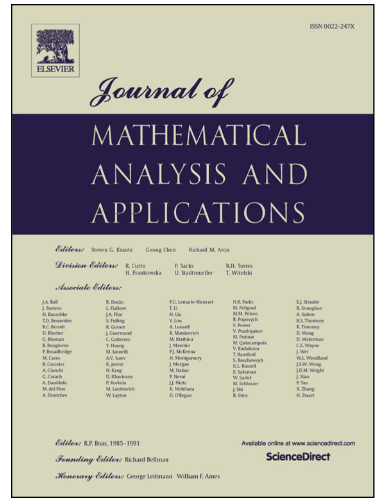
Accepted Manuscript



Exact rates of convergence in some martingale central limit theorems

Xiequan Fan

PII: S0022-247X(18)30798-4
DOI: https://doi.org/10.1016/j.jmaa.2018.09.049
Reference: YJMAA 22569

To appear in: *Journal of Mathematical Analysis and Applications*

Received date: 30 May 2017

Please cite this article in press as: X. Fan, Exact rates of convergence in some martingale central limit theorems, *J. Math. Anal. Appl.* (2018), https://doi.org/10.1016/j.jmaa.2018.09.049

Exact rates of convergence in some martingale central limit theorems

Xiequan Fan

Center for Applied Mathematics, Tianjin University, 300072 Tianjin, China;

Regularity Team, Inria, France

Abstract

Renz [14], Ouchti [12], El Machkouri and Ouchti [4] and Mourrat [13] have established some tight bounds on the rate of convergence in the central limit theorem for martingales. In the present paper a modification of the methods, developed by Bolthausen [1] and Grama and Haeusler [7], is applied for obtaining exact rates of convergence in the central limit theorem for martingales with differences having conditional moments of order $2+\rho, \rho>0$. Our results generalise and strengthen the bounds mentioned above.

Keywords: Martingales, Central limit theorem, Berry-Esseen bounds

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

1. Introduction

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

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

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

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

Define

$$D(X_n) = \sup_{x \in \mathbf{R}} \Big| \mathbf{P}(X_n \leq x) - \Phi(x) \Big|,$$

where $\Phi(x)$ is the distribution function of the standard normal random variable. Denote by $\xrightarrow{\mathbf{P}}$ convergence in probability. According to the basic results of martingale central limit theory (see the

*Corresponding author.
E-mail: fanxiequan@hotmail.com (X. Fan).

monograph Hall and Heyde [10]), the "conditional Lindeberg condition"

$$\sum_{i=1}^{n} \mathbf{E}[\xi_i^2 \mathbf{1}_{\{|\xi_i| \geq \varepsilon\}} | \mathcal{F}_{i-1}] \xrightarrow{\mathbf{P}} 0, \qquad \text{as } n \to \infty \text{ for each } \varepsilon > 0,$$

and the "conditional normalizing condition"

$$\langle X \rangle_n \xrightarrow{\mathbf{P}} 1, \qquad \text{as } n \to \infty,$$

together implies that

$$D(X_n) \longrightarrow 0, \qquad \text{as } \ n \to \infty.$$

In this paper we are interested in bounds of the speed of convergence in central limit theorem, usually termed "Berry-Esseen bounds".

For general martingales, we first recall the following Berry-Esseen bound due to Heyde and Brown [9]. For $1 < p \leq 2$, Heyde and Brown proved that

$$D(X_n) \leq C_p \Big(\mathbf{E}[|\langle X \rangle_n - 1|^p] + \sum_{i=1}^{n} \mathbf{E}[|\xi_i|^{2p}] \Big)^{1/(2p+1)}, \tag{3}$$

where C_p depends only on p. The proof of Heyde and Brown is based on the martingale version of the Skorokhod embedding scheme. This method seems to be unsuited to obtain (3) for $p > 2$. Using a method developed by Bolthausen [1], Haeusler [8] gave an extension of (3) to all $p > 1$. See also Joos [11]. Moreover, Haeusler also gave an example to show that the bound (3) is optimal under the stated condition, that is there exists a sequence of martingale differences $(\xi_k, \mathcal{F}_k)_{k \geq 0}$, such that for all n large enough,

$$D(X_n) \Big(\mathbf{E}[|\langle X \rangle_n - 1|^p] + \sum_{i=1}^{n} \mathbf{E}[|\xi_i|^{2p}] \Big)^{-1/(2p+1)} \ \geq \ c_p,$$

where c_p is a positive constant and does not depend on n. For more interesting Berry-Esseen bounds for martingales, we refer to Dedecker and Merlevède [2], where the authors consider the rates of convergence for linear statistics $X_n = \sum_{i \in \mathbb{Z}} c_{n,i} \xi_i$ based on stationary martingale differences. Their rates are (most of the time) optimal in term of Wasserstein distances. As an application, using the comparison between the uniform distance and the Kantorovith distance, it leads to a Berry-Esseen bound of order $n^{-1/4}\sqrt{\log n}$ when the ξ_i's have a moment of order 3 (see the rate (1.8) of [2]). This Berry-Esseen bound provides the best rate of convergence (up to the $\sqrt{\log n}$ term) under the stated condition. Indeed, Bolthausen [1] gave a counter-example showing that the rate $n^{-1/4}$ in the Berry-Esseen bound cannot be improved when ξ_i's have finite moments of order 3.

However, for martingales having bounded differences, the bound (3) is not the best possible. In fact, an earlier result of Bolthausen [1] sates that if $|\xi_i| \leq \epsilon$ and $\langle X \rangle_n = 1$ a.s., then

$$D(X_n) \leq C \epsilon^3 n \log n, \tag{4}$$

where C is a constant. Moreover, Bolthausen [1] also showed that there exists a sequence of martingale differences satisfying $|\xi_i| \leq 2/\sqrt{n}$ and $\langle X \rangle_n = 1$ a.s., such that for all n large enough,

$$D(X_n)\sqrt{n}/\log n \ \geq \ c, \tag{5}$$

where c is a positive constant and does not depend on n. This means the bound (4) is optimal in the case that ϵ is of order $1/\sqrt{n}$. Relaxing the condition $\langle X\rangle_n = 1$ a.s., Bolthausen [1] then proved that if $|\xi_i| \leq \epsilon$ a.s., then

$$D(X_n) \leq C\Big(\epsilon^3 n \log n + \min\{||\langle X\rangle_n - 1||_1^{1/3}, ||\langle X\rangle_n - 1||_\infty^{1/2}\}\Big). \tag{6}$$

It seems that the term $||\langle X\rangle_n - 1||_1^{1/3}$ in the last bound should be replaced by $||\langle X\rangle_n - 1||_1^{1/3} + \epsilon^{2/3}$; see Mourrat [13]. (Indeed, in the proof of Bolthausen's corollary, we found a term γ^2 is missing for the estimation of $\mathbf{E}[(\widehat{S} - S)^2]$; see [1] for details.)

