq 0, m \geq 1. \tag{3.7}$$

Let $\mathcal{F}_j$ and $\mathcal{F}_{j+k}^{\infty}$ be σ-fields generated respectively by $(\eta_i)_{i\leq j}$ and $(\eta_i)_{i\geq j+k}$. We say that $(\eta_i)_{i\geq 1}$ is *β-mixing* if

$$\beta(n) =: \sup_j \mathbf{E} \sup\{\left|\mathbf{P}(B|\mathcal{F}_j) - \mathbf{P}(B)\right| : \ B \in \mathcal{F}_{j+n}^{\infty}\} \to \infty, \qquad n \to \infty.$$

Assume that there exist positive numbers a_1, a_2 and τ such that

$$\beta(n) \leq a_1 \exp\{-a_2 n^{\tau}\}. \tag{3.8}$$

By Theorem 4.1 of Shao and Yu [26], it is known that (3.6) is implied by the condition that $\mathbf{E}|\eta_i|^{2+\rho'} \leq c_1^{2+\rho'}$ for a constant $\rho' > \rho$.

Set $\alpha \in (0, \frac{1}{2})$. Let $m = \lfloor n^{\alpha} \rfloor$ and $k = \lfloor n/(2m) \rfloor$ be respectively the integers part of n^{α} and $n/(2m)$. Let

$$Y_j = \sum_{i=1}^{m} \eta_{2m(j-1)+i} \qquad \text{and} \qquad S_n = \sum_{j=1}^{k} Y_j.$$

Note that S_n is an interlacing sum of $(\eta_i)_{i\geq 1}$, and that $\mathrm{Var}(S_n) = \mathbf{E}S_n^2$.

Theorem 3 *Assume conditions (3.5)-(3.8). Suppose that $\eta_i \leq c_3$ for all i. Then for all $0 \leq x = o(\min\{n^{\frac{1}{2}-\alpha},\, n^{\alpha\,\tau/2}\})$,*

$$\left| \ln \frac{\mathbf{P}(S_n/\sqrt{\mathbf{E}S_n^2} > x)}{1 - \Phi\,(x)} \right| \leq \ c_\rho \, \frac{(1+x)^{2+\rho}}{n^{\rho(\frac{1}{2}-\alpha)}}. \tag{3.9}$$

In particular, we have

$$\frac{\mathbf{P}(S_n/\sqrt{\mathbf{E}S_n^2} > x)}{1 - \Phi\,(x)} = 1 + o(1) \tag{3.10}$$

uniformly for $0 \leq x = o(\min\{n^{\rho(1-2\alpha)/(4+2\rho)},\, n^{\alpha\,\tau/2}\})$.

For a counterpart of Theorem 3 for interlacing self-normalized sums $W_n = S_n/\sqrt{\sum_{j=1}^{k} Y_j^2}$, we refer to Chen et al. [5].

The following MDP result is a consequence of the last theorem with $\alpha = 1/(2+\tau)$.

Corollary 3 *Assume the conditions of Theorem 3. Let a_n be any sequence of real numbers satisfying $a_n \to \infty$ and $a_n n^{-\tau/(2\tau+4)} \to 0$ as $n \to \infty$. Then for each Borel set $B \subset [0,\infty)$,*

$$\begin{aligned} - \inf_{x\in B^o} \frac{x^2}{2} &\leq \liminf_{n\to\infty} \frac{1}{a_n^2} \log \mathbf{P}\left(\frac{1}{a_n} \frac{S_n}{\sqrt{\mathbf{E}S_n^2}} \in B \right) \\ &\leq \limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbf{P}\left(\frac{1}{a_n} \frac{S_n}{\sqrt{\mathbf{E}S_n^2}} \in B \right) \leq - \inf_{x\in \overline{B}} \frac{x^2}{2}\,, \end{aligned}$$

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

3.3 ψ-mixing sequences

Recall the notations in Section 3.2. We say that $(\eta_i)_{i\geq 1}$ is ψ-*mixing* if

$$\psi(n) =: \sup_j \sup_B \{\big|\mathbf{P}(B|\mathcal{F}_j) - \mathbf{P}(B)\big|/\mathbf{P}(B):\ B \in \mathcal{F}_{j+n}^{\infty}\} \to 0, \qquad n \to \infty. \tag{3.11}$$

Set $\alpha \in (0, \frac{1}{2})$. Let $m = \lfloor n^{\alpha} \rfloor$ and $k = \lfloor n/(2m) \rfloor$ be respectively the integers part of n^{α} and $n/(2m)$, and let

$$Y_j = \sum_{i=1}^{m} \eta_{2m(j-1)+i} \qquad \text{and} \qquad S_n = \sum_{j=1}^{k} Y_j$$

as in Section 3.2.

Denote

$$\tau_n^2 = \psi(m) + n\psi^2(m) + k\psi^{1/2}(m). \tag{3.12}$$

We have the following Cramér moderate deviations for ψ-mixing sequences.

Theorem 4 *Assume conditions (3.5)-(3.7) with $\rho \in (0,1]$. Suppose that $\eta_i \leq c_3$ for all i, and that $\tau_n \to 0$ as $n \to \infty$.*

[i] *If $\rho \in (0,1)$, then for all $0 \leq x = o(n^{\frac{1}{2}-\alpha})$,*

$$\left| \ln \frac{\mathbf{P}(S_n/\sqrt{\mathbf{E}S_n^2} > x)}{1-\Phi(x)} \right| \leq c_\rho \left(\frac{x^{2+\rho}}{n^{\rho(\frac{1}{2}-\alpha)}} + x^2\tau_n^2 + (1+x)\Big(\frac{1}{n^{\rho(\frac{1}{2}-\alpha)}} + \tau_n\Big) \right). \tag{3.13}$$

[ii] *If $\rho = 1$, then for all $0 \leq x = o(n^{\frac{1}{2}-\alpha})$,*

$$\left| \ln \frac{\mathbf{P}(S_n/\sqrt{\mathbf{E}S_n^2} > x)}{1-\Phi(x)} \right| \leq c \left(\frac{x^{3}}{n^{\frac{1}{2}-\alpha}} + x^2\tau_n^2 + (1+x)\Big(\frac{|\ln n|}{n^{\frac{1}{2}-\alpha}} + \tau_n\Big) \right). \tag{3.14}$$

In particular, if

$$\psi(n) = O\big(n^{-(2+\rho)(1-\alpha)/\alpha}\big), \tag{3.15}$$

then

$$\tau_n = O(n^{-\rho(\frac{1}{2}-\alpha)}) \qquad \textit{and} \qquad \frac{\mathbf{P}(S_n/\sqrt{\mathbf{E}S_n^2} > x)}{1-\Phi(x)} = 1 + o(1) \tag{3.16}$$

uniformly for $0 \leq x = o(n^{\rho(1-2\alpha)/(4+2\rho)})$.

In the independent case, we have $\psi(n) = 0$ and $\tau_n = 0$. Let $\alpha \to 0$. Then (3.13) and (3.14) recover the optimal range of validity, that is $0 \leq x = o(n^{1/2})$.

The following MDP result is a consequence of the last theorem.

Corollary 4 *Assume the conditions of Theorem 4. Let a_n be any sequence of real numbers satisfying $a_n \to \infty$ and $a_n/n^{\frac{1}{2}-\alpha} \to 0$ as $n \to \infty$. Then for each Borel set $B \subset [0,\infty)$,*

$$\begin{aligned} -\inf_{x\in B^o} \frac{x^2}{2} &\leq \liminf_{n\to\infty} \frac{1}{a_n^2} \log \mathbf{P}\left(\frac{1}{a_n} \frac{S_n}{\sqrt{\mathbf{E}S_n^2}} \in B \right) \\ &\leq \limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbf{P}\left(\frac{1}{a_n} \frac{S_n}{\sqrt{\mathbf{E}S_n^2}} \in B \right) \leq -\inf_{x\in \overline{B}} \frac{x^2}{2}\, , \end{aligned}$$

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

4 Preliminary lemmas

Assume condition (A1). For any real $\lambda \in [0,\ \varepsilon_n^{-1}]$, define the exponential multiplicative martingale $Z(\lambda) = (Z_k(\lambda), \mathcal{F}_k)_{k=0,...,n}$, where

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

Then 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 the conjugate probability measure $\mathbf{P}_\lambda$ on $(\Omega, \mathcal{F})$ defined by

$$d\mathbf{P}_\lambda = Z_n(\lambda) d\mathbf{P}. \tag{4.1}$$

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

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

We thus have the following decomposition:

$$X_k = Y_k(\lambda) + B_k(\lambda), \qquad k = 1, ..., n, \tag{4.2}$$

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

$$Y_k(\lambda) = \sum_{i=1}^{k} \eta_i(\lambda), \qquad k = 1, ..., n, \tag{4.3}$$

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 theorem, we need a two-sided bound for the drift process $B_n(\lambda)$. To this end, we prove the following lemma.

Lemma 1 *If there exists an $s > 2$, such that*

$$\mathbf{E}[|\xi_i|^s e^{\varepsilon_n^{-1} \xi_i^+}|\mathcal{F}_{i-1}] \le \varepsilon_n^{s-2}\, \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}], \tag{4.4}$$

then

$$\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] \le \varepsilon_n^2. \tag{4.5}$$

In particular, condition (A1) implies (4.5).

Proof. By Jensen's inequality, it is easy to see that

$$(\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}])^{s/2} \le \mathbf{E}[|\xi_i|^s|\mathcal{F}_{i-1}] \le \mathbf{E}[|\xi_i|^s e^{\varepsilon_n^{-1} \xi_i^+}|\mathcal{F}_{i-1}] \le \varepsilon_n^{s-2}\, \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}].$$

Thus

$$(\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}])^{s/2-1} \le \varepsilon_n^{s-2},$$

which implies (4.5). □

Using the last lemma, we establish a two-sided bound for the drift process $B_n(\lambda)$.

Lemma 2 *Assume conditions (A1) and (A2). Then for all* $0 \leq \lambda \leq \varepsilon_n^{-1}$,

$$|B_n(\lambda) - \lambda| \leq \lambda\delta_n^2 + c\,\lambda^{1+\rho}\varepsilon_n^{\rho}. \tag{4.6}$$

Proof. Jensen's inequality and $\mathbf{E}[\xi_i|\mathcal{F}_{i-1}] = 0$ imply that $\mathbf{E}[e^{\lambda\xi_i}|\mathcal{F}_{i-1}] \geq 1, \lambda \geq 0$. Notice that

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

Using Taylor's expansion for e^x, we get

$$\begin{aligned} B_n(\lambda) &\leq \sum_{i=1}^n \mathbf{E}[\xi_i e^{\lambda\xi_i}|\mathcal{F}_{i-1}] \\ &= \lambda\langle X\rangle_n + \sum_{i=1}^n \mathbf{E}\big[\xi_i(e^{\lambda\xi_i} - 1 - \lambda\xi_i)\,\big|\mathcal{F}_{i-1}\big]. \end{aligned}$$

Recall $\rho \in (0,1]$. When $x \leq -1$, by Taylor's expansion, it is easy to see that $\big|x(e^x - 1 - x)\big| \leq \big|x(e^x - 1)\big| + x^2 \leq 2|x|^{2+\rho}$. When $x \in (-1,1)$, again by Taylor's expansion, we get $|x(e^x - 1 - x)| \leq \frac{1}{2}|x|^3 e^{x^+} \leq |x|^{2+\rho}e^{x^+}$. When $x \geq 1$, we have $|x(e^x - 1 - x)| \leq xe^x \leq x^{2+\rho}e^x$. Thus, it holds

$$|x(e^x - 1 - x)| \leq 2|x|^{2+\rho}e^{x^+}, \qquad x \in \mathbf{R}. \tag{4.7}$$

By inequality (4.7), we obtain for all $0 \leq \lambda \leq \varepsilon_n^{-1}$,

$$\begin{aligned} B_n(\lambda) &\leq \lambda\langle X\rangle_n + 2\lambda^{1+\rho}\sum_{i=1}^n \mathbf{E}[|\xi_i|^{2+\rho}e^{\lambda\xi_i^+}|\mathcal{F}_{i-1}] \\ &\leq \lambda\langle X\rangle_n + 2\lambda^{1+\rho}\sum_{i=1}^n \mathbf{E}[|\xi_i|^{2+\rho}e^{\varepsilon_n^{-1}\xi_i^+}|\mathcal{F}_{i-1}]. \end{aligned} \tag{4.8}$$

Condition (A2) implies that $\langle X\rangle_n \leq 2$. Combining (4.8), conditions (A1) and (A2) together, we get the upper bound of $B_n(\lambda)$:

$$B_n(\lambda) \leq \lambda\langle X\rangle_n + 2\lambda^{1+\rho}\varepsilon_n^{\rho}\langle X\rangle_n \leq \lambda + \lambda\delta_n^2 + 4\lambda^{1+\rho}\varepsilon_n^{\rho}.$$

When $x \leq -1$, by Taylor's expansion, it is easy to see that $\big|e^x - 1 - x - \frac{1}{2}x^2\big| \leq \big|e^x - 1 - x\big| + \frac{1}{2}x^2 \leq |x|^{2+\rho}$. When $x \in (-1,1)$, again by Taylor's expansion, we get $\big|e^x - 1 - x - \frac{1}{2}x^2\big| \leq \frac{1}{6}|x|^3 e^{x^+} \leq |x|^{2+\rho}e^{x^+}$. When $x \geq 1$, we have $\big|e^x - 1 - x - \frac{1}{2}x^2\big| \leq \big|e^x - 1 - x\big| + \frac{1}{2}x^2 \leq x^{2+\rho}e^x$. Thus, it holds

$$\Big|e^x - 1 - x - \frac{1}{2}x^2\Big| \leq |x|^{2+\rho}e^{x^+}, \quad \rho \in (0,1] \text{ and } x \in \mathbf{R}. \tag{4.9}$$