If $\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] = 1/n$ and $\mathbf{E}[|\xi_i|^{2+\rho}|\mathcal{F}_{i-1}] \leq 1/n^{1+\rho/2}$ a.s. for some number $\rho \in (0,1]$ and all $i = 1, ..., n$, Renz [14] has obtained the following Berry-Esseen bound:

$$D(X_n) \ \leq\ C_\rho \, \varepsilon_n, \tag{7}$$

where the constant C_ρ depends only on ρ and

$$\varepsilon_n = \begin{cases} n^{-\rho/2}, & \text{if } \rho \in (0,1), \\ n^{-1/2}\log n, & \text{if } \rho = 1. \end{cases}$$

Moreover, Renz also showed that there exists a sequence of martingale differences satisfying his conditions, such that for all n large enough,

$$D(X_n)\varepsilon_n^{-1} \ \geq\ c, \tag{8}$$

where c is a positive constant and does not depend on n. This means the bound (7) is exact.

With Bolthausen's method, El Machkouri and Ouchti [4] improved the term $\epsilon^3 n \log n$ in (6) to $\epsilon \log n$, that is if $\mathbf{E}[|\xi_i|^3|\mathcal{F}_{i-1}] \leq \epsilon\, \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}]$ a.s., then

$$D(X_n) \leq C\Big(\epsilon \, \log n + ||\langle X\rangle_n - 1||_\infty^{1/2}\Big). \tag{9}$$

They also proved a result with term $||\langle X\rangle_n - 1||_1^{1/3}$.

Following Bolthausen [1] again, Mourrat [13] has obtained that if $|\xi_i| \leq \epsilon$ a.s., then for $p \geq 1$,

$$D(X_n) \ \leq\ C_p \left(\epsilon^3 n \log n + \epsilon^{2p/(2p+1)} + \mathbf{E}[|\,\langle X\rangle_n - 1|^p]^{1/(2p+1)}\right), \tag{10}$$

where C_p is a constant depending only on p. Notice that Mourrat [13] has extended the term $\min\{||\langle X\rangle_n - 1||_1^{1/3}, ||\langle X\rangle_n - 1||_\infty^{1/2}\}$ of Bolthausen [1] to the more general term $\mathbf{E}[|\,\langle X\rangle_n - 1|^p]^{1/(2p+1)} + \epsilon^{2p/(2p+1)}$. Moreover, he also has justified the optimality of the term $\mathbf{E}[|\,\langle X\rangle_n - 1|^p]^{1/(2p+1)}$.

In this paper we give an improvement on the inequality of El Machkouri and Ouchti (9) and Mourrat's inequality (10). Our result also generalises the inequality of Renz (7). With the method of Grama and Haeusler [7], we prove that if there exist two positive numbers ρ and ϵ, such that

$$\mathbf{E}[|\xi_i|^{2+\rho}|\mathcal{F}_{i-1}] \leq \epsilon^\rho \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] \ \text{ a.s. for all } i = 1, ..., n, \tag{11}$$

then

$$D(X_n) \ \leq\ C_\rho \left(\gamma + ||\langle X\rangle_n - 1||_\infty^{1/2}\right), \tag{12}$$

where C_ρ is a constant depending only on ρ and

$$\gamma = \begin{cases} \epsilon^\rho, & \text{if } \rho \in (0,1), \\ \epsilon\,|\log \epsilon|, & \text{if } \rho \geq 1. \end{cases}$$

We also justify the optimality of the term γ. Then with the method of Bolthausen [1], we obtain a significant improvement of Mourrat's inequality (10) by dropping the term $\epsilon^3 n \log n$: If $|\xi_i| \leq \epsilon$ a.s., then for any $p \geq 1$,

$$D(X_n) \;\leq\; C_p \left(\epsilon^{2p/(2p+1)} + \mathbf{E}[|\left\langle X \right\rangle_n - 1|^p]^{1/(2p+1)}\right), \tag{13}$$

where C_p is a constant depending only on p.

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

Throughout the paper, 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.

2. Main Results

In the sequel we shall use the following conditions:

(A1) There exist two positive numbers ρ and $\epsilon \in (0, \frac{1}{2}]$, such that for all $1 \leq i \leq n$,

$$\mathbf{E}[|\xi_i|^{2+\rho}|\mathcal{F}_{i-1}] \leq \epsilon^\rho \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] \quad \text{a.s.;}$$

(A2) There exists a number $\delta \in [0, \frac{1}{2}]$, such that $|\left\langle X \right\rangle_n - 1| \leq \delta^2$ a.s.

Let us comment on conditions (A1) and (A2).

1. Note that in the case of normalized sums of i.i.d. random variables, conditions (A1) and (A2) are satisfied with $\epsilon = \frac{1}{\sigma\sqrt{n}}$ and $\delta = 0$.
2. In the case of martingales, ϵ and δ usually depend on n such that $\epsilon = \epsilon_n \to 0$ and $\delta = \delta_n \to 0$ as $n \to \infty$. It is also worth noting that the bounded differences, that is $|\xi_i| \leq \epsilon$ a.s. for all i, satisfy condition (A1).
3. Assume that $(Y_i, \mathcal{F}_i)_{i\geq 1}$ is a sequence of martingale differences satisfying
$$\mathbf{E}[|Y_i|^{2+\rho}|\mathcal{F}_{i-1}] \leq C^\rho \mathbf{E}[Y_i^2|\mathcal{F}_{i-1}]$$
for a positive absolute constant C and all $i \geq 1$. Let $S_n = \sum_{i=1}^n Y_i$ and $s_n = \sqrt{\mathbf{E}[S_n^2]}$. Then it is easy to verify that condition (A1) is satisfied with $\xi_i = Y_i/s_n$ and $\epsilon = C/s_n$. In particular, if $(Y_i, \mathcal{F}_i)_{i\geq 1}$ is a stationary sequence, then $\epsilon = O(1/\sqrt{n})$ as $n \to \infty$.
4. Condition (A1) is satisfied for separately Lipschitz functions of independent random variables. Let $f : \mathcal{X}^n \mapsto \mathbf{R}$ be separately Lipschitz, such that
$$|f(x_1, x_2, \ldots, x_n) - f(x_1', x_2', \ldots, x_n')| \leq d(x_1, x_1') + d(x_2, x_2') + \cdots + d(x_n, x_n')\,. \tag{14}$$

Let then

$$X_n := f(\eta_1, \dots, \eta_n) - \mathbf{E}[f(\eta_1, \dots, \eta_n)] \,, \tag{15}$$

where $\eta_1, \dots, \eta_n$ is a sequence of independent random variables. We also introduce the natural filtration of the chain, that is $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and for $k \in \mathbf{N}$, $\mathcal{F}_k = \sigma(\eta_1, \eta_2, \dots, \eta_k)$. Define then

$$g_k(\eta_1, \dots, \eta_k) = \mathbf{E}[f(\eta_1, \dots, \eta_n)|\mathcal{F}_k] \,, \tag{16}$$

and

$$\xi_k = g_k(\eta_1, \dots, \eta_k) - g_{k-1}(\eta_1, \dots, \eta_{k-1}) \,. \tag{17}$$

For $k \in [1, n-1]$, let

$$X_k := \xi_1 + \xi_2 + \cdots + \xi_k \,,$$

and note that, by definition of the ξ_k's, the functional X_n introduced in (15) satisfies

$$X_n = \xi_1 + \xi_2 + \cdots + \xi_n \,.$$

Hence X_k is a martingale adapted to the filtration $\mathcal{F}_k$. It is easy to verify that $\xi_1, ..., \xi_n$ satisfy condition (A1). Indeed, for all $1 \le i \le n$,

$$\begin{aligned} \mathbf{E}[|\xi_i|^{2+\rho}|\mathcal{F}_{i-1}] &= \mathbf{E}[|\mathbf{E}[f(\eta_1, \dots, \eta_n)|\mathcal{F}_i] - \mathbf{E}[f(\eta_1, \dots, \eta_n)|\mathcal{F}_{i-1}]|^\rho \xi_i^2 \,|\mathcal{F}_{i-1}] \\ &= \mathbf{E}[|\mathbf{E}[f(\eta_1, \dots, \eta_n)|\mathcal{F}_i] - \mathbf{E}[f(\eta_1, \dots, \eta_i', \dots, \eta_n)|\mathcal{F}_i]|^\rho \xi_i^2 \,|\mathcal{F}_{i-1}] \\ &\le \mathbf{E}[(\mathbf{E}[d(\eta_i, \eta_i')|\mathcal{F}_i])^\rho \xi_i^2 \,|\mathcal{F}_{i-1}] \\ &= \mathbf{E}[(\mathbf{E}[d(\eta_i, \eta_i')|\eta_i])^\rho] \; \mathbf{E}[\xi_i^2 \,|\mathcal{F}_{i-1}], \end{aligned} \tag{18}$$

where $(\eta_1', \dots, \eta_n')$ is an independent copy of $(\eta_1, \dots, \eta_n)$. Hence, condition (A1) is satisfied with $\epsilon = \max_{1 \le i \le n} \mathbf{E}[(\mathbf{E}[d(\eta_i, \eta_i')|\eta_i])^\rho]^{1/\rho}$. In particular, by Jensen's inequality, it holds $\epsilon \le \max_{1 \le i \le n} \mathbf{E}[d(\eta_i, \eta_i')]$ for $\rho \in (0, 1]$.

Our first result is the following Berry-Esseen bounds for martingales.

Theorem 2.1. *Assume conditions (A1) and (A2).*

- *If $\rho \in (0, 1)$, then*

$$D(X_n) \le c_\rho \Big(\epsilon^\rho + \delta \Big). \tag{19}$$

- *If $\rho \in [1, \infty)$, then*

$$D(X_n) \le c \, \Big(\epsilon |\log \epsilon| + \delta \Big). \tag{20}$$

We justify the optimality of the term ϵ^ρ of (19). Let $n = \lfloor \epsilon^{-2} \rfloor$ be the integer part of ϵ^{-2} and $\rho \in (0, 1)$. Renz's inequality (5) shows that there exists a sequence of martingale differences satisfying condition (A1) and $\langle X \rangle_n = 1$ a.s., such that for all ϵ small enough,

$$\epsilon^{-\rho} D(X_n) \ge n^{\rho/2} D(X_n) \ge c, \tag{21}$$

where the constant $c > 0$ does not depend on ϵ.

Notice that, for bounded martingale differences, condition (A1) holds with $\rho = 1$. By Bolthausen's inequality (5) with $n = [\epsilon^{-2}]$, there exists a sequence of martingale differences satisfying $|\xi_i| \leq 3\epsilon$ and $\langle X \rangle_n = 1$ a.s., such that for all ϵ small enough,

$$(3\epsilon\,|\log 3\epsilon|)^{-1} D(X_n) \geq \frac{1}{4} D(X_n)\sqrt{n}/\log n \geq c, \tag{22}$$

where the constant $c > 0$ does not depend on ϵ. Thus the term $\epsilon\,|\log \epsilon|$ of (20) is exact even for bounded martingale differences.

Under the conditions (A1) and (A2), the order of the term $\epsilon\,|\log \epsilon|$ in (20) is less than the order of the term $\epsilon^3 n \log n$ in Bolthausen's inequality (6). Indeed, by condition (A2), we have $3/4 \leq \langle X \rangle_n \leq n\epsilon^2$ a.s. (see Lemma 3.2) and then $\epsilon \geq \sqrt{3/(4n)}$. For $\epsilon \leq 1/2$, it is easy to see that $\epsilon^3 n \log n \geq 3\,\epsilon|\log \epsilon|/4$. Moreover, $\epsilon^3 n \log n$ may converge to infinity while $\epsilon\,|\log \epsilon|$ converges to 0 as $\epsilon \to 0$ and $n \to \infty$. For instance, if ϵ is of the order $n^{-1/3}$ as $n \to \infty$, then it is obvious that $\epsilon\,|\log \epsilon| = O(n^{-1/3} \log n)$ while $\epsilon^3 n \log n \geq \log n$. Thus the term $\epsilon\,|\log \epsilon|$ is much smaller than $\epsilon^3 n \log n$. Similarly, the order of $\epsilon\,|\log \epsilon|$ is also better than the order of $\epsilon \log n$ in (9) of El Machkouri and Ouchti [4].

For martingales with bounded differences, inequality (20) has been established earlier in Grama [5, 6]. Under the conditional Bernstein condition, that is

$$|\mathbf{E}[\xi_i^k|\mathcal{F}_{i-1}]| \leq \frac{1}{2} k! \epsilon^{k-2} \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] \ \text{ a.s. for } k \geq 3 \ \text{ and } \ 1 \leq i \leq n,$$

instead of condition (A1), Fan, Grama and Liu [3] have obtained the Berry-Esseen bound (20). Note that the conditional Bernstein condition implies that ξ_i has conditional exponential moment. Now we only assume that ξ_i has conditional moment of order 3.

Using Theorem 2.2, we have the following Berry-Esseen bounds similar to the results of Ouchti [12]. Following the notations of Ouchti [12], let $v(n)$ denote either

$$\sup\{k : \ \langle X \rangle_k \leq 1\} \quad \text{ or } \quad \inf\{k : \ \langle X \rangle_k \geq 1\}.$$

Corollary 2.1. *Assume conditions (A1) and $\langle X \rangle_n \geq 1$ a.s.*

- *If $\rho \in (0, 1)$, then*
$$D(X_{v(n)}) \leq c_\rho\, \epsilon^\rho. \tag{23}$$
- *If $\rho \in [1, \infty)$, then*
$$D(X_{v(n)}) \leq c\, \epsilon |\log \epsilon|. \tag{24}$$

Inequality (24) significantly improves an earlier result of Ouchti [12] under the following condition

$$\mathbf{E}[|\xi_i|^3|\mathcal{F}_{i-1}] \leq n^{-1/2} \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] \ \text{ a.s. for all } i \geq 1.$$

Ouchti has obtained a convergence rate in central limit theorem of order $n^{-1/4}$, while (24) gives a convergence rate of order $n^{-1/2} \log n$.

Relaxing condition (A2), we have the following estimation.