Using inequality (4.9), condition (A1) and Lemma 1, we have for all $0 \leq \lambda \leq \varepsilon_n^{-1}$,

$$\begin{aligned} \mathbf{E}[e^{\lambda\xi_i}|\mathcal{F}_{i-1}] &= 1 + \frac{1}{2}\lambda^2\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] + \mathbf{E}\big[e^{\lambda\xi_i} - 1 - \lambda\xi_i - \frac{1}{2}\lambda^2\xi_i^2\big|\mathcal{F}_{i-1}\big] \\ &\leq 1 + \frac{1}{2}\lambda^2\,\mathbf{E}\big[\xi_i^2|\mathcal{F}_{i-1}\big] + \lambda^{2+\rho}\mathbf{E}\big[|\xi_i|^{2+\rho}e^{\varepsilon_n^{-1}\xi_i^+}\big|\mathcal{F}_{i-1}\big] \\ &\leq 1 + \big(\frac{1}{2}\lambda^2 + \lambda^{2+\rho}\varepsilon_n^{\rho}\big)\mathbf{E}\big[\xi_i^2|\mathcal{F}_{i-1}\big] \qquad (4.10) \\ &\leq 1 + 2(\lambda\varepsilon_n)^2. \qquad (4.11) \end{aligned}$$

By inequality (4.7) and the fact $\langle X\rangle_n \leq 2$, we deduce that for all $0 \leq \lambda \leq \varepsilon_n^{-1}$,

$$\begin{aligned}
\sum_{i=1}^n \mathbf{E}[\xi_i e^{\lambda \xi_i}|\mathcal{F}_{i-1}] &= \lambda \langle X\rangle_n + \sum_{i=1}^n \mathbf{E}\big[\xi_i (e^{\lambda\xi_i} - 1 - \lambda \xi_i)\,\big|\mathcal{F}_{i-1}\big] \\
&\geq \lambda\langle X\rangle_n - 2\lambda^{1+\rho} \sum_{i=1}^n \mathbf{E}[|\xi_i|^{2+\rho} e^{\varepsilon_n^{-1}\xi_i^+}|\mathcal{F}_{i-1}] \\
&\geq \lambda\langle X\rangle_n - 2\lambda^{1+\rho}\varepsilon_n^\rho \langle X\rangle_n \\
&\geq \lambda - \lambda\delta_n^2 - 4\,\lambda^{1+\rho}\varepsilon_n^\rho.
\end{aligned}$$

The last inequality together with (4.11) imply the lower bound of $B_n(\lambda)$: for all $0 \leq \lambda \leq \varepsilon_n^{-1}$,

$$\begin{aligned}
B_n(\lambda) &\geq \Big(\lambda - \lambda\delta_n^2 - 4\,\lambda^{1+\rho}\varepsilon_n^\rho\Big)\Big(1 + 2\,(\lambda\varepsilon_n)^2\Big)^{-1} \\
&\geq \lambda - \lambda\delta_n^2 - 6\lambda^{1+\rho}\varepsilon_n^\rho,
\end{aligned}$$

where the last line follows from the following inequality

$$\begin{aligned}
\lambda - \lambda\delta_n^2 - 4\,\lambda^{1+\rho}\varepsilon_n^\rho &\geq \lambda - \lambda\delta_n^2 - (6 - 2(\lambda\varepsilon_n)^{2-\rho})\lambda^{1+\rho}\varepsilon_n^\rho \\
&\geq \Big(\lambda - \lambda\delta_n^2 - 6\lambda^{1+\rho}\varepsilon_n^\rho\Big)\Big(1 + 2(\lambda\varepsilon_n)^2\Big).
\end{aligned}$$

The proof of Lemma 2 is finished. □

Next, we consider the following predictable cumulant process $\Psi(\lambda) = (\Psi_k(\lambda), \mathcal{F}_k)_{k=0,\ldots,n}$:

$$\Psi_k(\lambda) = \sum_{i=1}^k \log \mathbf{E}\big[e^{\lambda\xi_i}|\mathcal{F}_{i-1}\big]. \tag{4.12}$$

The following lemma gives a two-sided bound for the process $\Psi(\lambda)$.

Lemma 3 *Assume conditions (A1) and (A2). Then for all $0 \leq \lambda \leq \varepsilon_n^{-1}$,*

$$\left|\Psi_n(\lambda) - \frac{\lambda^2}{2}\right| \leq c\,\lambda^{2+\rho}\varepsilon_n^\rho + \frac{\lambda^2\delta_n^2}{2}. \tag{4.13}$$

Proof. Using a two-term Taylor's expansion of $\log(1+x), x \geq 0$, we have

$$\begin{aligned}
\Psi_n(\lambda) - \frac{\lambda^2}{2}\langle X\rangle_n &= \sum_{i=1}^n \Big(\mathbf{E}[e^{\lambda\xi_i}|\mathcal{F}_{i-1}] - 1 - \lambda\mathbf{E}[\xi_i|\mathcal{F}_{i-1}] - \frac{\lambda^2}{2}\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}]\Big) \\
&\quad - \sum_{i=1}^n \frac{1}{2\left(1 + \theta_i\left(\mathbf{E}[e^{\lambda\xi_i}|\mathcal{F}_{i-1}] - 1\right)\right)^2}\left(\mathbf{E}[e^{\lambda\xi_i}|\mathcal{F}_{i-1}] - 1\right)^2,
\end{aligned}$$

where $\theta_i \in (0,1)$. Since $\mathbf{E}[\xi_i|\mathcal{F}_{i-1}] = 0$ and $\mathbf{E}[e^{\lambda\xi_i}|\mathcal{F}_{i-1}] \geq 1$ for all $0 \leq \lambda \leq \varepsilon_n^{-1}$, we deduce that for all $0 \leq \lambda \leq \varepsilon_n^{-1}$,

$$\begin{aligned}
\left|\Psi_n(\lambda) - \frac{\lambda^2}{2}\langle X\rangle_n\right| &\leq \sum_{i=1}^n \left|\mathbf{E}[e^{\lambda\xi_i}|\mathcal{F}_{i-1}] - 1 - \lambda\mathbf{E}[\xi_i|\mathcal{F}_{i-1}] - \frac{\lambda^2}{2}\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}]\right| \\
&\quad + \frac{1}{2}\sum_{i=1}^n \left(\mathbf{E}[e^{\lambda\xi_i}|\mathcal{F}_{i-1}] - 1\right)^2.
\end{aligned}$$

Using condition (A1) and the inequalities (4.9)-(4.11), we get for all $0 \leq \lambda \leq \varepsilon_n^{-1}$,

$$\begin{aligned}\Big|\Psi_n(\lambda) - \frac{\lambda^2}{2}\langle X\rangle_n\Big| &\leq \sum_{i=1}^n \mathbf{E}[e^{\lambda\xi_i^+}|\lambda\xi_i|^{2+\rho}|\mathcal{F}_{i-1}] + \frac{1}{2}\sum_{i=1}^n \big(\mathbf{E}[e^{\lambda\xi_i}|\mathcal{F}_{i-1}] - 1\big)^2 \\ &\leq \lambda^{2+\rho}\varepsilon_n^\rho \sum_{i=1}^n \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] + (\lambda\varepsilon_n)^2 \sum_{i=1}^n \big(\mathbf{E}[e^{\lambda\xi_i}|\mathcal{F}_{i-1}] - 1\big) \\ &\leq \lambda^{2+\rho}\varepsilon_n^\rho\langle X\rangle_n + c_1\lambda^4\varepsilon_n^2\langle X\rangle_n.\end{aligned}$$

Thus

$$\Big|\Psi_n(\lambda) - \frac{\lambda^2}{2}\langle X\rangle_n\Big| \leq \Big(1 + c_1(\lambda\varepsilon_n)^{2-\rho}\Big)\lambda^{2+\rho}\varepsilon_n^\rho\langle X\rangle_n.$$

Combining the last inequality with condition (A2) and the fact $\langle X\rangle_n \leq 2$, we get for all $0 \leq \lambda \leq \varepsilon_n^{-1}$,

$$\Big|\Psi_n(\lambda) - \frac{\lambda^2}{2}\Big| \leq 2\Big(1 + c_1(\lambda\varepsilon_n)^{2-\rho}\Big)\lambda^{2+\rho}\varepsilon_n^\rho + \frac{\lambda^2\delta_n^2}{2},$$

which completes the proof of Lemma 3. □

In the proof of Theorem 1, we make use of the following lemma, which gives us some rates of convergence in the central limit theorem for the conjugate martingale $Y(\lambda)$ under the probability measure $\mathbf{P}_\lambda$. The proof of this lemma is given in Section 10.

Lemma 4 *Assume conditions (A1) and (A2).*

[i] *If $\rho \in (0,1)$, then there is a positive constant α such that for all $0 \leq \lambda \leq \alpha\,\varepsilon_n^{-1}$,*

$$\sup_x \Big|\mathbf{P}_\lambda(Y_n(\lambda) \leq x) - \Phi(x)\Big| \leq c_{\alpha,\rho}\Big((\lambda\varepsilon_n)^\rho + \varepsilon_n^\rho + \delta_n\Big).$$

In particular, it implies that

$$\sup_x \Big|\mathbf{P}(X_n \leq x) - \Phi(x)\Big| \leq c_{\alpha,\rho}\Big(\varepsilon_n^\rho + \delta_n\Big). \tag{4.14}$$

[ii] *If $\rho = 1$, then there is a positive constant α such that for all $0 \leq \lambda \leq \alpha\,\varepsilon_n^{-1}$,*

$$\sup_x \Big|\mathbf{P}_\lambda(Y_n(\lambda) \leq x) - \Phi(x)\Big| \leq c_\alpha\Big(\lambda\varepsilon_n + \varepsilon_n|\ln\varepsilon_n| + \delta_n\Big).$$

In particular, it implies that

$$\sup_x \Big|\mathbf{P}(X_n \leq x) - \Phi(x)\Big| \leq c_\alpha\Big(\varepsilon_n|\log\varepsilon_n| + \delta_n\Big). \tag{4.15}$$

5 Proof of Theorem 1

Theorem 1 will be deduced by the combination of the following two propositions (1 and 2), which are stated and proved respectively in Subsections 5.1 and 5.2. The proof of the propositions are similar to the proofs of Theorems 2.1 and 2.2 of Fan et al. [11]. However, Fan et al. [11] considered the particular case where $\rho = 1$.

5.1 Proof of upper bound in Theorem 1

The following assertion gives an upper bound for moderate deviation probabilities.

Proposition 1 *Assume conditions (A1) and (A2).*

[i] *If* $\rho \in (0,1)$, *then there is a constant* $\alpha > 0$ *such that for all* $0 \leq x \leq \alpha \varepsilon_n^{-1}$,

$$\frac{\mathbf{P}(X_n > x)}{1-\Phi(x)} \leq \exp\left\{ c_{\alpha,\rho,1}\Big(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2 + (1+x)\left(\varepsilon_n^\rho + \delta_n\right)\Big)\right\}. \tag{5.1}$$

[ii] *If* $\rho = 1$, *then there is a constant* $\alpha > 0$ *such that for all* $0 \leq x \leq \alpha \varepsilon_n^{-1}$,

$$\frac{\mathbf{P}(X_n > x)}{1-\Phi(x)} \leq \exp\left\{ c_{\alpha,1,1}\Big(x^3\varepsilon_n + x^2\delta_n^2 + (1+x)\left(\varepsilon_n|\ln\varepsilon_n| + \delta_n\right)\Big)\right\}. \tag{5.2}$$

Proof. For all $0 \leq x < 1$, the assertion follows from (4.14) and (4.15). It remains to prove Proposition 1 for all $1 \leq x \leq \alpha\,\varepsilon_n^{-1}$. Changing the probability measure according to (4.1), we get for all $0 \leq \lambda \leq \varepsilon_n^{-1}$,

$$\begin{aligned}\mathbf{P}(X_n > x) &= \mathbf{E}_\lambda\left[Z_n(\lambda)^{-1}\mathbf{1}_{\{X_n>x\}}\right]\\ &= \mathbf{E}_\lambda\left[\exp\left\{-\lambda X_n + \Psi_n(\lambda)\right\}\mathbf{1}_{\{X_n>x\}}\right]\\ &= \mathbf{E}_\lambda\left[\exp\left\{-\lambda Y_n(\lambda) - \lambda B_n(\lambda) + \Psi_n(\lambda)\right\}\mathbf{1}_{\{Y_n(\lambda)+B_n(\lambda)>x\}}\right].\end{aligned} \tag{5.3}$$

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

$$\lambda + \lambda\delta_n^2 + c\lambda^{1+\rho}\varepsilon_n^\rho = x, \tag{5.4}$$

where c is given by inequality (4.6). The definition of $\overline{\lambda}$ implies that there exist $c_{\alpha,0}, c_{\alpha,1} > 0$, such that for all $1 \leq x \leq \alpha\,\varepsilon_n^{-1}$,

$$c_{\alpha,0}\, x \leq \overline{\lambda} \leq x \tag{5.5}$$

and

$$\overline{\lambda} = x - c_{\alpha,1}|\theta|(x^{1+\rho}\varepsilon_n^\rho + x\delta_n^2) \in [c_{\alpha,0},\ \alpha\,\varepsilon_n^{-1}\,]. \tag{5.6}$$

By Lemma 2, it follows that $B_n(\overline{\lambda}) \leq x$. From (5.3), by Lemma 3 and equality (5.4), we deduce that for all $1 \leq x \leq \alpha\,\varepsilon_n^{-1}$,

$$\mathbf{P}(X_n > x) \leq e^{c_{\alpha,2}\,(\overline{\lambda}^{2+\rho}\varepsilon_n^\rho + \overline{\lambda}^2\delta_n^2) - \overline{\lambda}^2/2}\,\mathbf{E}_{\overline{\lambda}}\big[e^{-\overline{\lambda}Y_n(\overline{\lambda})}\mathbf{1}_{\{Y_n(\overline{\lambda})>0\}}\big]. \tag{5.7}$$

Clearly, it holds

$$\mathbf{E}_{\overline{\lambda}}\big[e^{-\overline{\lambda}Y_n(\overline{\lambda})}\mathbf{1}_{\{Y_n(\overline{\lambda})>0\}}\big] = \int_0^\infty \overline{\lambda}e^{-\overline{\lambda}y}\mathbf{P}_{\overline{\lambda}}(0 < Y_n(\overline{\lambda}) \leq y)dy. \tag{5.8}$$