Theorem 2.2. *Assume condition (A1). Let $p \geq 1$.*

- *If $\rho \in (0,1)$, then*

$$D(X_n) \le c_{p,\rho} \left(\epsilon^{\rho} + \Big(\mathbf{E}[|\langle X\rangle_n - 1|^p] + \mathbf{E}[\max_{1\le i\le n} |\xi_i|^{2p}] \Big)^{1/(2p+1)} \right). \tag{25}$$

- *If $\rho \in [1,\infty)$, then*

$$D(X_n) \le c_{p} \left(\epsilon |\log \epsilon| + \Big(\mathbf{E}[|\langle X\rangle_n - 1|^p] + \mathbf{E}[\max_{1\le i\le n} |\xi_i|^{2p}] \Big)^{1/(2p+1)} \right). \tag{26}$$

Notice that $\mathbf{E}[\max_{1\le i\le n} |\xi_i|^{2p}] \le \sum_{i=1}^n \mathbf{E}[|\xi_i|^{2p}]$. Therefore, our bounds are usually smaller than the bound of Haeusler (3). For instance, if $|\xi_i| \le \epsilon$ a.s. for all $i = 1, ..., n$, then $\mathbf{E}[\max_{1\le i\le n} |\xi_i|^{2p}] \le \epsilon^{2p}$, while $\sum_{i=1}^n \mathbf{E}[|\xi_i|^{2p}] \le n\epsilon^{2p}$.

For martingales having bounded differences, Theorem 2.2 implies the following corollary.

Corollary 2.2. *Assume $|\xi_i| \le \epsilon$ a.s. for all $i \in [0,n]$. Then for any $p \ge 1$,*

$$D(X_n) \le c_p \Big(\epsilon^{2p} + \mathbf{E}[|\langle X\rangle_n - 1|^p] \Big)^{1/(2p+1)}. \tag{27}$$

Clearly, the term $\epsilon^3 n \log n$ appearing in Mourrat's inequality (10) does not appear any more in (27). When $\epsilon \to 0$ and $\epsilon \ge \sqrt[3/7]{1/(n\log n)}$, it holds $\epsilon^{2p/(2p+1)} \le \epsilon^3 n \log n$ for any $p \ge 1$. Moreover, when $\epsilon \to 0$ and $\epsilon \ge 1/\sqrt[3]{n}$, we have $\epsilon^3 n \log n \to \infty$ and $\epsilon^{2p/(2p+1)} \to 0$. Thus our bound (27) is significantly smaller than the bound of Mourrat (10).

3. Proofs of Theorems

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

In the proofs of theorems, we will make use of the following two lemmas. The first lemma shows that we may assume $\rho \in (0,1]$ in condition (A1).

Lemma 3.1. *If there exists an $s > 2$, such that*

$$\mathbf{E}[|\xi_i|^s | \mathcal{F}_{i-1}] \le \epsilon^{s-2}\, \mathbf{E}[\xi_i^2 | \mathcal{F}_{i-1}], \tag{28}$$

then, for any $t \in [2,s)$,

$$\mathbf{E}[|\xi_i|^t | \mathcal{F}_{i-1}] \le \epsilon^{t-2}\, \mathbf{E}[\xi_i^2 | \mathcal{F}_{i-1}]. \tag{29}$$

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

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

Solving the last equations, we get

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

By Hölder's inequality and (28), it is easy to see that

$$\begin{aligned}
\mathbf{E}[|\xi_i|^t|\mathcal{F}_{i-1}] &= \mathbf{E}[|\xi_i|^l|\xi_i|^{t-l}|\mathcal{F}_{i-1}] \\
&\leq (\mathbf{E}[|\xi_i|^{lp}|\mathcal{F}_{i-1}])^{1/p}(\mathbf{E}[|\xi_i|^{(t-l)q}|\mathcal{F}_{i-1}])^{1/q} \\
&\leq (\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}])^{1/p}(\mathbf{E}[|\xi_i|^s|\mathcal{F}_{i-1}])^{1/q} \\
&\leq (\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}])^{1/p}(\epsilon^{s-2}\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}])^{1/q} \\
&\leq \epsilon^{(s-2)/q}\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] \\
&= \epsilon^{t-2}\ \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}].
\end{aligned}$$

This completes the proof of lemma. □

The following lemma shows that under condition (A1), ξ_i has a bounded conditional variance.

Lemma 3.2. *If there exists an $s > 2$, such that*

$$\mathbf{E}[|\xi_i|^s|\mathcal{F}_{i-1}] \leq \epsilon^{s-2}\ \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}], \tag{30}$$

then

$$\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] \leq \epsilon^2. \tag{31}$$

In particular, condition (A1) implies (31).

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

$$\begin{aligned}
(\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}])^{s/2} &\leq \mathbf{E}[|\xi_i|^s|\mathcal{F}_{i-1}] \\
&\leq \epsilon^{s-2}\ \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}].
\end{aligned}$$

Thus

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

which implies (31). □

3.1. Proof of Theorem 2.1

Theorem 2.1 is a refinement of Lemma 3.3 of Grama and Haeusler [7] where it is assumed that ξ_i's are bounded, which is a particular case of condition (A1). See also Lemma 3.1 of Fan, Grama and Liu [3]. Compared to the proofs of Grama and Haeusler [7] and Fan, Grama and Liu [3], the main challenge of our proof comes from the control of I_1 defined in (38).

By Lemma 3.1, we only need to consider the case of $\rho \in (0, 1]$. Set $T = 1 + \delta^2$, and introduce a modification of the conditional variance $\langle X \rangle$ as follows:

$$V_k = \langle X \rangle_k\, \mathbf{1}_{\{k<n\}} + T\mathbf{1}_{\{k=n\}}. \tag{32}$$

It is obvious that $V_0 = 0$, $V_n = T$, and that $(V_k, \mathcal{F}_k)_{k=0,\dots,n}$ is a predictable process. For simplicity of notations, denote

$$\gamma = \begin{cases} \epsilon + \delta, & \text{if } \rho \in (0, 1), \\ \epsilon\, |\log \epsilon| + \delta, & \text{if } \rho = 1. \end{cases}$$

Let c_* be a constant depending on ρ, whose value will be chosen later. Define the following non-increasing discrete time predictable process

$$A_k = c_*^2\gamma^2 + T - V_k, \qquad k = 1, ..., n.$$

In particular, we have $A_0 = c_*^2\gamma^2 + T$ and $A_n = c_*^2\gamma^2$. Moreover, for $u, x \in \mathbf{R}$, and $y > 0$, set, for brevity,

$$\Phi_u(x, y) = \Phi\Big(\frac{u - x}{\sqrt{y}}\Big). \tag{33}$$

Let $\mathcal{N} = \mathcal{N}(0, 1)$ be a standard normal random variable, which is independent of X_n. Using a smoothing procedure, by Lemma 3.4, we get

$$\begin{aligned}
\sup_u \Big|\mathbf{P}(X_n \le u) - \Phi(u)\Big| &\le c_1 \sup_u \Big|\mathbf{P}(X_n + c_*\gamma\mathcal{N} \le u) - \Phi(u)\Big| + c_2\gamma \\
&= c_1 \sup_u \Big|\mathbf{E}[\Phi_u(X_n, A_n)] - \Phi(u)\Big| + c_2\gamma \\
&\le c_1 \sup_u \Big|\mathbf{E}[\Phi_u(X_n, A_n)] - \mathbf{E}[\Phi_u(X_0, A_0)]\Big| \\
&\quad + c_1 \sup_u \Big|\mathbf{E}[\Phi_u(X_0, A_0)] - \Phi(u)\Big| + c_2\gamma \\
&= c_1 \sup_u \Big|\mathbf{E}[\Phi_u(X_n, A_n)] - \mathbf{E}[\Phi_u(X_0, A_0)]\Big| \\
&\quad + c_1 \sup_u \Big|\Phi\Big(\frac{u}{\sqrt{c_*^2\gamma^2 + T}}\Big) - \Phi(u)\Big| + c_2\gamma. \qquad (34)
\end{aligned}$$

Since $T = 1 + \delta^2$, it is easy to see that

$$\Big|\Phi\Big(\frac{u}{\sqrt{c_*^2\gamma^2 + T}}\Big) - \Phi(u)\Big| \le c_3\Big|\frac{1}{\sqrt{c_*^2\gamma^2 + T}} - 1\Big| \le c_4\gamma. \tag{35}$$

Returning to (34), we obtain

$$\sup_u \Big|\mathbf{P}(X_n \le u) - \Phi(u)\Big| \le c_1 \sup_u \Big|\mathbf{E}[\Phi_u(X_n, A_n)] - \mathbf{E}[\Phi_u(X_0, A_0)]\Big| + c_5\gamma. \tag{36}$$

By a simple telescoping, we deduce that

$$\mathbf{E}[\Phi_u(X_n, A_n)] - \mathbf{E}[\Phi_u(X_0, A_0)] = \mathbf{E}\Big[\sum_{k=1}^{n}\Big(\Phi_u(X_k, A_k) - \Phi_u(X_{k-1}, A_{k-1})\Big)\Big].$$

Using the fact

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

we obtain

$$\mathbf{E}[\Phi_u(X_n, A_n)] - \mathbf{E}[\Phi_u(X_0, A_0)] = I_1 + I_2 - I_3, \tag{37}$$

where

$$I_1 = \mathbf{E}\Big[\sum_{k=1}^{n}\Big(\Phi_u(X_k, A_k) - \Phi_u(X_{k-1}, A_k) - \frac{\partial}{\partial x}\Phi_u(X_{k-1}, A_k)\xi_k - \frac{1}{2}\frac{\partial^2}{\partial x^2}\Phi_u(X_{k-1}, A_k)\xi_k^2\Big)\Big], \tag{38}$$

$$I_2 = \frac{1}{2}\mathbf{E}\Big[\sum_{k=1}^{n}\frac{\partial^2}{\partial x^2}\Phi_u(X_{k-1}, A_k)\Big(\Delta\langle X\rangle_k - \Delta V_k\Big)\Big], \tag{39}$$

$$I_3 = \mathbf{E}\Big[\sum_{k=1}^{n}\Big(\Phi_u(X_{k-1}, A_{k-1}) - \Phi_u(X_{k-1}, A_k) - \frac{\partial}{\partial y}\Phi_u(X_{k-1}, A_k)\Delta V_k\Big)\Big], \tag{40}$$

where $\Delta\langle X\rangle_k = \langle X\rangle_k - \langle X\rangle_{k-1}$.

Next, we give the estimates of I_1, I_2 and I_3. To this end, we introduce the following notations. Denote by φ the density function of the standard normal random variable. Moreover, ϑ_i's stand for some values or random variables satisfying $0 \leq \vartheta_i \leq 1$, which may represent different values at different places.

a) *Control of* I_1. To shorten notations, set $T_{k-1} = (u - X_{k-1})/\sqrt{A_k}$. It is obvious that

$$\begin{aligned} R_k =&: \Phi_u(X_k, A_k) - \Phi_u(X_{k-1}, A_k) - \frac{\partial}{\partial x}\Phi_u(X_{k-1}, A_k)\xi_k - \frac{1}{2}\frac{\partial^2}{\partial x^2}\Phi_u(X_{k-1}, A_k)\xi_k^2 \\ =&\ \Phi\Big(T_{k-1} + \frac{\xi_k}{\sqrt{A_k}}\Big) - \Phi(T_{k-1}) - \Phi'(T_{k-1})\frac{\xi_k}{\sqrt{A_k}} - \frac{1}{2}\Phi''(T_{k-1})\Big(\frac{\xi_k}{\sqrt{A_k}}\Big)^2. \end{aligned}$$

We distinguish two cases as follows.

Case 1: $|\xi_k/\sqrt{A_k}| \leq 1 + |T_{k-1}|/2$. By a three-term Taylor expansion, it is easy to see that if $|\xi_k/\sqrt{A_k}| \leq 1$, then

$$\begin{aligned} \Big|R_k\Big| &= \Big|\frac{1}{6}\Phi'''\Big(T_{k-1} + \vartheta\frac{\xi_k}{\sqrt{A_k}}\Big)\Big|\frac{\xi_k}{\sqrt{A_k}}\Big|^3\Big| \\ &\leq \Big|\Phi'''\Big(T_{k-1} + \vartheta\frac{\xi_k}{\sqrt{A_k}}\Big)\Big|\Big|\frac{\xi_k}{\sqrt{A_k}}\Big|^{2+\rho}. \end{aligned}$$

It is also easy to see that if $|\xi_k/\sqrt{A_k}| > 1$, then

$$\begin{aligned} \Big|R_k\Big| &\leq \frac{1}{2}\Big(\Big|\Phi''\Big(T_{k-1} + \vartheta\frac{\xi_k}{\sqrt{A_k}}\Big)\Big| + \Big|\Phi''(T_{k-1})\Big|\Big)\Big(\frac{\xi_k}{\sqrt{A_k}}\Big)^2 \\ &\leq \Big|\Phi''\Big(T_{k-1} + \vartheta'\frac{\xi_k}{\sqrt{A_k}}\Big)\Big|\Big(\frac{\xi_k}{\sqrt{A_k}}\Big)^2 \\ &\leq \Big|\Phi''\Big(T_{k-1} + \vartheta'\frac{\xi_k}{\sqrt{A_k}}\Big)\Big|\Big|\frac{\xi_k}{\sqrt{A_k}}\Big|^{2+\rho}, \end{aligned}$$

where

$$\vartheta' = \begin{cases} \vartheta, & \text{if } \Big|\Phi''\Big(T_{k-1} + \vartheta\frac{\xi_k}{\sqrt{A_k}}\Big)\Big| \geq \Big|\Phi''(T_{k-1})\Big|, \\ 0, & \text{if } \Big|\Phi''\Big(T_{k-1} + \vartheta\frac{\xi_k}{\sqrt{A_k}}\Big)\Big| < \Big|\Phi''(T_{k-1})\Big|. \end{cases}$$

By the inequality $\max\{|\Phi''(t)|, |\Phi'''(t)|\} \leq \varphi(t)(1+t^2)$, it follows that

$$\begin{aligned}\Big|R_k \mathbf{1}_{\{|\xi_k/\sqrt{A_k}|\leq 1+|T_{k-1}|/2\}}\Big| &\leq \varphi\Big(T_{k-1}+\vartheta_1 \frac{\xi_k}{\sqrt{A_k}}\Big)\Big(1+\Big(T_{k-1}+\vartheta_1\frac{\xi_k}{\sqrt{A_k}}\Big)^2\Big)\\ &\leq g_1(T_{k-1}),\end{aligned}$$

where

$$g_1(z) = \sup_{|t-z|\leq 1+|z|/2} \varphi(t)(1+t^2).$$

It is easy to see that $g_1(z)$ is a non-increasing in $z \geq 0$, and that $g_1(z)$ satisfies

$$\Big|R_k \mathbf{1}_{\{|\xi_k/\sqrt{A_k}|\leq 1+|T_{k-1}|/2\}}\Big| \leq g_1(T_{k-1})\Big|\frac{\xi_k}{\sqrt{A_k}}\Big|^{2+\rho}\mathbf{1}_{\{|\xi_k/\sqrt{A_k}|\leq 1+|T_{k-1}|/2\}}. \tag{41}$$