Similarly, for a standard normal random variable $\mathcal{N}$, we have

$$\mathbf{E}\big[e^{-\overline{\lambda}\mathcal{N}}\mathbf{1}_{\{\mathcal{N}>0\}}\big] = \int_0^\infty \overline{\lambda}e^{-\overline{\lambda}y}\mathbf{P}(0 < \mathcal{N} \leq y)dy. \tag{5.9}$$

From (5.8) and (5.9), it is easy to see that

$$\Big|\mathbf{E}_{\overline{\lambda}}\big[e^{-\overline{\lambda}Y_n(\overline{\lambda})}\mathbf{1}_{\{Y_n(\overline{\lambda})>0\}}\big] - \mathbf{E}\big[e^{-\overline{\lambda}\mathcal{N}}\mathbf{1}_{\{\mathcal{N}>0\}}\big]\Big| \leq 2\sup_y\Big|\mathbf{P}_{\overline{\lambda}}(Y_n(\overline{\lambda}) \leq y) - \Phi(y)\Big|.$$

Using Lemma 4, we get the following bound: for all $1 \leq x \leq \alpha\, \varepsilon_n^{-1}$,

$$\Big| \mathbf{E}_{\overline{\lambda}} \big[e^{-\overline{\lambda} Y_n(\overline{\lambda})} \mathbf{1}_{\{Y_n(\overline{\lambda})>0\}} \big] - \mathbf{E}\big[e^{-\overline{\lambda}\mathcal{N}} \mathbf{1}_{\{\mathcal{N}>0\}} \big] \Big| \leq c_\rho \Big((\overline{\lambda}\varepsilon_n)^\rho + \widetilde{\varepsilon}_n + \delta_n \Big), \tag{5.10}$$

where

$$\widetilde{\varepsilon}_n = \begin{cases} \varepsilon_n^\rho, & \text{if } \rho \in (0,1), \\ \varepsilon_n |\ln \varepsilon_n|, & \text{if } \rho = 1. \end{cases} \tag{5.11}$$

From (5.7) and (5.10), we deduce that for all $1 \leq x \leq \alpha\, \varepsilon_n^{-1}$,

$$\mathbf{P}(X_n > x) \leq e^{c_{\alpha,2}\, (\overline{\lambda}^{2+\rho} \varepsilon_n^\rho + \overline{\lambda}^2 \delta_n^2) - \overline{\lambda}^2/2} \Big(\mathbf{E}\big[e^{-\overline{\lambda}\mathcal{N}} \mathbf{1}_{\{\mathcal{N}>0\}} \big] + c_\rho \Big((\overline{\lambda}\varepsilon_n)^\rho + \widetilde{\varepsilon}_n + \delta_n \Big) \Big).$$

Since

$$e^{-\lambda^2/2} \mathbf{E}\left[e^{-\lambda \mathcal{N}} \mathbf{1}_{\{\mathcal{N}>0\}} \right] = \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-(y+\lambda)^2/2} dy = 1 - \Phi(\lambda) \tag{5.12}$$

and

$$1 - \Phi(\lambda) \geq \frac{1}{\sqrt{2\pi}(1+\lambda)}\, e^{-\lambda^2/2} \geq \frac{c_{\alpha,0}}{\sqrt{2\pi}(1+c_{\alpha,0})} \frac{1}{\lambda} e^{-\lambda^2/2}, \quad \lambda \geq c_{\alpha,0}, \tag{5.13}$$

we have the following upper bound for moderate deviation probabilities: for all $1 \leq x \leq \alpha\, \varepsilon_n^{-1}$,

$$\frac{\mathbf{P}(X_n > x)}{1 - \Phi\left(\overline{\lambda}\right)} \leq e^{c_{\alpha,2}\, (\overline{\lambda}^{2+\rho} \varepsilon_n^\rho + \overline{\lambda}^2 \delta_n^2)} \Big(1 + c_{\alpha,\rho,3}\, (\overline{\lambda}^{1+\rho} \varepsilon_n^\rho + \overline{\lambda}\widetilde{\varepsilon}_n + \overline{\lambda}\delta_n\,) \Big). \tag{5.14}$$

Next, we would like to make a comparison between $1 - \Phi(\overline{\lambda})$ and $1 - \Phi(x)$. By (5.5), (5.6) and (5.13), it follows that

$$\begin{aligned} 1 \leq \frac{\int_{\overline{\lambda}}^\infty \exp\{-t^2/2\} dt}{\int_x^\infty \exp\{-t^2/2\} dt} &= 1 + \frac{\int_{\overline{\lambda}}^x \exp\{-t^2/2\} dt}{\int_x^\infty \exp\{-t^2/2\} dt} \\ &\leq 1 + c_{\alpha,4} x(x - \overline{\lambda}) \exp\{(x^2 - \overline{\lambda}^2)/2\} \\ &\leq \exp\{c_{\alpha,5}\, (x^{2+\rho} \varepsilon_n^\rho + x^2 \delta_n^2)\}. \end{aligned} \tag{5.15}$$

So, it holds

$$1 - \Phi\left(\overline{\lambda}\right) = \big(1 - \Phi(x)\big) \exp\left\{ |\theta_1| c_{\alpha,5}\, (x^{2+\rho} \varepsilon_n^\rho + x^2 \delta_n^2) \right\}. \tag{5.16}$$

Implementing (5.16) in (5.14) and using (5.5), we obtain for all $1 \leq x \leq \alpha\, \varepsilon_n^{-1}$,

$$\begin{aligned} \frac{\mathbf{P}(X_n > x)}{1 - \Phi(x)} &\leq \exp\Big\{ c_{\alpha,6} (x^{2+\rho} \varepsilon_n^\rho + x^2 \delta_n^2) \Big\} \Big(1 + c_{\alpha,\rho,7} \left(x^{1+\rho} \varepsilon_n^\rho + x \widetilde{\varepsilon}_n + x\delta_n \right) \Big) \\ &\leq \exp\Big\{ c_{\alpha,\rho,8} \Big(x^{2+\rho} \varepsilon_n^\rho + x^2 \delta_n^2 + x\, (\widetilde{\varepsilon}_n + \delta_n) \Big) \Big\}. \end{aligned}$$

This completes the proof of Proposition 1. □

5.2 Proof of lower bound in Theorem 1

The following assertion gives a lower bound for moderate deviation probabilities.

Proposition 2 *Assume conditions (A1) and (A2).*

[i] *If $\rho \in (0,1)$, then there is a constant $\alpha > 0$ such that for all $0 \le x \le \alpha \varepsilon_n^{-1}$,*

$$\frac{\mathbf{P}(X_n > x)}{1 - \Phi(x)} \ge \exp\Big\{ - c_{\alpha,\rho,2} \Big(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2 + (1+x)\left(\varepsilon_n^\rho + \delta_n\right) \Big) \Big\}. \tag{5.17}$$

[ii] *If $\rho = 1$, then there is a constant $\alpha > 0$ such that for all $0 \le x \le \alpha \varepsilon_n^{-1}$,*

$$\frac{\mathbf{P}(X_n > x)}{1 - \Phi(x)} \ge \exp\Big\{ - c_{\alpha,1,2} \Big(x^3\varepsilon_n + x^2\delta_n^2 + (1+x)\left(\varepsilon_n|\ln\varepsilon_n| + \delta_n\right) \Big) \Big\}. \tag{5.18}$$

Proof. For all $0 \le x < 1$, the assertion follows from (4.14) and (4.15). It remains to prove Proposition 2 for all $1 \le x \le \alpha\varepsilon_n^{-1}$, where $\alpha > 0$ is a small constant. Let $\underline{\lambda} = \underline{\lambda}(x)$ be the smallest positive solution of the following equation

$$\lambda - \lambda\delta_n^2 - c\lambda^{1+\rho}\varepsilon_n^\rho = x, \tag{5.19}$$

where c is given by inequality (4.6). The definition of $\underline{\lambda}$ implies that for all $1 \le x \le \alpha\varepsilon_n^{-1}$,

$$x \le \underline{\lambda} \le c_{\alpha,1}\, x \tag{5.20}$$

and

$$\underline{\lambda} = x + c_{\alpha,2}|\theta|(x^{1+\rho}\varepsilon_n^\rho + x\delta_n^2) \in [1,\ \varepsilon_n^{-1}]. \tag{5.21}$$

From (5.3), using Lemmas 2, 3 and equality (5.19), we have for all $1 \le x \le \alpha\varepsilon_n^{-1}$,

$$\mathbf{P}(X_n > x) \ge e^{-c_1\,(\underline{\lambda}^{2+\rho}\varepsilon_n^\rho + \underline{\lambda}^2\delta_n^2) - \underline{\lambda}^2/2}\, \mathbf{E}_{\underline{\lambda}}\left[e^{-\underline{\lambda} Y_n(\underline{\lambda})} \mathbf{1}_{\{Y_n(\underline{\lambda}) > 0\}} \right]. \tag{5.22}$$

In the subsequent we distinguish $\underline{\lambda}$ into two cases. First, let $1 \le \underline{\lambda} \le \alpha_1 \min\{\varepsilon_n^{-\rho/(1+\rho)}, \delta_n^{-1}\}$, where $\alpha_1 > 0$ is a small positive constant whose exact value will be given later. Note that inequality (5.10) can be established with $\overline{\lambda}$ replaced by $\underline{\lambda}$, which, in turn, implies that

$$\mathbf{P}(X_n > x) \ge e^{-c_1\,(\underline{\lambda}^{2+\rho}\varepsilon_n^\rho + \underline{\lambda}^2\delta_n^2) - \underline{\lambda}^2/2} \Big(\mathbf{E}\left[e^{-\underline{\lambda}\mathcal{N}} \mathbf{1}_{\{\mathcal{N} > 0\}} \right] - c_{\rho,2}\Big((\underline{\lambda}\varepsilon_n)^\rho + \widetilde{\varepsilon}_n + \delta_n \Big) \Big),$$

where $\widetilde{\varepsilon}_n$ is defined by (5.11). By (5.12) and (5.13), we get the following lower bound on tail probabilities:

$$\frac{\mathbf{P}(X_n > x)}{1 - \Phi(\underline{\lambda})} \ge e^{-c_1\,(\underline{\lambda}^{2+\rho}\varepsilon_n^\rho + \underline{\lambda}^2\delta_n^2)} \Big(1 - c_{\rho,2}\left(\underline{\lambda}^{1+\rho}\varepsilon_n^\rho + \underline{\lambda}\widetilde{\varepsilon}_n + \underline{\lambda}\delta_n^2 \right) \Big). \tag{5.23}$$

Taking $\alpha_1 = \min\{\frac{1}{(8c_{\rho,2})^{1/(1+\rho)}}, \frac{1}{8c_{\rho,2}}\}$, we have for all $1 \le \underline{\lambda} \le \alpha_1 \min\{\varepsilon_n^{-\rho/(1+\rho)}, \delta_n^{-1}\}$,

$$1 - c_{\rho,2}\left(\underline{\lambda}^{1+\rho}\varepsilon_n^\rho + \underline{\lambda}\widetilde{\varepsilon}_n + \underline{\lambda}\delta_n \right) \ge \exp\left\{ -2c_{\rho,2}\left(\underline{\lambda}^{1+\rho}\varepsilon_n^\rho + \underline{\lambda}\widetilde{\varepsilon}_n + \underline{\lambda}\delta_n \right) \right\}. \tag{5.24}$$

Implementing (5.24) in (5.23), we get

$$\frac{\mathbf{P}(X_n > x)}{1 - \Phi(\underline{\lambda})} \ge \exp\Big\{ - c_{\rho,3}\left(\underline{\lambda}^{2+\rho}\varepsilon_n^\rho + \underline{\lambda}\widetilde{\varepsilon}_n + \underline{\lambda}\delta_n + \underline{\lambda}^2\delta_n^2 \right) \Big\} \tag{5.25}$$

which holds for all $1 \le \underline{\lambda} \le \alpha_1 \min\{\varepsilon_n^{-\rho/(1+\rho)}, \delta_n^{-1}\}$.

Next, consider the case of $\alpha_1 \min\{\varepsilon_n^{-\rho/(1+\rho)}, \delta_n^{-1}\} \leq \underline{\lambda} \leq \alpha\varepsilon_n^{-1}$. Let $K \geq 1$ be a constant, whose exact value will be chosen later. It is obvious that

$$\begin{aligned}\mathbf{E}_{\underline{\lambda}}\left[e^{-\underline{\lambda} Y_n(\underline{\lambda})}\mathbf{1}_{\{Y_n(\underline{\lambda})>0\}}\right] &\geq \mathbf{E}_{\underline{\lambda}}\left[e^{-\underline{\lambda} Y_n(\underline{\lambda})}\mathbf{1}_{\{0<Y_n(\underline{\lambda})\leq K\tau\}}\right] \\ &\geq e^{-\underline{\lambda} K\tau}\mathbf{P}_{\underline{\lambda}}\Big(0 < Y_n(\underline{\lambda}) \leq K\tau\Big), \end{aligned}\tag{5.26}$$

where $\tau = (\underline{\lambda}\varepsilon_n)^\rho + \widetilde{\varepsilon}_n + \delta_n$. From Lemma 4, we get

$$\begin{aligned}\mathbf{P}_{\underline{\lambda}}\Big(0 < Y_n(\underline{\lambda}) \leq K\tau\Big) &\geq \mathbf{P}\Big(0 < \mathcal{N} \leq K\tau\Big) - c_{\rho,5}\tau \\ &\geq K\tau e^{-K^2\tau^2/2} - c_{\rho,5}\tau \\ &\geq \Big(K e^{-8K^2\alpha} - c_{\rho,5}\Big)\,\tau.\end{aligned}$$

Taking $\alpha = 1/(16K^2)$, we obtain

$$\mathbf{P}_{\underline{\lambda}}\Big(0 < Y_n(\underline{\lambda}) \leq K\tau\Big) \geq \left(\frac{1}{2}K - c_{\rho,5}\right)\tau.$$

Letting $K \geq 8c_{\rho,5}$, we deduce that

$$\mathbf{P}_{\underline{\lambda}}\Big(0 < Y_n(\underline{\lambda}) \leq K\tau\Big) \geq \frac{3}{8}K\tau \geq \frac{3}{8}K\frac{\max\left\{\underline{\lambda}^{1+\rho}\varepsilon_n^\rho, \underline{\lambda}\delta_n\right\}}{\underline{\lambda}}.$$

Choosing $K = \max\Big\{8c_{\rho,5}, \frac{16\alpha_1^{-1-\rho}}{3\sqrt{\pi}}\Big\}$ and taking into account that $\alpha_1 \min\{\varepsilon_n^{-\rho/(1+\rho)}, \delta_n^{-1}\} \leq \underline{\lambda} \leq \alpha\varepsilon_n^{-1}$, we get

$$\mathbf{P}_{\underline{\lambda}}\Big(0 < Y_n(\underline{\lambda}) \leq K\tau\Big) \geq \frac{2}{\sqrt{\pi}\underline{\lambda}}.$$

Since the inequality

$$\frac{2}{\sqrt{\pi}\lambda}e^{-\lambda^2/2} \geq 1 - \Phi(\lambda)$$

is valid for all $\lambda > 0$, it follows that for all $\alpha_1 \min\{\varepsilon_n^{-\rho/(1+\rho)}, \delta_n^{-1}\} \leq \underline{\lambda} \leq \alpha\varepsilon_n^{-1}$,

$$\mathbf{P}_{\underline{\lambda}}\Big(0 < Y_n(\underline{\lambda}) \leq K\tau\Big) \geq \Big(1 - \Phi(\underline{\lambda})\Big)e^{\underline{\lambda}^2/2}. \tag{5.27}$$

From (5.22), (5.26) and (5.27), we get

$$\frac{\mathbf{P}(X_n > x)}{1 - \Phi(\underline{\lambda})} \geq \exp\left\{ - c_{\alpha,6}\left(\underline{\lambda}^{2+\rho}\varepsilon_n^\rho + \underline{\lambda}\widetilde{\varepsilon}_n + \underline{\lambda}\delta_n + \underline{\lambda}^2\delta_n^2\right)\right\} \tag{5.28}$$

which holds for all $\alpha_1 \min\{\varepsilon_n^{-\rho/(1+\rho)}, \delta_n^{-1}\} \leq \underline{\lambda} \leq \alpha\varepsilon_n^{-1}$.

Combining (5.25) and (5.28) together, we obtain for all $1 \leq \underline{\lambda} \leq \alpha\varepsilon_n^{-1}$,

$$\frac{\mathbf{P}(X_n > x)}{1 - \Phi(\underline{\lambda})} \geq \exp\left\{ - c_{\alpha,\rho,7}\left(\underline{\lambda}^{2+\rho}\varepsilon_n^\rho + \underline{\lambda}\widetilde{\varepsilon}_n + \underline{\lambda}\delta_n + \underline{\lambda}^2\delta_n^2\right)\right\}. \tag{5.29}$$

By a similar argument as in (5.15), it is easy to see that

$$1 - \Phi(\underline{\lambda}) = \Big(1 - \Phi(x)\Big)\exp\left\{-|\theta|c_3\left(x^{2+\rho}\varepsilon_n^\rho + x^2\delta_n^2\right)\right\}. \tag{5.30}$$

Combining (5.20), (5.29) and (5.30) together, we find that for all $1 \leq x \leq \alpha\varepsilon_n^{-1}$,

$$\frac{\mathbf{P}(X_n > x)}{1 - \Phi(x)} \geq \exp\left\{ - c_{\alpha,\rho,8}\left(x^{2+\rho}\varepsilon_n^\rho + x\widetilde{\varepsilon}_n + x\delta_n + x^2\delta_n^2\right)\right\}, \tag{5.31}$$

which gives the conclusion of Proposition 2. $\square$

6 Proof of Corollary 2

To prove Corollary 2, we need the following two-sides bound on tail probabilities of the standard normal random variable:

$$\frac{1}{\sqrt{2\pi}(1+x)}e^{-x^2/2} \leq 1-\Phi(x) \leq \frac{1}{\sqrt{\pi}(1+x)}e^{-x^2/2}, \qquad x \geq 0. \tag{6.1}$$

First, we prove that for any given Borel set $B \subset [0,\infty)$,

$$\limsup_{n\to\infty} \frac{1}{a_n^2} \log \mathbf{P}\left(\frac{1}{a_n}X_n \in B\right) \leq -\inf_{x\in \overline{B}} \frac{x^2}{2}. \tag{6.2}$$

Let $x_0 = \inf_{x\in B} x$. Then it is obvious that $x_0 \geq 0$ and $x_0 \geq \inf_{x\in \overline{B}} x$. By Theorem 1, we deduce that

$$\begin{aligned}
&\mathbf{P}\left(\frac{1}{a_n}X_n \in B\right) \\
&\leq \mathbf{P}\left(X_n \geq a_n x_0\right) \\
&\leq \Big(1-\Phi\left(a_n x_0\right)\Big) \exp\left\{c_\alpha\Big(\left(a_n x_0\right)^{2+\rho} \varepsilon_n^\rho + \left(a_n x_0\right)^2 \delta_n^2 + \left(1+\left(a_n x_0\right)\right)\left(\varepsilon_n^\rho |\ln \varepsilon_n| + \delta_n\right)\Big)\right\}.
\end{aligned}$$

Using (6.1) and the assumption $a_n\epsilon_n \to 0$, we have

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

which gives (6.2).

Next, we prove that for any given Borel set $B \subset [0,\infty)$,

$$\liminf_{n\to\infty} \frac{1}{a_n^2} \log \mathbf{P}\left(\frac{1}{a_n}X_n \in B\right) \geq -\inf_{x\in B^o} \frac{x^2}{2}. \tag{6.3}$$

For any $\varepsilon_1 > 0$, there exists an $x_0 \in B^o$, such that

$$\frac{x_0^2}{2} \leq \inf_{x\in B^o} \frac{x^2}{2} + \varepsilon_1. \tag{6.4}$$

For $x_0 \in B^o$, there exists an $\varepsilon_2 > 0$, such that $(x_0-\varepsilon_2, x_0+\varepsilon_2] \subset B$. Then it is obvious that $x_0 \geq \inf_{x\in B^o} x$. By Theorem 1, we deduce that

$$\begin{aligned}
\mathbf{P}\left(\frac{1}{a_n}X_n \in B\right) \ &\geq\ \mathbf{P}\Big(X_n \in (a_n(x_0-\varepsilon_2), a_n(x_0+\varepsilon_2)]\Big) \\
&\geq \mathbf{P}\Big(X_n > a_n(x_0-\varepsilon_2)\Big) - \mathbf{P}\Big(X_n > a_n(x_0+\varepsilon_2)\Big) \\
&\geq \Big(1-\Phi\left(a_n(x_0-\varepsilon_2)\right)\Big) \exp\bigg\{-c_\alpha\Big(\left(a_n(x_0-\varepsilon_2)\right)^{2+\rho} \varepsilon_n^\rho + \left(a_n(x_0-\varepsilon_2)\right)^2 \delta_n^2 \\
&\qquad\qquad + \left(1+\left(a_n(x_0-\varepsilon_2)\right)\right)\left(\varepsilon_n^\rho |\ln \varepsilon_n| + \delta_n\right)\Big)\bigg\} \\
&\quad -\Big(1-\Phi\left(a_n(x_0+\varepsilon_2)\right)\Big) \exp\bigg\{c_\alpha\Big(\left(a_n(x_0+\varepsilon_2)\right)^{2+\rho} \varepsilon_n^\rho + \left(a_n(x_0+\varepsilon_2)\right)^2 \delta_n^2 \\
&\qquad\qquad + \left(1+\left(a_n(x_0+\varepsilon_2)\right)\right)\left(\varepsilon_n^\rho |\ln \varepsilon_n| + \delta_n\right)\Big)\bigg\} \\
&=: P_{1,n} - P_{2,n}.
\end{aligned}$$

Since $a_n\epsilon_n \to 0$, it is easy to see that $\lim_{n\to\infty} P_{2,n}/P_{1,n} = 0$. Thus for n large enough, it holds

$$\mathbf{P}\left(\frac{1}{a_n}X_n \in B\right) \geq \frac{1}{2}P_{1,n}.$$

Using (6.1) and the assumption $a_n\epsilon_n \to 0$ again, it follows that

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

Letting $\varepsilon_2 \to 0$, we get

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

Since ε_1 can be arbitrary small, we obtain (6.3). □

7 Proof of Theorem 2

To prove Theorem 2, we need the following lemma.

Lemma 5 *Assume the conditions of Theorem 2. Then for all $x \geq 0$,*

$$\mathbf{P}\Big(|W_n| > x\Big) \leq 2\exp\left\{-\frac{x^2}{2(1+\frac{M}{n}+\frac{xL}{3\sqrt{n}})}\right\}. \tag{7.1}$$

Proof. Let $T_0 = \min\{K, \frac{L}{3}\}$. It is easy to see that for all $0 \leq \lambda < T_0$,

$$\begin{aligned}
\mathbf{E}[e^{\lambda\eta_i}|\mathcal{F}_{i-1}] &\leq 1 + \lambda\mathbf{E}[\eta_i|\mathcal{F}_{i-1}] + \frac{\lambda^2}{2}\mathbf{E}[\eta_i^2|\mathcal{F}_{i-1}] + \frac{\lambda^3}{3!}\mathbf{E}[|\eta_i|^3e^{K|\eta_i|}|\mathcal{F}_{i-1}] \\
&\leq 1 + \frac{\lambda^2}{2}(1+\frac{1}{3}\lambda L)\mathbf{E}[\eta_i^2|\mathcal{F}_{i-1}] \\
&\leq \exp\left\{\frac{\lambda^2}{2}(1+\frac{1}{3}\lambda L)\mathbf{E}[\eta_i^2|\mathcal{F}_{i-1}]\right\} \\
&\leq \exp\left\{\frac{\lambda^2}{2(1-\lambda T_0)}\mathbf{E}[\eta_i^2|\mathcal{F}_{i-1}]\right\},
\end{aligned}$$

which implies that for all $0 \leq \lambda < T_0$,

$$\begin{aligned}
&\mathbf{E}\left[\exp\left\{\lambda\sum_{i=1}^n \eta_i - \frac{\lambda^2\,\Xi_n}{2(1-\lambda T_0)}\right\}\right] \\
&\leq \mathbf{E}\left[\exp\left\{\lambda\sum_{i=1}^{n-1}\eta_i - \frac{\lambda^2\,\Xi_{n-1}}{2(1-\lambda T_0)}\right\}\mathbf{E}\left[\exp\left\{\lambda\eta_n - \frac{\lambda^2\mathbf{E}[\eta_n^2|\mathcal{F}_{n-1}]}{2(1-\lambda T_0)}\right\}\Big|\mathcal{F}_{n-1}\right]\right] \\
&\leq \mathbf{E}\left[\exp\left\{\lambda\sum_{i=1}^{n-1}\eta_i - \frac{\lambda^2\,\Xi_{n-1}}{2(1-\lambda T_0)}\right\}\right] \\
&\leq 1,
\end{aligned}$$

where

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

Since $\Xi_n \leq n + M$ a.s., we have for all $x \geq 0$ and all $0 \leq \lambda < T_0$,

$$\begin{aligned}
\mathbf{P}\Big(W_n > x\Big) &= \mathbf{P}\Big(\sum_{i=1}^n \eta_i > x\sqrt{n}\Big) \\
&\leq \mathbf{E}\Big[\exp\Big\{-\lambda x\sqrt{n} + \lambda \sum_{i=1}^n \eta_i - \frac{\lambda^2\, \Xi_n}{2(1-\lambda T_0)} + \frac{\lambda^2(n+M)}{2(1-\lambda T_0)}\Big\}\Big] \\
&\leq \mathbf{E}\Big[\exp\Big\{-\lambda x\sqrt{n} + \frac{\lambda^2(n+M)}{2(1-\lambda T_0)}\Big\}\Big].
\end{aligned}$$

Thus for all $x \geq 0$,

$$\begin{aligned}
\mathbf{P}\Big(W_n > x\Big) &\leq \inf_{0\leq\lambda<T_0} \mathbf{E}\Big[\exp\Big\{-\lambda x\sqrt{n} + \frac{\lambda^2(n+M)}{2(1-\lambda T_0)}\Big\}\Big] \\
&\leq \exp\bigg\{-\frac{x^2}{2(1+M/n+xT_0/\sqrt{n})}\bigg\} \\
&\leq \exp\bigg\{-\frac{x^2}{2(1+\frac{M}{n}+\frac{xL}{3\sqrt{n}})}\bigg\}. \qquad (7.2)
\end{aligned}$$

Similarly, we have for all $x \geq 0$,

$$\mathbf{P}\Big(W_n < -x\Big) \leq \exp\bigg\{-\frac{x^2}{2(1+\frac{M}{n}+\frac{xL}{3\sqrt{n}})}\bigg\}. \qquad (7.3)$$

Combining (7.2) and (7.3) together, we obtain the desired inequality. $\square$

Now we are in position to prove Theorem 2. By Theorem 1, there exist constants $\alpha \in (0,1]$ and $C \geq 1$ such that for all $0 \leq x \leq \alpha\, n^{1/2}$,

$$\frac{\mathbf{P}\left(W_n > x\right)}{1-\Phi\left(x\right)} = \exp\left\{\theta C(1+x^3)\frac{\ln n}{\sqrt{n}}\right\} \qquad (7.4)$$

and

$$\frac{\mathbf{P}\left(W_n < -x\right)}{\Phi\left(-x\right)} = \exp\left\{\theta C(1+x^3)\frac{\ln n}{\sqrt{n}}\right\}, \qquad (7.5)$$

where $|\theta| \leq 1$. By Theorem 1 of Mason and Zhou [18] with $\varepsilon_n = \alpha$ and $K_n = C\ln n$, then whenever $n \geq 64C^2(\ln n)^2$ and

$$|W_n| \leq \frac{\sqrt{n}}{8\ln n},$$

we have

$$|W_n - Z| \leq 2C\Big(W_n^2 + 1\Big)\frac{\ln n}{\sqrt{n}},$$

which gives (3.3). Notice that there exists an integer n_0 such that $n \geq 64C^2(\ln n)^2$ for all $n \geq n_0$.