Case 2: $|\xi_k/\sqrt{A_k}| > 1+|T_{k-1}|/2$. It is easy to see that for $|\Delta x| > 1+|x|/2$,

$$\begin{aligned}&\Big|\Phi(x+\Delta x)-\Phi(x)-\Phi'(x)\Delta x-\frac{1}{2}\Phi''(x)(\Delta x)^2\Big|\\ &= \Big(\Big|\frac{\Phi(x+\Delta x)-\Phi(x)}{|\Delta x|^{2+\rho}}\Big| + |\Phi'(x)|+|\Phi''(x)|\Big)\,|\Delta x|^{2+\rho}\\ &\leq \Big(4\Big|\frac{\Phi(x+\Delta x)-\Phi(x)}{(2+|x|)^2}\Big| + |\Phi'(x)|+|\Phi''(x)|\Big)\,|\Delta x|^{2+\rho}\\ &\leq \Big(\frac{c_1}{(2+|x|)^2} + |\Phi'(x)|+|\Phi''(x)|\Big)\,|\Delta x|^{2+\rho}\\ &\leq \frac{c_2}{(2+|x|)^2}\,|\Delta x|^{2+\rho}.\end{aligned}$$

Therefore,

$$\Big|R_k \mathbf{1}_{\{|\xi_k/\sqrt{A_k}|> 1+|T_{k-1}|/2\}}\Big| \leq g_2(T_{k-1})\Big|\frac{\xi_k}{\sqrt{A_k}}\Big|^{2+\rho}\mathbf{1}_{\{|\xi_k/\sqrt{A_k}|> 1+|T_{k-1}|/2\}}, \tag{42}$$

where

$$g_2(z) = \frac{c_2}{(2+|z|)^2}.$$

Set

$$G(z) = g_1(z) + g_2(z).$$

Combining (41) and (42) together, we obtain

$$\Big|R_k\Big| \leq G(T_{k-1})\Big|\frac{\xi_k}{\sqrt{A_k}}\Big|^{2+\rho}, \tag{43}$$

and thus

$$\Big|I_1\Big| = \Big|\mathbf{E}\Big[\sum_{k=1}^n R_k\Big]\Big| \leq \mathbf{E}\Big[\sum_{k=1}^n G(T_{k-1})\Big|\frac{\xi_k}{\sqrt{A_k}}\Big|^{2+\rho}\Big]. \tag{44}$$

Now we consider the conditional expectation of $|\xi_k|^{2+\rho}$. Using condition (A1), we have

$$\mathbf{E}[|\xi_k|^{2+\rho}|\mathcal{F}_{k-1}] \leq \epsilon^\rho \, \Delta\langle X\rangle_k,$$

where $\Delta\langle X\rangle_k = \langle X\rangle_k - \langle X\rangle_{k-1}$. It is obvious that

$$\Delta \left\langle X\right\rangle_k = \Delta V_k = V_k - V_{k-1}, \quad 1 \leq k < n, \quad \Delta \left\langle X\right\rangle_n \leq \Delta V_n,$$

and that

$$\mathbf{E}[|\xi_k|^{2+\rho}|\mathcal{F}_{k-1}] \leq \epsilon^\rho \, \Delta V_k. \tag{45}$$

Combining (44) and (45) together, we obtain

$$\Big|I_1\Big| \leq J_1 := \epsilon^\rho \Big[\sum_{k=1}^n \frac{1}{A_k^{1+\rho/2}} G(T_{k-1}) \, \Delta V_k\Big]. \tag{46}$$

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

$$\tau_t = \min\{k \leq n : V_k > t\}, \quad \text{where } \min \emptyset = n. \tag{47}$$

Clearly, for any $t \in [0,T]$, the stopping time τ_t is predictable. Denote by $(\sigma_k)_{k=1,\ldots,n+1}$ the increasing sequence of moments when the increasing and 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 have

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

Set $a_t = c_*^2\gamma^2 + T - t$. Since $\Delta V_{\tau_t} \leq \epsilon^2 + 2\delta^2$ (cf. Lemma 3.2), we see that

$$t \leq V_{\tau_t} \leq V_{\tau_t-1} + \Delta V_{\tau_t} \leq t + \epsilon^2 + 2\delta^2, \quad t \in [0,T]. \tag{48}$$

Assume that $c_* \geq 4$. Then we have

$$\frac{1}{2}\, a_t \leq A_{\tau_t} = c_*^2\gamma^2 + T - V_{\tau_t} \leq a_t, \quad t \in [0,T]. \tag{49}$$

Notice that $G(z)$ is symmetric and is non-increasing in $z \geq 0$. The last bound implies that

$$J_1 \leq 2^{1+\rho/2}\, \epsilon^\rho \int_0^T \frac{1}{a_t^{1+\rho/2}}\, \mathbf{E}\Big[G\Big(\frac{u - X_{\tau_t-1}}{a_t^{1/2}}\Big)\Big] dt. \tag{50}$$

Notice also that $G(z)$ is a symmetric integrable function of bounded variation. By Lemma 3.5, it is easy to see that

$$\mathbf{E}\left[G\left(\frac{u - X_{\tau_t-1}}{a_t{}^{1/2}}\right)\right] \leq c_6 \sup_z \Big|\mathbf{P}(X_{\tau_t-1} \leq z) - \Phi(z)\Big| + c_7\sqrt{a_t}. \tag{51}$$

Since $V_{\tau_t - 1} = V_{\tau_t} - \Delta V_{\tau_t}$, $V_{\tau_t} \geq t$ (cf. (48)) and $\Delta V_{\tau_t} \leq \epsilon^2 + 2\,\delta^2$, we get

$$V_n - V_{\tau_t - 1} \leq V_n - V_{\tau_t} + \Delta V_{\tau_t} \leq 2(\epsilon^2 + \delta^2) + T - t \leq a_t. \tag{52}$$

Thus

$$\begin{aligned}
\mathbf{E}\left[(X_n - X_{\tau_t - 1})^2 | \mathcal{F}_{\tau_t - 1}\right] &= \mathbf{E}\Big[\sum_{k=\tau_t}^{n} \mathbf{E}[\xi_k^2 | \mathcal{F}_{k-1}] \Big| \mathcal{F}_{\tau_t - 1}\Big] \\
&= \mathbf{E}[\,\langle X\rangle_n - \langle X\rangle_{\tau_t - 1} | \mathcal{F}_{\tau_t - 1}] \\
&\leq \mathbf{E}[V_n - V_{\tau_t - 1} | \mathcal{F}_{\tau_t - 1}] \\
&\leq a_t.
\end{aligned}$$