Next we give the proof of (3.4). By (3.3), we have for all $0 \leq x \leq \frac{C}{32}\, n/(\ln n)^2$,

$$\begin{aligned}\mathbf{P}\Big(\sqrt{n}|W_n - Z|/\ln n > x\Big) &\leq \mathbf{P}\Big(\sqrt{n}|W_n - Z|/\ln n > x, |W_n| \leq \frac{1}{8}\sqrt{n}/\ln n\Big)\\ &\quad + \mathbf{P}\Big(|W_n| > \frac{1}{8}\sqrt{n}/\ln n\Big)\\ &\leq \mathbf{P}\Big(2C\big(W_n^2+1\big) > x\Big) + \mathbf{P}\Big(|W_n| > \frac{1}{8}\sqrt{n}/\ln n\Big)\\ &\leq \mathbf{P}\Big(|W_n| > \sqrt{x/(2C)}\Big) + \mathbf{P}\Big(|W_n| > \frac{1}{8}\sqrt{n}/\ln n\Big). \quad (7.6)\end{aligned}$$

Notice that

$$1 - \Phi(x) \leq \exp\{-x^2/2\}, \quad x \geq 0.$$

When $0 \leq x \leq 2C\alpha^2 n/(8C\ln n)^2, n \geq 2$, by the inequalities (7.4) and (7.5), it holds that

$$\begin{aligned}\mathbf{P}\Big(|W_n| > \sqrt{x/(2C)}\Big) &\leq 2\exp\Big\{-\frac{1}{4}(\sqrt{x/(2C)})^2\Big\}\\ &= \exp\Big\{1 - \frac{1}{8C}x\Big\}, \quad (7.7)\end{aligned}$$

and that

$$\begin{aligned}\mathbf{P}\Big(|W_n| > \frac{1}{8}\sqrt{n}/\ln n\Big) &\leq 2\exp\Big\{-\frac{n}{8\cdot 32(\ln n)^2}\Big\}\\ &\leq \exp\Big\{1 - \frac{C}{8\alpha^2}x\Big\}. \quad (7.8)\end{aligned}$$

Returning to (7.6), we obtain for all $0 \leq x \leq 2C\alpha^2 n/(8C\ln n)^2$,

$$\mathbf{P}\Big(\sqrt{n}|W_n - Z|/\ln n > x\Big) \leq 2\exp\Big\{1 - c'x\Big\}, \quad (7.9)$$

where $c' = \min\{\frac{1}{8C}, \frac{C}{8\alpha^2}\}$. When $x > 2C\alpha^2 n/(8C\ln n)^2$, it holds

$$\mathbf{P}\Big(\sqrt{n}|W_n - Z|/\ln n > x\Big) \leq \mathbf{P}\Big(\sqrt{n}|W_n|/\ln n > x/2\Big) + \mathbf{P}\Big(\sqrt{n}|Z|/\ln n > x/2\Big). \quad (7.10)$$

By Lemma 5, there exists a positive constant λ such that for all $x > 2C\alpha^2 n/(8C\ln n)^2$,

$$\begin{aligned}\mathbf{P}\Big(\sqrt{n}|W_n|/\ln n > x/2\Big) &\leq 2\exp\Big\{-\frac{3}{8L}x\sqrt{n}\frac{\ln n}{\sqrt{n}}\Big\}\\ &\leq \exp\Big\{1 - \frac{3}{8L}x\Big\},\end{aligned}$$

and that

$$\begin{aligned}\mathbf{P}\Big(\sqrt{n}|Z|/\ln n > x/2\Big) &\leq 2\exp\Big\{-\frac{1}{8}x^2\frac{\ln n}{\sqrt{n}}\Big\}\\ &\leq \exp\Big\{1 - \frac{\alpha^2}{256C}x\Big\}.\end{aligned}$$

Returning to (7.10), we have for all $x > 2C\alpha^2 n/(8C\ln n)^2$,

$$\mathbf{P}\Big(\sqrt{n}|W_n - Z|/\ln n > x\Big) \leq 2\exp\Big\{1 - c''x\Big\}, \quad (7.11)$$

where $c'' = \min\{\frac{3}{8L}, \frac{\alpha^2}{256C}\}$. Combining (7.9) and (7.11) together, we get (3.4).

8 Proof of Theorem 3

The main idea of the proof of Theorem 3 is to use m-dependence approximation. We make use of the following lemma of Berbee [1].

Lemma 6 *Let $(Y_i)_{1\leq i\leq n}$ be a sequence of random variables on some probability space and define $\beta^{(i)} = \beta(Y_i, (Y_{i+1}, ..., Y_n))$. Then the probability space can be extended with random variables $\widetilde{Y}_i$ distributed as Y_i such that $(\widetilde{Y}_i)_{1\leq i\leq n}$ are independent and*

$$\mathbf{P}(Y_i \neq \widetilde{Y}_i \textit{ for some } 1 \leq i \leq n) \leq \beta^{(1)} + ... + \beta^{(n-1)}.$$

Now we are in position to prove Theorem 3. Recall $m = \lfloor n^\alpha \rfloor$ and $k = \lfloor n/(2m) \rfloor$. By Lemma 6, there exists a sequence of independent random variables $(\widetilde{Y}_j)_{1\leq j\leq k}$ such that $\widetilde{Y}_j$ and Y_j have the same distribution for each $1 \leq j \leq k$ and

$$\mathbf{P}(Y_i \neq \widetilde{Y}_i \text{ for some } 1 \leq i \leq k) \leq k\beta(m) \leq a_1 \exp\{-0.5a_2 n^{\alpha\,\tau}\}. \tag{8.1}$$

Therefore, we have

$$\left|\mathbf{P}\big(S_n/\sqrt{\mathbf{E}S_n^2} > x\big) - \mathbf{P}\big(\widetilde{S}_n/\sqrt{\mathbf{E}S_n^2} > x\big)\right| \leq a_1 \exp\big\{ - 0.5a_2 n^{\alpha\,\tau}\big\}, \tag{8.2}$$

where $\widetilde{S}_n = \sum_{j=1}^k \widetilde{Y}_j$. By (3.6) and (3.7), we have

$$\mathbf{E}|\widetilde{Y}_i|^{2+\rho} \leq c_1^{2+\rho} c_2^{-2} m^{\rho/2} \mathbf{E}\widetilde{Y}_i^2$$

for all $1 \leq j \leq k$, and

$$\mathrm{Var}(\widetilde{S}_n) \asymp n.$$

By (8.1) and (3.6), it is easy to see that

$$\begin{aligned}
\Big|\mathbf{E}\widetilde{S}_n^2 - \mathbf{E}S_n^2\Big| &= \Big|\mathbf{E}[(\widetilde{S}_n^2 - S_n^2)\mathbf{1}_{\{Y_i\neq\widetilde{Y}_i \text{ for some } 1\leq i\leq k\}}]\Big| \\
&\leq 2\mathbf{E}[e^{\frac{1}{4}a_2 n^{\alpha\tau}} \mathbf{1}_{\{Y_i\neq\widetilde{Y}_i \text{ for some } 1\leq i\leq k\}}] \\
&\quad + \mathbf{E}[\widetilde{S}_n^2 \mathbf{1}_{\{\widetilde{S}_n > e^{\frac{1}{8}a_2 n^{\alpha\tau}}\}}] + \mathbf{E}[S_n^2 \mathbf{1}_{\{S_n > e^{\frac{1}{8}a_2 n^{\alpha\tau}}\}}] \\
&\leq 2e^{\frac{1}{4}a_2 n^{\alpha\tau}} \mathbf{P}(Y_i \neq \widetilde{Y}_i \text{ for some } 1 \leq i \leq k) \\
&\quad + e^{-\frac{\rho}{8}a_2 n^{\alpha\tau}} \mathbf{E}|\widetilde{S}_n|^{2+\rho} + e^{-\frac{\rho}{8}a_2 n^{\alpha\tau}} \mathbf{E}|S_n|^{2+\rho} \\
&\leq O(1) \exp\Big\{ - \frac{\rho}{16} a_2 n^{\alpha\tau}\Big\} \\
&= O(n^{-2}).
\end{aligned}$$

It is obvious that $\widetilde{Y}_j \leq n^\alpha c_3$ a.s. Applying Theorem 1 to $\widetilde{S}_n/\sqrt{\mathbf{E}S_n^2}$, we deduce that there is a constant $\alpha > 0$, such that for all $0 \leq x = o(n^{\frac{1}{2}-\alpha})$,

$$\frac{\mathbf{P}(\widetilde{S}_n/\sqrt{\mathbf{E}S_n^2} > x)}{1 - \Phi(x)} = \exp\left\{\theta_1 c_\rho \frac{(1+x)^{2+\rho}}{n^{\rho(\frac{1}{2}-\alpha)}}\right\}. \tag{8.3}$$

The inequalities (8.2) and (8.3) together implies that

$$\begin{aligned}
\frac{\mathbf{P}(S_n/\sqrt{\mathbf{E}S_n^2} > x)}{1 - \Phi(x)} &= \exp\left\{\theta_1 c_{1,\rho} \frac{(1+x)^{2+\rho}}{n^{\rho(\frac{1}{2}-\alpha)}}\right\} + a_1 \frac{\exp\{-0.5a_2 n^{\alpha\,\tau}\}}{1 - \Phi(x)} \\
&= \exp\left\{\theta_2 c_{2,\rho} \frac{(1+x)^{2+\rho}}{n^{\rho(\frac{1}{2}-\alpha)}}\right\}
\end{aligned}$$

uniformly for $0 \leq x = o(\min\{n^{\frac{1}{2}-\alpha},\, n^{\alpha\tau/2}\})$.

9 Proof of Theorem 4

We only give a proof for the case of $\rho \in (0,1)$. The proof for the case of $\rho = 1$ is similar to the case of $\rho \in (0,1)$. In the proof of theorem, we use the following lemma. The proof of the lemma is similar to the proof of Theorem A.6 of Hall and Heyde [17].

Lemma 7 *Suppose that X and Y are random variables which are $\mathcal{F}_{j+n}^{\infty}$- and $\mathcal{F}_j$-measurable, respectively, and that $\mathbf{E}|X|^p < \infty$, $\mathbf{E}|Y|^q < \infty$, where $p, q > 1, p^{-1} + q^{-1} = 1$. Then*

$$\Big|\mathbf{E}XY - \mathbf{E}X\mathbf{E}Y\Big| \leq 2[\psi(n)]^{1/p}\big(\mathbf{E}|X|^p\big)^{1/p}\big(\mathbf{E}|Y|^q\big)^{1/q}.$$

Denote by $\mathcal{F}_l = \sigma\{\eta_i, 1 \leq i \leq 2ml - m\}$. Then Y_j is $\mathcal{F}_j$-measurable. Since $\mathbf{E}\eta_i = 0$ for all i, it is easy to see that for $1 \leq j \leq k$,

$$\begin{aligned}\Big|\mathbf{E}[Y_j|\mathcal{F}_{j-1}]\Big| &= \Big|\sum_{i=1}^{m}\Big(\mathbf{E}[\eta_{2m(j-1)+i}|\mathcal{F}_{j-1}] - \mathbf{E}\eta_{2m(j-1)+i}\Big)\Big| \\ &\leq \sum_{i=1}^{m}\psi(m+i)\mathbf{E}|\eta_{2m(j-1)+i}| \\ &\leq \sum_{i=1}^{m}\psi(m+i)(\mathbf{E}|\eta_{2m(j-1)+i}|^{2+\rho})^{1/(2+\rho)} \\ &\leq \sum_{i=1}^{m}\psi(m+i)c_1,\end{aligned}$$

where c_1 is defined in (3.6). Thus

$$\Big|\sum_{j=1}^{k}\mathbf{E}[Y_j|\mathcal{F}_{j-1}]\Big| \leq c_1\sum_{j=1}^{k}\sum_{i=1}^{m}\psi(m+i) \leq n\psi(m)c_1.$$

By (3.6), we have

$$\begin{aligned}\mathbf{E}[|Y_j - \mathbf{E}[Y_j|\mathcal{F}_{j-1}]|^{2+\rho}|\mathcal{F}_{j-1}] &\leq 2^{1+\rho}\mathbf{E}[|Y_j|^{2+\rho} + |\mathbf{E}[Y_j|\mathcal{F}_{j-1}]|^{2+\rho}|\mathcal{F}_{j-1}] \\ &\leq 2^{2+\rho}\mathbf{E}[|Y_j|^{2+\rho}|\mathcal{F}_{j-1}] \\ &\leq 2^{2+\rho}(1+\psi(m))\mathbf{E}|Y_j|^{2+\rho} \\ &\leq 2^{2+\rho}(1+\psi(m))m^{1+\rho/2}c_1^{2+\rho}. \end{aligned} \tag{9.1}$$

Notice that $\tau_n \to 0$ implies that $m\psi^2(m) \to 0$ as $n \to \infty$. Similarly, by (3.7), it holds

$$\begin{aligned}\mathbf{E}[(Y_j - \mathbf{E}[Y_j|\mathcal{F}_{j-1}])^2|\mathcal{F}_{j-1}] &= \mathbf{E}[Y_j^2|\mathcal{F}_{j-1}] - (\mathbf{E}[Y_j|\mathcal{F}_{j-1}])^2 \\ &\geq (1-\psi(m))\mathbf{E}Y_j^2 - (\mathbf{E}[Y_j|\mathcal{F}_{j-1}])^2 \\ &\geq \frac{1}{2}c_2^2 m. \end{aligned} \tag{9.2}$$

Combining (9.1) and (9.2), we deduce that

$$\sum_{j=1}^{k}\mathbf{E}[|Y_j - \mathbf{E}[Y_j|\mathcal{F}_{j-1}]|^2|\mathcal{F}_{j-1}] \asymp n,$$

$$\mathbf{E}[|Y_j - \mathbf{E}[Y_j|\mathcal{F}_{j-1}]|^{2+\rho}|\mathcal{F}_{j-1}] \leq c_\rho\, m^{\rho/2}\mathbf{E}[(Y_j - \mathbf{E}[Y_j|\mathcal{F}_{j-1}])^2|\mathcal{F}_{j-1}]$$

and, by Lemma 7,

$$\begin{aligned}
&\Big|\sum_{j=1}^{k}\mathbf{E}[(Y_j-\mathbf{E}[Y_j|\mathcal{F}_{j-1}])^2|\mathcal{F}_{j-1}]-\mathbf{E}S_n^2\Big| \\
&\leq \Big|\sum_{j=1}^{k}\mathbf{E}[(Y_j-\mathbf{E}[Y_j|\mathcal{F}_{j-1}])^2|\mathcal{F}_{j-1}]-\sum_{j=1}^{k}\mathbf{E}Y_j^2\Big|+\Big|\mathbf{E}S_n^2-\sum_{j=1}^{k}\mathbf{E}Y_j^2\Big| \\
&\leq \sum_{j=1}^{k}\Big|\mathbf{E}[Y_j^2|\mathcal{F}_{j-1}]-\mathbf{E}Y_j^2\Big|+\sum_{j=1}^{k}\Big|\mathbf{E}[Y_j|\mathcal{F}_{j-1}]\Big|^2+\sum_{j\neq l}\Big|\mathbf{E}Y_jY_l\Big| \\
&\leq k\psi(m)\mathbf{E}Y_j^2+k\Big|\sum_{i=1}^{m}\psi(m+i)c_1\Big|^2+2\psi(m)^{1/2}\sum_{j\neq l}\sqrt{\mathbf{E}Y_j^2}\sqrt{\mathbf{E}Y_l^2} \\
&\leq 2n\psi(m)c_1^2+nm\psi^2(m)c_1^2+2n\psi(m)^{1/2}kc_1^2.
\end{aligned}$$

Denote by

$$\epsilon_n^2=\psi(m)+m\psi^2(m)+k\psi(m)^{1/2}.$$

Applying Theorem 1 to $X_n := \sum_{j=1}^{k}(Y_j-\mathbf{E}[Y_j|\mathcal{F}_{j-1}])/\sqrt{\mathbf{E}S_n^2}$, we have for all $0\leq x=o(n^{\frac{1}{2}-\alpha})$,

$$\left|\ln\frac{\mathbf{P}(X_n>x)}{1-\Phi(x)}\right|\leq c_\rho\left(\frac{(1+x)^{2+\rho}}{n^{\rho(\frac{1}{2}-\alpha)}}+x^2\epsilon_n^2+(1+x)\Big(\frac{1}{n^{\rho(\frac{1}{2}-\alpha)}}+\epsilon_n\Big)\right). \tag{9.3}$$

Notice that for $x\geq 0$ and $|\varepsilon|\leq 1$,

$$\frac{1-\Phi(x+\varepsilon)}{1-\Phi(x)}=\exp\Big\{O(1)(1+x)|\varepsilon|\Big\}$$

and

$$\Big|\frac{1}{\sqrt{n}}\sum_{j=1}^{k}\mathbf{E}[Y_j|\mathcal{F}_{j-1}]\Big|\leq\sqrt{n}\psi(m)c_1.$$

Thus

$$\left|\ln\frac{\mathbf{P}(S_n/\sqrt{\mathbf{E}S_n^2}>x)}{1-\Phi(x)}\right|\leq c_\rho\left(\frac{(1+x)^{2+\rho}}{n^{\rho(\frac{1}{2}-\alpha)}}+x^2\tau_n^2+(1+x)\Big(\frac{1}{n^{\rho(\frac{1}{2}-\alpha)}}+\tau_n\Big)\right), \tag{9.4}$$

where τ_n^2 is defined by (3.12).

10 Proof of Lemma 4

The proof is a refinement of the proof of Lemma 3.1 of Fan et al. [11]. See also Grama and Haeusler [16] for an earlier result. Notice that Grama and Haeusler assumed that $|\eta_i|\leq 2\varepsilon_n$, while we only assume that η_i has moments of order $2+\rho, \rho>0$ (cf. condition (B1)). Compared to the case of Fan et al. [11], the main difference comes from the control of I_1 defined in (10.7) and the rest of this section is similar to the proof of Lemma 3.1 of Fan et al. [11].

In this section, α denotes any a given positive number, ϑ denotes a real number satisfying $0 \leq \vartheta \leq 1$, and $\varphi(t)$ denotes the density function of the standard normal distribution. For the sake of simplicity, we also denote $Y(\lambda), Y_n(\lambda)$ and $\eta(\lambda)$ by Y, Y_n and η, respectively.

Notice that $|\eta_i|^t \leq 2^{t-1}(|\xi_i|^t + \mathbf{E}_\lambda[|\xi_i||\mathcal{F}_{i-1}]^t)$ for any $2 \leq t \leq 2+\rho$. We have for all $0 \leq \lambda \leq \varepsilon_n^{-1}$,

$$\begin{aligned}\mathbf{E}_\lambda\left[|\eta_i|^{2+\rho}\left|\mathcal{F}_{i-1}\right.\right] &\leq 2^{1+\rho}\,\mathbf{E}_\lambda\left[|\xi_i|^{2+\rho} + \mathbf{E}_\lambda[|\xi_i||\mathcal{F}_{i-1}]^{2+\rho}\left|\mathcal{F}_{i-1}\right.\right] \\ &\leq 2^{2+\rho}\,\mathbf{E}_\lambda[|\xi_i|^{2+\rho}\left|\mathcal{F}_{i-1}\right.].\end{aligned}$$

Using the fact $\mathbf{E}[e^{\lambda\xi_i}|\mathcal{F}_{i-1}] \geq 1$ and condition (A1), we get for $\rho \in (0,1]$ and all $0 \leq \lambda \leq \varepsilon_n^{-1}$,

$$\begin{aligned}\mathbf{E}_\lambda\left[|\eta_i|^{2+\rho}\left|\mathcal{F}_{i-1}\right.\right] &\leq 2^{2+\rho}\,\mathbf{E}[|\xi_i|^{2+\rho}e^{\lambda\xi_i}\left|\mathcal{F}_{i-1}\right.] \\ &\leq 8\,\mathbf{E}[|\xi_i|^{2+\rho}e^{\varepsilon_n^{-1}\xi_i^+}\left|\mathcal{F}_{i-1}\right.].\end{aligned}$$

Denote $\langle Y\rangle_k = \sum_{i\leq k}\mathbf{E}_\lambda[\eta_i^2|\mathcal{F}_{i-1}]$, $\Delta\langle Y\rangle_k = \mathbf{E}_\lambda[\eta_k^2|\mathcal{F}_{k-1}]$ and $\Delta\langle X\rangle_k = \mathbf{E}[\xi_k^2|\mathcal{F}_{k-1}]$. It is easy to see that for $k = 1, ..., n$,

$$\begin{aligned}\Delta\langle Y\rangle_k &= \mathbf{E}_\lambda\left[(\xi_k - b_k(\lambda))^2|\mathcal{F}_{k-1}\right] \\ &= \frac{\mathbf{E}[\xi_k^2 e^{\lambda\xi_k}|\mathcal{F}_{k-1}]}{\mathbf{E}[e^{\lambda\xi_k}|\mathcal{F}_{k-1}]} - \frac{\mathbf{E}[\xi_k e^{\lambda\xi_k}|\mathcal{F}_{k-1}]^2}{\mathbf{E}[e^{\lambda\xi_k}|\mathcal{F}_{k-1}]^2}.\end{aligned} \tag{10.1}$$

Since $\mathbf{E}[e^{\lambda\xi_i}|\mathcal{F}_{i-1}] \geq 1$, we deduce that for all $0 \leq \lambda \leq \varepsilon_n^{-1}$,

$$\begin{aligned}|\Delta\langle Y\rangle_k - \Delta\langle X\rangle_k| &\leq \left|\frac{\mathbf{E}[\xi_k^2 e^{\lambda\xi_k}|\mathcal{F}_{k-1}]}{\mathbf{E}[e^{\lambda\xi_k}|\mathcal{F}_{k-1}]} - \mathbf{E}[\xi_k^2|\mathcal{F}_{k-1}]\right| + \left|\frac{\mathbf{E}[\xi_k e^{\lambda\xi_k}|\mathcal{F}_{k-1}]^2}{\mathbf{E}[e^{\lambda\xi_k}|\mathcal{F}_{k-1}]^2}\right| \\ &\leq \left|\mathbf{E}[\xi_k^2 e^{\lambda\xi_k}|\mathcal{F}_{k-1}] - \mathbf{E}[\xi_k^2|\mathcal{F}_{k-1}]\mathbf{E}[e^{\lambda\xi_k}|\mathcal{F}_{k-1}]\right| \\ &\quad + \Big(\mathbf{E}[\xi_k e^{\lambda\xi_k}|\mathcal{F}_{k-1}]\Big)^2 \\ &= \left|\mathbf{E}[\xi_k^2(e^{\lambda\xi_k} - 1)|\mathcal{F}_{k-1}] - \mathbf{E}[\xi_k^2|\mathcal{F}_{k-1}]\mathbf{E}[(e^{\lambda\xi_k} - 1 - \lambda\xi_k - \frac{1}{2}\lambda^2\xi_k^2)|\mathcal{F}_{k-1}]\right| \\ &\quad + \frac{1}{2}\Big(\lambda\mathbf{E}[\xi_k^2|\mathcal{F}_{k-1}]\Big)^2 + \Big(\lambda\mathbf{E}[\xi_k^2|\mathcal{F}_{k-1}] + \mathbf{E}[\xi_k(e^{\lambda\xi_k} - 1 - \lambda\xi_k)|\mathcal{F}_{k-1}]\Big)^2.\end{aligned}$$

Using condition (A1) and Lemma 1, we have for all $0 \leq \lambda \leq \varepsilon_n^{-1}$,

$$\begin{aligned}|\Delta\langle Y\rangle_k - \Delta\langle X\rangle_k| &\leq c_{\rho,1}\bigg(\lambda^{-2}\mathbf{E}[e^{\lambda\xi_k}|\lambda\xi_k|^{2+\rho}|\mathcal{F}_{k-1}] + \mathbf{E}[\xi_k^2|\mathcal{F}_{k-1}]\mathbf{E}[e^{\lambda\xi_k}|\lambda\xi_k|^{2+\rho}|\mathcal{F}_{k-1}] \\ &\quad + \lambda^2\big(\mathbf{E}[\xi_k^2|\mathcal{F}_{k-1}]\big)^2 + \Big(\mathbf{E}[e^{\lambda\xi_k}\lambda^{1+\rho}|\xi_k|^{2+\rho}|\mathcal{F}_{k-1}]\Big)^2\bigg) \\ &\leq c_{\rho,1}\bigg(\lambda^{-2}\mathbf{E}[e^{\varepsilon_n^{-1}\xi_i^+}|\lambda\xi_k|^{2+\rho}|\mathcal{F}_{k-1}] + \mathbf{E}[\xi_k^2|\mathcal{F}_{k-1}]\mathbf{E}[e^{\varepsilon_n^{-1}\xi_i^+}|\lambda\xi_k|^{2+\rho}|\mathcal{F}_{k-1}] \\ &\quad + \lambda^2\big(\mathbf{E}[\xi_k^2|\mathcal{F}_{k-1}]\big)^2 + \Big(\mathbf{E}[e^{\varepsilon_n^{-1}\xi_i^+}\lambda^{1+\rho}|\xi_k|^{2+\rho}|\mathcal{F}_{k-1}]\Big)^2\bigg) \\ &\leq c_{\rho,2}\Big(\lambda^\rho\varepsilon_n^\rho\mathbf{E}[\xi_k^2|\mathcal{F}_{k-1}] + \lambda^2\varepsilon_n^2\mathbf{E}[\xi_k^2|\mathcal{F}_{k-1}]\Big) \\ &\leq c_\rho(\lambda\varepsilon_n)^\rho\,\Delta\langle X\rangle_k.\end{aligned} \tag{10.2}$$

Therefore

$$|\langle Y\rangle_n - 1| \leq |\langle Y\rangle_n - \langle X\rangle_n| + |\langle X\rangle_n - 1| \leq c_\rho(\lambda\varepsilon_n)^\rho\langle X\rangle_n + \delta_n^2.$$

Thus the martingale Y satisfies the following conditions: for all $0 \leq \lambda \leq \alpha\,\varepsilon_n^{-1}$,

(B1) $\mathbf{E}_\lambda\big[|\eta_i|^{2+\rho}\,\big|\mathcal{F}_{i-1}\big] \leq c\,\varepsilon_n^\rho\,\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}]$;
(B2) $|\langle Y\rangle_n - 1| \leq c_\rho\left((\lambda\varepsilon_n)^\rho + \delta_n^2\right)$.

We first prove Lemma 4 for all $1 \leq \lambda \leq \alpha\,\varepsilon_n^{-1}$. Without loss of generality, we can further assume that $\alpha \leq \frac{1}{4}$, otherwise we take $c \geq 4$ in the assertion of the lemma. Set $T = 1 + \delta_n^2$. We introduce the following modification of the quadratic characteristic $\langle X\rangle$:

$$V_k = \langle X\rangle_k\,\mathbf{1}_{\{k<n\}} + T\mathbf{1}_{\{k=n\}}. \tag{10.3}$$

It is obvious that $V_0 = 0$, $V_n = T$ and that $(V_k, \mathcal{F}_k)_{k=0,\ldots,n}$ is a predictable process. Set

$$\gamma = (\lambda\varepsilon_n)^\rho + \delta_n.$$

Let $c_* \geq 4$ be a constant, whose exact value will be chosen later. Consider the non-increasing discrete time predictable process $A_k = c_*^2\gamma^2 + T - V_k, k = 1, ..., n$. For any fixed $u, x \in \mathbf{R}$ and $y > 0$, set

$$\Phi_u(x, y) = \Phi\Big(\frac{u-x}{\sqrt{y}}\Big). \tag{10.4}$$

We need the following two technical lemmas of Bolthausen [2].

Lemma 8 *Let X and Y be random variables. Then*

$$\sup_u \Big|\mathbf{P}\left(X \leq u\right) - \Phi\left(u\right)\Big| \leq c_1 \sup_u \Big|\mathbf{P}\left(X + Y \leq u\right) - \Phi\left(u\right)\Big| + c_2\Big\|\mathbf{E}\left[Y^2|X\right]\Big\|_\infty^{1/2}.$$

Lemma 9 *Let $G(x)$ be an integrable function on $\mathbf{R}$ of bounded variation $||G||_V$, X be a random variable and a, $b \neq 0$ are real numbers. Then*

$$\mathbf{E}\left[G\left(\frac{X+a}{b}\right)\right] \leq ||G||_V \sup_u \Big|\mathbf{P}\left(X \leq u\right) - \Phi\left(u\right)\Big| + ||G||_1\,|b|,$$

where $||G||_1$ is the $L^1(\mathbf{R})$-norm of $G(x)$.