Then, by Lemma 3.4, we deduce that for any $t \in [0, T]$,

$$\sup_z \Big|\mathbf{P}(X_{\tau_t - 1} \leq z) - \Phi(z)\Big| \leq c_8 \sup_z \Big|\mathbf{P}(X_n \leq z) - \Phi(z)\Big| + c_9 \sqrt{a_t}. \tag{53}$$

Combining (46), (50), (51) and (53) together, we obtain

$$\Big|I_1\Big| \leq c_{10}\,\epsilon^\rho \int_0^T \frac{dt}{a_t^{1+\rho/2}} \sup_z \Big|\mathbf{P}(X_n \leq z) - \Phi(z)\Big| + c_{11}\,\epsilon^\rho \int_0^T \frac{dt}{a_t^{(1+\rho)/2}}. \tag{54}$$

By some elementary computations, it follows that

$$\int_0^T \frac{dt}{a_t^{1+\rho/2}} \leq \int_0^T \frac{dt}{(c_*^2\gamma^2 + T - t)^{1+\rho/2}} \leq \frac{1}{c_*^\rho \gamma^\rho} \tag{55}$$

and

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

Thus

$$|I_1| \leq \frac{c_{12}}{c_*^\rho} \sup_z \Big|\mathbf{P}(X_n \leq z) - \Phi(z)\Big| + c_{\rho,1}\,\widehat{\epsilon}, \tag{56}$$

where

$$\widehat{\epsilon} = \begin{cases} \epsilon^\rho + \delta, & \text{if } \rho \in (0,1), \\ \epsilon\,|\log \epsilon| + \delta, & \text{if } \rho = 1. \end{cases}$$

b) *Control of* I_2. Note that $0 \leq \Delta V_k - \Delta\,\langle X\rangle_k \leq 2\delta^2 \mathbf{1}_{\{k=n\}}$. We have

$$|I_2| \leq \mathbf{E}\Big[\frac{1}{2A_n}\,\big|\varphi'\,(T_{n-1})\,(\Delta V_n - \Delta\,\langle X\rangle_n)\big|\Big].$$

Set $\widetilde{G}(z) = \sup_{|z-t|\leq 1} |\varphi'(t)|$. Then $|\varphi'(z)| \leq \widetilde{G}(z)$ for any real z. Note that $A_n = c_*^2\gamma^2$. Then we get the following estimation:

$$|I_2| \leq \frac{1}{c_*^2}\,\mathbf{E}[\widetilde{G}\,(T_{n-1})].$$

Notice that $\widetilde{G}(z)$ is non-increasing in $z \geq 0$, and thus it has bounded variation on $\mathbf{R}$. By Lemmas 3.2 and 3.5, we obtain

$$|I_2| \leq \frac{c_{13}}{c_*^2} \sup_z \Big| \mathbf{P}(X_n \leq z) - \Phi(z) \Big| + c_{\rho,2}\, \widehat{\epsilon}. \tag{57}$$

c) *Control of* I_3. By a two-term Taylor expansion, it follows that

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

Since $c_* \geq 4$, $\Delta A_k \leq 0$ and $|\Delta A_k| = \Delta V_k \leq \epsilon^2 + 2\,\delta^2$, we have

$$A_k \leq A_k - \vartheta_k \Delta A_k \leq c_*^2 \gamma^2 + T - V_k + \epsilon^2 + 2\,\delta^2 \leq 2A_k. \tag{58}$$

Set $\widehat{G}(z) = \sup_{|t-z|\leq 2} |\varphi'''(t)|$. Then $\widehat{G}(z)$ is symmetric, and is non-increasing in $z \geq 0$. By (58), we obtain

$$|I_3| \leq (\epsilon^2 + 2\,\delta^2) \mathbf{E}\Big[\sum_{k=1}^{n} \frac{1}{A_k^2}\, \widehat{G} \left(\frac{T_{k-1}}{\sqrt{2}} \right) \Delta V_k \Big].$$

By an argument similar to the proof of (56), we get

$$\begin{aligned} |I_3| &\leq \frac{\epsilon^2 + 2\,\delta^2}{c_* \gamma} \sup_z \Big| \mathbf{P}(X_n \leq z) - \Phi(z) \Big| + c_{\rho,3}\, \widehat{\epsilon} \\ &\leq \frac{2}{c_*} \sup_z \Big| \mathbf{P}(X_n \leq z) - \Phi(z) \Big| + c_{\rho,3}\, \widehat{\epsilon}. \end{aligned} \tag{59}$$

From (37), using (56), (57) and (59), we have

$$\Big| \mathbf{E}[\Phi_u(X_n, A_n)] - \mathbf{E}[\Phi_u(X_0, A_0)] \Big| \leq \frac{c_{14}}{c_*^{\rho}} \sup_z \Big| \mathbf{P}(X_n \leq z) - \Phi(z) \Big| + c_{\rho,4}\, \widehat{\epsilon}.$$

Implementing the last bound in (36), we deduce that

$$\sup_z \Big| \mathbf{P}(X_n \leq z) - \Phi(z) \Big| \leq \frac{c_{15}}{c_*^{\rho}} \sup_z \Big| \mathbf{P}(X_n \leq z) - \Phi(z) \Big| + c_{\rho,5}\, \widehat{\epsilon},$$

from which, choosing $c_*^{\rho} = \max\{2c_{15}, 4^{\rho}\}$, we get

$$\sup_z \Big| \mathbf{P}(X_n \leq z) - \Phi(z) \Big| \leq 2c_{\rho,5}\, \widehat{\epsilon}, \tag{60}$$

which completes the proof of theorem. □

3.2. Proof of Corollary 2.1

Define $\eta_i = \xi_i$ if $i \leq v(n)$, $\eta_i = 0$ if $i > v(n)$. Then $(\eta_i, \mathcal{F}_i)_{i=0,\dots,n}$ is also a sequence of martingale differences. It is easy to see that

$$\mathbf{E}[|\eta_i|^{2+\rho} | \mathcal{F}_{i-1}] \leq \epsilon^{\rho} \mathbf{E}[\eta_i^2 | \mathcal{F}_{i-1}].$$

If $v(n) = \sup\{k : \langle X\rangle_k \leq 1\}$, then

$$1 - \mathbf{E}[\xi^2_{v(n)+1}|\mathcal{F}_{v(n)}] \leq \sum_{i=1}^{n} \mathbf{E}[\eta_i^2|\mathcal{F}_{i-1}] = \sum_{i=1}^{v(n)} \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] \leq 1.$$

If $v(n) = \inf\{k : \langle X\rangle_k \geq 1\}$, then

$$1 \leq \sum_{i=1}^{n} \mathbf{E}[\eta_i^2|\mathcal{F}_{i-1}] = \sum_{i=1}^{v(n)} \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] \leq 1 + \mathbf{E}[\xi^2_{v(n)}|\mathcal{F}_{v(n)-1}].$$

Since $\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] \leq \epsilon^2$ for all i (cf. Lemma 3.2), we always have

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

Notice that $\sum_{i=1}^{n} \eta_i = X_{v(n)}$. Applying Theorem 2.1 to $(\eta_i, \mathcal{F}_i)_{i=0,\ldots,n}$, we obtain the desired inequalities. This completes the proof of Corollary 2.1. □

3.3. Proof of Theorem 2.2.

To prove Theorem 2.2, we use the following technical lemma of El Machkouri and Ouchti [4]; see Lemma 1 therein.