Let $\mathcal{N}_{c_*^2\gamma^2} = \mathcal{N}(0, c_*\gamma)$ be a normal random variable independent of Y_n. Using a smoothing procedure (which employs Lemma 8), we deduce that

$$\begin{aligned}
\sup_u \Big|\mathbf{P}_\lambda(Y_n \leq u) - \Phi(u)\Big| &\leq c_1 \sup_u \Big|\mathbf{E}_\lambda[\Phi_u(Y_n, A_n)] - \Phi(u)\Big| + c_2\gamma \\
&\leq c_1 \sup_u \Big|\mathbf{E}_\lambda[\Phi_u(Y_n, A_n)] - \mathbf{E}_\lambda[\Phi_u(Y_0, A_0)]\Big| \\
&\quad + c_1 \sup_u \Big|\mathbf{E}_\lambda[\Phi_u(Y_0, A_0)] - \Phi(u)\Big| + c_2\gamma \\
&= c_1 \sup_u \Big|\mathbf{E}_\lambda[\Phi_u(Y_n, A_n)] - \mathbf{E}_\lambda[\Phi_u(Y_0, A_0)]\Big| \\
&\quad + c_1 \sup_u \Big|\Phi\Big(\frac{u}{\sqrt{c_*^2\gamma^2 + T}}\Big) - \Phi(u)\Big| + c_2\gamma \\
&\leq c_1 \sup_u \Big|\mathbf{E}_\lambda[\Phi_u(Y_n, A_n)] - \mathbf{E}_\lambda[\Phi_u(Y_0, A_0)]\Big| + c_3\gamma,
\end{aligned} \tag{10.5}$$

where

$$\mathbf{E}_\lambda[\Phi_u(Y_n, A_n)] = \mathbf{P}_\lambda(Y_n + \mathcal{N}_{c_*^2\gamma^2} \leq u) \quad \text{and} \quad \mathbf{E}_\lambda[\Phi_u(Y_0, A_0)] = \mathbf{P}_\lambda(\mathcal{N}_{c_*^2\gamma^2+T} \leq u).$$

By telescoping, we get

$$\mathbf{E}_\lambda[\Phi_u(Y_n, A_n)] - \mathbf{E}_\lambda[\Phi_u(Y_0, A_0)] = \mathbf{E}_\lambda\Big[\sum_{k=1}^n \Big(\Phi_u(Y_k, A_k) - \Phi_u(Y_{k-1}, A_{k-1})\Big)\Big].$$

Taking into account that $(\eta_i, \mathcal{F}_i)_{i=0,...,n}$ is a $\mathbf{P}_\lambda$-martingale and that

$$\frac{\partial^2}{\partial x^2}\Phi_u(x,y) = 2\frac{\partial}{\partial y}\Phi_u(x,y),$$

we get

$$\mathbf{E}_\lambda[\Phi_u(Y_n, A_n)] - \mathbf{E}_\lambda[\Phi_u(Y_0, A_0)] = I_1 + I_2 - I_3, \tag{10.6}$$

where

$$\begin{aligned} I_1 &= \mathbf{E}_\lambda\Big[\sum_{k=1}^n \Big(\ \Phi_u(Y_k, A_k) - \Phi_u(Y_{k-1}, A_k) \\ &\qquad\qquad - \frac{\partial}{\partial x}\Phi_u(Y_{k-1}, A_k)\eta_k - \frac{1}{2}\frac{\partial^2}{\partial x^2}\Phi_u(Y_{k-1}, A_k)\eta_k^2\Big)\Big], &(10.7)\\ I_2 &= \frac{1}{2}\mathbf{E}_\lambda\Big[\sum_{k=1}^n \frac{\partial^2}{\partial x^2}\Phi_u(Y_{k-1}, A_k)\Big(\Delta\langle Y\rangle_k - \Delta V_k\Big)\Big], &(10.8)\end{aligned}$$

$$I_3 = \mathbf{E}_\lambda\Big[\sum_{k=1}^n \Big(\Phi_u(Y_{k-1}, A_{k-1}) - \Phi_u(Y_{k-1}, A_k) - \frac{\partial}{\partial y}\Phi_u(Y_{k-1}, A_k)\Delta V_k\Big)\Big]. \tag{10.9}$$

Next, we give estimates of I_1, I_2 and I_3. To shorten notations, set

$$T_{k-1} = \frac{u - Y_{k-1}}{\sqrt{A_k}}.$$

a) *Control of* I_1. Let f be a three times differentiable function on $\mathbf{R}$. Then it is easy to see that for all $|\Delta x| \le 1$,

$$\begin{aligned}\Big|f(x+\Delta x) - f(x) - f'(x)\Delta x - \frac{1}{2}f''(x)(\Delta x)^2\Big| &= \Big|\frac{1}{6}f'''(x+\vartheta_1\Delta x)(\Delta x)^3\Big| \\ &\le |f'''(x+\vartheta_1\Delta x)|\,|\Delta x|^{2+\rho}\end{aligned}$$

and

$$\begin{aligned}&\Big|f(x+\Delta x) - f(x) - f'(x)\Delta x - \frac{1}{2}f''(x)(\Delta x)^2\Big| \\ &= \Big|\frac{1}{2}f''(x+\vartheta_2\Delta x)(\Delta x)^2 - \frac{1}{2}f''(x)(\Delta x)^2\Big| \\ &\le \frac{1}{2}\Big(|f''(x+\vartheta_2\Delta x)| + |f''(x)|\Big)\,|\Delta x|^2 \\ &\le |f''(x+\vartheta_3\Delta x)|\,|\Delta x|^2 \\ &\le |f''(x+\vartheta_3\Delta x)|\,|\Delta x|^{2+\rho}\end{aligned}$$

for all $|\Delta x| > 1$. Moreover, if f is bounded, then for all $|\Delta x| > \frac{1}{2}(2+|x|)$,

$$\begin{aligned}
&\Big|f(x+\Delta x) - f(x) - f'(x)\Delta x - \frac{1}{2}f''(x)(\Delta x)^2\Big| \\
&= \Big(\Big|\frac{f(x+\Delta x)-f(x)}{|\Delta x|^{2+\rho}}\Big| + |f'(x)| + \Big|\frac{1}{2}f''(x)\Big|\Big)\,|\Delta x|^{2+\rho} \\
&\le \Big(4\Big|\frac{f(x+\Delta x)-f(x)}{(2+|x|)^2}\Big| + |f'(x)| + |f''(x)|\Big)\,|\Delta x|^{2+\rho} \\
&\le \Big(\frac{c}{(2+|x|)^2} + |f'(x)| + |f''(x)|\Big)\,|\Delta x|^{2+\rho}.
\end{aligned}$$

Taking $f(u) = \Phi(x)$, $x = T_{k-1}$ and $\Delta x = \frac{\eta_k}{\sqrt{A_k}}$, we have

$$\begin{aligned}
|I_1| \le \mathbf{E}_\lambda\Big[\sum_{k=1}^{n} F\Big(T_{k-1} + \frac{\vartheta\eta_k}{\sqrt{A_k}}\Big)\Big|\frac{\eta_k}{\sqrt{A_k}}\Big|^{2+\rho}\mathbf{1}_{\{|\eta_k/\sqrt{A_k}|\le\frac{1}{2}(2+|T_{k-1}|)\}}\Big] \\
+\mathbf{E}_\lambda\Big[\sum_{k=1}^{n} H(T_{k-1})\Big|\frac{\eta_k}{\sqrt{A_k}}\Big|^{2+\rho}\mathbf{1}_{\{|\eta_k/\sqrt{A_k}|>\frac{1}{2}(2+|T_{k-1}|)\}}\Big],
\end{aligned}\tag{10.10}$$

where

$$F\big(t\big) = \max\Big\{\Big|\Phi'''\big(t\big)\Big|, \Big|\Phi''\big(t\big)\Big|\Big\},$$

$$H(x) = \frac{c}{(2+|x|)^2} + |\Phi'(x)| + |\Phi''(x)|$$

and

$$\vartheta = \vartheta_1\mathbf{1}_{\{|\Delta x|\le 1\}} + \vartheta_3\mathbf{1}_{\{|\Delta x|>1\}}.$$

In order to bound $|I_1|$ we distinguish two cases as follows.

Case 1: Assume $|\eta_k/\sqrt{A_k}| \le \frac{1}{2}(2+|T_{k-1}|)$. By the inequality $F(t) \le \varphi(t)(1+t^2)$, we deduce that

$$\begin{aligned}
F\Big(T_{k-1} + \frac{\vartheta\eta_k}{\sqrt{A_k}}\Big) &\le \varphi\Big(T_{k-1} + \frac{\vartheta\eta_k}{\sqrt{A_k}}\Big)\Big(1 + \Big(T_{k-1} + \frac{\vartheta\eta_k}{\sqrt{A_k}}\Big)^2\Big) \\
&\le \sup_{|t-T_{k-1}|\le\frac{1}{2}(2+|T_{k-1}|)} \varphi(t)(1+t^2).
\end{aligned}$$

Define

$$g_1(z) = \sup_{|t-z|\le\frac{1}{2}(2+|z|)} f_1(t),$$

where $f_1(t) = \varphi(t)(1+t^2)$. Then

$$F\Big(T_{k-1} + \frac{\vartheta\eta_k}{\sqrt{A_k}}\Big)\mathbf{1}_{\{|\eta_k/\sqrt{A_k}|\le\frac{1}{2}(2+|T_{k-1}|)\}} \le g_1(T_{k-1}). \tag{10.11}$$

Case 2: Assume $|\eta_k/\sqrt{A_k}| > \frac{1}{2}(2+|T_{k-1}|)$. It is easy to see that

$$H(T_{k-1})\big|\eta_k\big|^{2+\rho}\mathbf{1}_{\{|\eta_k/\sqrt{A_k}|>\frac{1}{2}(2+|T_{k-1}|)\}} \le \frac{c_0\,|\eta_k|^{2+\rho}}{(2+|T_{k-1}|)^2}. \tag{10.12}$$

Next we bound the conditional expectation of $|\eta_k|^{2+\rho}$. Using condition (B1), we get

$$\mathbf{E}_\lambda[|\eta_k|^{2+\rho}|\mathcal{F}_{k-1}] \le c\,\Delta\langle X\rangle_k\,\varepsilon_n^\rho.$$

By the definition of the process V (cf. (10.3)), it follows that $\Delta \langle X\rangle_k \leq \Delta V_k = V_k - V_{k-1}$ and that

$$\mathbf{E}_\lambda[|\eta_k|^{2+\rho}|\mathcal{F}_{k-1}] \leq c\,\Delta V_k\,\varepsilon_n^\rho. \tag{10.13}$$

Thus, from (10.11) and (10.13), we have

$$\mathbf{E}_\lambda\Big[F\Big(T_{k-1}+\frac{\vartheta\eta_k}{\sqrt{A_k}}\Big)\big|\eta_k\big|^{2+\rho}\mathbf{1}_{\{|\eta_k/\sqrt{A_k}|\leq\frac{1}{2}(2+|T_{k-1}|)\}}\Big|\mathcal{F}_{k-1}\Big] \leq c\,g_1(T_{k-1})\Delta V_k\,\varepsilon_n^\rho \tag{10.14}$$

Similarly, from (10.12) and (10.13), we obtain

$$\begin{aligned}
&\mathbf{E}_\lambda\Big[H(T_{k-1})\big|\eta_k\big|^{2+\rho}\mathbf{1}_{\{|\eta_k/\sqrt{A_k}|>\frac{1}{2}(2+|T_{k-1}|)\}}\Big|\mathcal{F}_{k-1}\Big]\\
&\leq \mathbf{E}_\lambda\bigg[\frac{c\,|\eta_k|^{2+\rho}}{(1+|T_{k-1}|)^2}\bigg|\mathcal{F}_{k-1}\bigg]\\
&\leq cg_2(T_{k-1})\Delta V_k\,\varepsilon_n^\rho,
\end{aligned}\tag{10.15}$$

where $g_2(t) = \frac{c_0}{(1+|t|)^2}$. Set $G(t) = c\,(g_1(t)+g_2(t))$. Returning to (10.10), by (10.14) and (10.15), we deduce that

$$|I_1| \leq J_1, \tag{10.16}$$

where

$$J_1 = c\,\varepsilon_n^\rho\,\mathbf{E}_\lambda\Big[\sum_{k=1}^n \frac{1}{A_k^{1+\rho/2}}G\,(T_{k-1})\,\Delta V_k\Big]. \tag{10.17}$$

To bound J_1, we introduce the time change τ_t as follows: for any real $t\in[0,T]$,

$$\tau_t = \min\{k\leq n : V_k > t\},\quad \text{where}\quad \min\emptyset = n. \tag{10.18}$$

Clearly, for any $t\in[0,T]$, the stopping time τ_t is predictable. Let $(\sigma_k)_{k=1,\dots,n+1}$ be the increasing sequence of moments when the increasing stepwise function τ_t, $t\in[0,T]$, has jumps. It is obvious that $\Delta V_k = \int_{[\sigma_k,\sigma_{k+1})} dt$ and that $k=\tau_t$, for $t\in[\sigma_k,\sigma_{k+1})$. Since $\tau_T = n$, we get

$$\begin{aligned}
\sum_{k=1}^n \frac{1}{A_k^{1+\rho/2}}\,G\,(T_{k-1})\,\Delta V_k &= \sum_{k=1}^n \int_{[\sigma_k,\sigma_{k+1})} \frac{1}{A_{\tau_t}^{1+\rho/2}}\,G\,(T_{\tau_t-1})\,dt\\
&= \int_0^T \frac{1}{A_{\tau_t}^{1+\rho/2}}\,G\,(T_{\tau_t-1})\,dt.
\end{aligned}$$

Set $a_t = c_*^2\gamma^2 + T - t$. Since $\Delta V_{\tau_t} \leq \varepsilon_n^2 + 2\delta_n^2$ (cf. Lemma 1), we deduce that

$$t \leq V_{\tau_t} \leq V_{\tau_t-1} + \Delta V_{\tau_t} \leq t + \varepsilon_n^2 + 2\delta_n^2, \quad t\in[0,T]. \tag{10.19}$$

Taking into account that $c_*\geq 4$, we obtain

$$\frac{1}{2}\,a_t \leq A_{\tau_t} = c_*^2\gamma^2 + T - V_{\tau_t} \leq a_t, \quad t\in[0,T]. \tag{10.20}$$

Note that $G(z)$ is symmetric and is non-increasing in $z\geq 0$. The bounds (10.20) implies that

$$J_1 \leq c\,\varepsilon_n^\rho \int_0^T \frac{1}{a_t^{1+\rho/2}}\,\mathbf{E}_\lambda\Big[G\Big(\frac{u-Y_{\tau_t-1}}{a_t^{1/2}}\Big)\Big]dt. \tag{10.21}$$

Notice that $G(t)$ is also an integrable function of bounded variation on $\mathbf{R}$. By Lemma 9, it is easy to see that

$$\mathbf{E}_\lambda\left[G\left(\frac{u-Y_{\tau_t-1}}{a_t{}^{1/2}}\right)\right] \leq c_1 \sup_z \Big|\mathbf{P}_\lambda(Y_{\tau_t-1}\leq z)-\Phi(z)\Big| + c_2\sqrt{a_t}. \qquad (10.22)$$

Since $V_{\tau_t-1}=V_{\tau_t}-\Delta V_{\tau_t}$, $V_{\tau_t}\geq t$ (cf. (10.19)) and $\Delta V_{\tau_t}\leq \varepsilon_n^2+2\,\delta_n^2$, we get

$$V_n - V_{\tau_t-1} \leq V_n - V_{\tau_t} + \Delta V_{\tau_t} \leq c_*(\varepsilon_n^2+\delta_n^2)+T-t \leq a_t. \qquad (10.23)$$

Thus

$$\begin{aligned}
\mathbf{E}_\lambda\left[(Y_n-Y_{\tau_t-1})^2|\mathcal{F}_{\tau_t-1}\right] &= \mathbf{E}_\lambda\Bigg[\sum_{k=\tau_t}^n \mathbf{E}_\lambda[\eta_k^2|\mathcal{F}_{k-1}]\Bigg|\mathcal{F}_{\tau_t-1}\Bigg]\\
&\leq c\,\mathbf{E}_\lambda\Bigg[\sum_{k=\tau_t}^n \Delta\left\langle X\right\rangle_k\Bigg|\mathcal{F}_{\tau_t-1}\Bigg]\\
&= c\,\mathbf{E}_\lambda\big[\left\langle X\right\rangle_n - \left\langle X\right\rangle_{\tau_t-1}|\mathcal{F}_{\tau_t-1}\big]\\
&\leq c\,\mathbf{E}_\lambda[V_n - V_{\tau_t-1}|\mathcal{F}_{\tau_t-1}]\\
&\leq c\,a_t.
\end{aligned}$$

Then, by Lemma 8, we find that for any $t\in[0,T]$,

$$\sup_z\Big|\mathbf{P}_\lambda(Y_{\tau_t-1}\leq z)-\Phi(z)\Big| \leq c_3\sup_z\Big|\mathbf{P}_\lambda(Y_n\leq z)-\Phi(z)\Big| + c_4\sqrt{a_t}. \qquad (10.24)$$

Combining (10.21), (10.22) and (10.24) together, we obtain

$$J_1 \leq c_5\,\varepsilon_n^\rho\int_0^T\frac{dt}{a_t^{1+\rho/2}}\sup_z\Big|\mathbf{P}_\lambda(Y_n\leq z)-\Phi(z)\Big| + c_6\,\varepsilon_n^\rho\int_0^T\frac{dt}{a_t^{(1+\rho)/2}}. \qquad (10.25)$$

By (10.23) and elementary computations, we see that (since $\lambda\geq 1$ and $\rho\in(0,1]$)

$$\int_0^T\frac{dt}{a_t^{1+\rho/2}} \leq \int_0^T\frac{dt}{(c_*(\varepsilon_n^2+\delta_n^2)+T-t)^{1+\rho/2}} \leq \frac{c_\rho}{c_*^\rho\varepsilon_n^\rho}. \qquad (10.26)$$

and

$$\int_0^T\frac{dt}{a_t^{(1+\rho)/2}} = \begin{cases} c_\rho, & \text{if } \rho\in(0,1),\\ c\,|\log\varepsilon_n|, & \text{if } \rho=1.\end{cases}$$

Then

$$|I_1| \leq J_1 \leq \frac{c_{\rho,1}}{c_*^\rho}\sup_z\Big|\mathbf{P}(Y_n\leq z)-\Phi(z)\Big| + c_{\rho,2}\,\widetilde{\varepsilon}_n, \qquad (10.27)$$

where

$$\widetilde{\varepsilon}_n = \begin{cases} \varepsilon_n^\rho, & \text{if } \rho\in(0,1),\\ \varepsilon_n|\ln\varepsilon_n|, & \text{if } \rho=1.\end{cases} \qquad (10.28)$$

***b)** Control of* I_2. Set $\widetilde{G}(z)=\sup_{|v|\leq 2}\psi(z+v)$, where $\psi(z)=\varphi(z)(1+z^2)^{3/2}$. Since $\Delta A_k=-\Delta V_k$, we have

$$|I_2| \leq I_{2,1}+I_{2,2},$$

where

$$I_{2,1} = \mathbf{E}_\lambda \Bigg[\sum_{k=1}^n \frac{1}{2A_k} \left| \varphi'\left(T_{k-1}\right) \left(\Delta V_k - \Delta \left\langle X \right\rangle_k\right) \right| \Bigg],$$

$$I_{2,2} = \mathbf{E}_\lambda \Bigg[\sum_{k=1}^n \frac{1}{2A_k} \left| \varphi'\left(T_{k-1}\right) \left(\Delta \left\langle Y \right\rangle_k - \Delta \left\langle X \right\rangle_k\right) \right| \Bigg].$$

Proceeding in the same way as for estimating I_1, we get the estimations

$$|I_{2,1}| \leq \frac{c_1}{c_*} \sup_z \Big| \mathbf{P}_\lambda(Y_n \leq z) - \Phi(z) \Big| + c_2 \gamma$$

and

$$|I_{2,2}| \leq \frac{c_3}{c_*} \sup_z \Big| \mathbf{P}_\lambda(Y_n \leq z) - \Phi(z) \Big| + c_4 \left(\lambda \varepsilon_n\right)^\rho.$$

Collecting the bounds for $I_{2,1}$ and $I_{2,2}$, we get

$$|I_2| \leq \frac{c_5}{c_*} \sup_z \Big| \mathbf{P}_\lambda(Y_n \leq z) - \Phi(z) \Big| + c_6 \gamma. \tag{10.29}$$

c) *Control of* I_3. By Taylor's expansion, it holds

$$I_3 = \frac{1}{8} \mathbf{E}_\lambda \Bigg[\sum_{k=1}^n \frac{1}{(A_k - \vartheta_k \Delta A_k)^2} \varphi''' \left(\frac{u - Y_{k-1}}{\sqrt{A_k - \vartheta_k \Delta A_k}} \right) \Delta A_k^2 \Bigg].$$

Since $|\Delta A_k| = \Delta V_k \leq \varepsilon_n^2 + 2\,\delta_n^2$ and $c_* \geq 4$, it follows that

$$A_k \leq A_k - \vartheta_k \Delta A_k \leq c_*^2 \gamma^2 + T - V_k + 12\gamma^2 \leq 2A_k. \tag{10.30}$$

Using (10.30) and the inequalities $|\varphi'''(z)| \leq \widetilde{G}(z)$, we get

$$|I_3| \leq c\,(\varepsilon_n^2 + 2\,\delta_n^2) \mathbf{E}_\lambda \Bigg[\sum_{k=1}^n \frac{1}{A_k^2}\ \widetilde{G} \left(\frac{T_{k-1}}{\sqrt{2}} \right) \Delta V_k \Bigg].$$

Proceeding in the same way as for estimating J_1 in (10.17), we obtain

$$|I_3| \leq \frac{c_1}{c_*} \sup_z \Big| \mathbf{P}_\lambda(Y_n \leq z) - \Phi(z) \Big| + c_2 \gamma. \tag{10.31}$$

We are now in a position to end the proof of Lemma 4. From (10.6), using (10.27), (10.29) and (10.31), we obtain

$$\begin{aligned} &\Big| \mathbf{E}_\lambda[\Phi_u(Y_n, A_n)] - \mathbf{E}_\lambda[\Phi_u(Y_0, A_0)] \Big| \\ &\qquad\qquad \leq \frac{c_{\rho,3}}{c_*^\rho} \sup_z \Big| \mathbf{P}_\lambda(Y_n \leq z) - \Phi(z) \Big| + c_{\rho,4} \Big(\left(\lambda \varepsilon_n\right)^\rho + \widetilde{\varepsilon}_n + \delta_n \Big), \end{aligned}$$

where $\widetilde{\varepsilon}_n$ is defined by (10.28). Implementing the last bound into (10.5), we get

$$\sup_z \Big| \mathbf{P}_\lambda(Y_n \leq z) - \Phi(z) \Big| \leq \frac{c_{\rho,3}}{c_*^\rho} \sup_z \Big| \mathbf{P}_\lambda(Y_n \leq z) - \Phi(z) \Big| + c_{\rho,4} \Big(\left(\lambda \varepsilon_n\right)^\rho + \widetilde{\varepsilon}_n + \delta_n \Big),$$

from which, choosing $c_*^\rho = \max\{2c_{\rho,3}, 4\}$, we deduce that

$$\sup_z \Big| \mathbf{P}_\lambda(Y_n \leq z) - \Phi(z) \Big| \leq 2c_{\rho,4} \Big(\left(\lambda \varepsilon_n\right)^\rho + \widetilde{\varepsilon}_n + \delta_n \Big), \tag{10.32}$$

which proves Lemma 4 for all $1 \leq \lambda \leq \alpha\, \varepsilon_n^{-1}$. For all $0 \leq \lambda < 1$, we can prove Lemma 4 similarly by taking $\gamma = \varepsilon_n^\rho + \delta_n$. □

Acknowledgements

We would like to thank the associated editor for his helpful suggestions on the structure of paper. We also deeply indebted to the anonymous referees for their helpful comments.

References

1. Berbee, H. 1987. Convergence rates in the strong law for bounded mixing sequences. *Probab. Theory Relat. Fields*, **74**, 255–270.
2. Bolthausen, E., 1982. Exact convergence rates in some martingale central limit theorems. *Ann. Probab.*, **10**, 672–688.
3. Bose, A., 1986a. Certain non-uniform rates of convergence to normality for martingale differences. *J. Statist. Plann. Inference*, **14**, 155–167.
4. Bose, A., 1986b. Certain non-uniform rates of convergence to normality for a restricted class of martingales. *Stochastics*, **16**, 279–294.
5. Chen, X., Shao, Q.M., Wu, W.B., Xu, L.H., 2016. Self-normalized Cramér-type moderate deviations under dependence. *Ann. Statist.*, **44**(4): 1593–1617.
6. Cramér, H., 1938. Sur un nouveau théorème-limite de la théorie des probabilités. *Actualite's Sci. Indust.*, **736**, 5–23.
7. Djellout, H., 2002. Moderate deviations for martingale differences and applications to $\phi-$mixing sequences. *Stochastic Stochastic Rep.*, **73**, 37–63.
8. Dembo, A., 1996. Moderate deviations for martingales with bounded jumps. *Electronic Commun. Probab.* **1**: 11–17.
9. Eichelsbacher, P., Löwe, M., 2017. Lindeberg's method for moderate deviations and random summation. arXiv:1705.03837.
10. Esscher, F., 1932. On the probability function in the collective theory of risk. *Skand. Aktuarie Tidskr.* **15**, 175–195.
11. Fan, X., Grama, I., Liu, Q., 2013. Cramér large deviation expansions for martingales under Bernstein's condition. *Stochastic Process. Appl.* **123**, 3919–3942.
12. Fan, X., 2017. Sharp large deviation for sums of bounded from above random variables. *Sci. China Math.* **60**(12), 2465–2480.
13. Fan, X., 2019. Exact Berry-Esseen bounds for martingales. *J. Math. Anal. Appl.* **469**(2), 1028–1044.
14. Feller, W., 1943. Generalization of a probability limit theorem of Cramér. *Trans. Amer. Math. Soc.*, 361–372.
15. Gao, F.Q., 1996. Moderate deviations for martingales and mixing random processes. *Stochastic Process. Appl.*, **61**, 263–275.
16. Grama, I., Haeusler, E., 2000. Large deviations for martingales via Cramér's method. *Stochastic Process. Appl.*, **85**, 279–293.
17. Hall, P., Heyde, C.C., 1980. *Martingale Limit Theory and Its Application.* Academic Press.
18. Mason, D.M., Zhou, H.H. 2012. Quantile coupling inequalities and their applications. *Probab. Surveys* **9**: 439–479.
19. Petrov, V.V., 1954. A generalization of Cramér's limit theorem. *Uspekhi Math. Nauk*, **9**, 195–202.
20. Petrov, V.V., 1975. *Sums of Independent Random Variables.* Springer-Verlag. Berlin.
21. Račkauskas, A., 1990. On probabilities of large deviations for martingales. *Liet. Mat. Rink.*, **30**, 784–795.
22. Račkauskas, A., 1995. Large deviations for martingales with some applications. *Acta Appl. Math.*, **38**, 109–129.
23. Račkauskas, A., 1997. Limit theorems for large deviations probabilites of certain quadratic forms. *Lithuanian Math. J.*, **37**, 402–415.
24. Sakhanenko, A.I., 1985. Convergence rate in the invariance principle for non-identically distributed variables with exponential moments. In: *Advance in Probability Theory: Limit Theorems for Sums of Random Variables* (A.A. Brovkov, ed.) 2-73. Springer, New York.
25. Saulis, L., Statulevičius, V.A., 1978. *Limit theorems for large deviations.* Kluwer Academic Publishers.
26. Shao, Q.M., Yu, H. 1996. Weak convergence for weighted empirical processes of depedent sequences. *Ann. Probab.*, **24**, 2098–2127.