Lemma 3.3. *Let X and Y be random variables. Then for any $p \geq 1$,*

$$D(X+Y) \leq 2D(X) + 3\Big|\Big|\mathbf{E}[|Y|^{2p}|X]\Big|\Big|_1^{1/(2p+1)}. \qquad (61)$$

Following Bolthausen [1], consider the stopping time

$$\tau = \sup\{0 \leq k \leq n : \langle X\rangle_k \leq 1\}.$$

Assume that $0 < \varepsilon \leq \epsilon$. Let $r = \lfloor (1 - \langle X\rangle_\tau)/\varepsilon^2 \rfloor$, where $\lfloor x \rfloor$ stands for the largest integer less than x. Then $r \leq \lfloor 1/\varepsilon^2 \rfloor$. Let $N = n + \lfloor 1/\varepsilon^2 \rfloor + 1$. Let $(\eta_i)_{i\geq 1}$ be a sequence of independent Rademacher random variables, which is also independent of the martingale differences $(\xi_i)_{1\leq i\leq n}$. For any $i = 1, \ldots, N$, define $\xi_i' = \xi_i$ if $i \leq \tau$, $\xi_i' = \varepsilon\eta_i$ if $\tau < i \leq \tau + r$, $\xi_i' = (1 - \langle X\rangle_\tau - r\varepsilon^2)^{1/2}\eta_i$ if $i = \tau + r + 1$, and $\xi_i' = 0$ if $\tau + r + 1 < i \leq N$. Clearly, $X_k' = \sum_{i=1}^{k} \xi_i'$, $k = 0, \ldots, N$ (with $X_0' = 0$) is also a martingale sequence with respect to the enlarged probability space and the enlarged filtration. Moreover, it holds $\langle X'\rangle_N = 1$ a.s. and condition (A1) is satisfied for $(\xi_k')_{k=1,\ldots,N}$. Denote by

$$\gamma = \begin{cases} \epsilon^\rho, & \text{if } \rho \in (0,1), \\ \epsilon\,|\log \epsilon|\,, & \text{if } \rho \geq 1. \end{cases}$$

By Theorem 2.2, it holds, for all $x \in \mathbf{R}$,

$$\Big|\mathbf{P}(X_N' \leq x) - \Phi(x)\Big| \leq c_\rho\, \gamma. \qquad (62)$$

Using Lemma 3.3, we get

$$\begin{aligned} D(X_n) &\leq 2D(X_N') + 3\Big\|\mathbf{E}[|X_n - X_N'|^{2p}|X_N']\Big\|_1^{1/(2p+1)} \\ &\leq 2\,c_\rho\,\gamma + 3\Big(\mathbf{E}[|X_n - X_N'|^{2p}]\Big)^{1/(2p+1)}. \end{aligned} \tag{63}$$

As τ is a stoping time, conditionally on τ, the $(\xi_i - \xi_i')_{i \geq \tau+1}$ still forms a martingale difference sequence. Using Burkholder's inequality (cf. Theorem 2.11 of Hall and Heyde [10]), we have

$$\mathbf{E}[|X_N' - X_n|^{2p}] \leq c_p \left(\mathbf{E}\Big[\Big|\sum_{i=\tau+1}^{N} \mathbf{E}[(\xi_i - \xi_i')^2|\mathcal{F}_{i-1}]\Big|^p\Big] + \mathbf{E}[\max_{\tau+1\leq i\leq N} |\xi_i - \xi_i'|^{2p}]\right). \tag{64}$$

It is easy to see that

$$\begin{aligned} \sum_{i=\tau+1}^{N} \mathbf{E}[(\xi_i - \xi_i')^2|\mathcal{F}_{i-1}] &= \sum_{i=\tau+1}^{n} \mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] + \sum_{i=\tau+1}^{N} \mathbf{E}[\xi_i'^{\,2}|\mathcal{F}_{i-1}] \\ &= \langle X\rangle_n + 1 - 2\langle X\rangle_\tau. \end{aligned}$$

Notice that

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

Hence

$$\sum_{i=\tau+1}^{N} \mathbf{E}[(\xi_i - \xi_i')^2|\mathcal{F}_{i-1}] \leq \langle X\rangle_n - 1 + 2\mathbf{E}[\xi_{\tau+1}^2|\mathcal{F}_\tau]. \tag{65}$$

Using the inequality $|a+b|^k \leq 2^{k-1}(|a|^k + |b|^k)$, $k \geq 1$, we get

$$\begin{aligned} \mathbf{E}[\max_{\tau+1\leq i\leq N} |\xi_i - \xi_i'|^{2p}] &\leq 2^{2p-1}\Big(\mathbf{E}[\max_{\tau+1\leq i\leq n} |\xi_i|^{2p}] + \varepsilon^{2p}\Big) \\ &\leq 2^{2p-1}\Big(\mathbf{E}[\max_{1\leq i\leq n} |\xi_i|^{2p}] + \varepsilon^{2p}\Big). \end{aligned} \tag{66}$$

Combining (64), (65) and (66) together, we deduce that

$$\mathbf{E}[|X_N' - X_n|^{2p}] \leq c_p \left(\mathbf{E}[|\langle X\rangle_n - 1|^p] + \mathbf{E}[\max_{1\leq i\leq n} |\xi_i|^{2p}] + \varepsilon^{2p}\right).$$

Returning to (63) and letting $\varepsilon \to 0$, we obtain

$$D(X_n) \leq c_{p,\rho}\left(\gamma + \Big(\mathbf{E}[|\langle X\rangle_n - 1|^p] + \mathbf{E}[\max_{1\leq i\leq n} |\xi_i|^{2p}]\Big)^{1/(2p+1)}\right).$$

This completes the proof of Theorem 2.2. $\square$

Appendix

In the proof of Theorem 2.1, we make use of the following two technical lemmas due to Bolthausen (cf. Lemmas 1 and 2 of [1]).

Lemma 3.4. *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 3.5. *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*

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

where $||G||_1$ is the $L_1(\mathbf{R})$ norm of $G(x)$.

Acknowledgements

The author would like to thank editor and reviewers for their comments. This work has been partially supported by the National Natural Science Foundation of China (Grant nos. 11601375 and 11626250).

References

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

[2] J. Dedecker, F. Merlevède, Rates of convergence in the central limit theorem for linear statistics of martingale differences, Stochastic Process. Appl. 121 (2011) 1013–1043.

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

[4] M. El Machkouri, L. Ouchti, Exat convergence rates in the central limit theorem for a class of martingales, Bernoulli 13 (4) (2007) 981–999.

[5] I. Grama, On the improved rate of convergence in the CLT for semimartingales, Stokhasticeskii Analiz. Matematiceskie Issledovania 97 (1987a) pp. 34–40, Kishinev: Stiinza (in Russian).

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

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

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

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

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

[11] K. Joos, Nonuniform convergence rates in the central limit theorem for martingales, Studia Sci. Math. Hungar. 28 (1-2) (1993) 145–158.

[12] L. Ouchti, On the rate of convergence in the central limit theorem for martingale difference sequences, Ann. Inst. H. Poincaré Probab. Statist. 41 (1) (2005) 35–43.

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

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