Exact rates of convergence in some martingale central limit theorems

Xiequan Fan

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

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. An application to Lipschitz functionals of independent random variables is also given.

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

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

1. Introduction

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

$$X_0 = 0, \quad X_k = \sum_{i=1}^{k} \xi_i, \quad k = 1, \ldots, n.$$  (1)

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

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

Define

$$D(X_n) = \sup_{x \in \mathbb{R}} \left| P(X_n \leq x) - \Phi(x) \right|,$$

where $\Phi(x)$ is the distribution function of the standard normal random variable. Denote by $\xrightarrow{P}$ 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”
\[ \sum_{i=1}^{n} E[\xi_i^2 \mathbf{1}_{|\xi_i| \geq \varepsilon}] |F_{i-1}| \xrightarrow{P} 0, \quad \text{as } n \to \infty \text{ for each } \varepsilon > 0, \]
and the “conditional normalizing condition”
\[ \langle X \rangle_n \xrightarrow{P} 1, \quad \text{as } n \to \infty, \]
together implies that
\[ D(X_n) \to 0, \quad \text{as } n \to \infty. \]
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 ≤ 2, Heyde and Brown proved that
\[ D(X_n) \leq C_p \left( E[|\langle X \rangle_n - 1|^p] + \sum_{i=1}^{n} E[|\xi_i|^{2p}] \right)^{1/(2p+1)}, \tag{3} \]
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 \( C_p \) for \( p > 2 \). Using a method developed by Bolthausen [1], Haeusler [8] gave an extension of \( C_p \) to all \( p > 1 \). See also Joos [11]. Moreover, Haeusler also gave an example to show that the bound \( C_p \) is optimal under the stated condition, that is there exists a sequence of martingale differences \( (\xi_k, F_k)_{k \geq 0} \), such that for all \( n \) large enough,
\[ D(X_n) \left( E[|\langle X \rangle_n - 1|^p] + \sum_{i=1}^{n} E[|\xi_i|^{2p}] \right)^{-1/(2p+1)} \geq c_p, \]
where \( c_p \) is a positive constant and does not depend on \( n \).

However, for martingales having bounded differences, the bound \( C_p \) is not the best possible. In fact, an earlier result of Bolthausen [1] states that if \( |\xi_i| \leq \varepsilon \) and \( \langle X \rangle_n = 1 \) a.s., then
\[ D(X_n) \leq C \varepsilon^3 n \log n, \tag{4} \]
where \( C \) is a constant. Moreover, Bolthausen [4] 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 \( C \varepsilon^3 n \log n \) is optimal in the case that \( \varepsilon \) is of order \( 1/\sqrt{n} \). Relaxing the condition \( \langle X \rangle_n = 1 \) a.s., Bolthausen [1] then proved that if \( |\xi_i| \leq \varepsilon \) a.s., then
\[ D(X_n) \leq C \left( \varepsilon^3 n \log n + \min\{||\langle X \rangle_n - 1||_1^{1/3}, ||\langle X \rangle_n - 1||_\infty^{1/2}\} \right). \tag{6} \]
It seems that the item $||⟨X⟩_n - 1||_1^{1/3}$ in the last bound should be replaced by $||⟨X⟩_n - 1||_1^{1/3} + \epsilon^{2/3}$; see Mourrat [13]. (Indeed, in the proof of Corollary of Bolthausen [1], we found an item $\gamma^2$ is missing for the estimation of $\mathbf{E}[\langle \hat{S} - S \rangle^2]$.)

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

$$D(X_n) \leq C_\rho \epsilon_n,$$

where the constant $C_\rho$ depends only on $\rho$ and

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

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

$$D(X_n)\epsilon_n^{-1} \geq c,$$

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 item $\epsilon^3 n \log n$ in (6) to $\epsilon \log n$, that is if $\mathbf{E}[||\xi_i|^3|F_{i-1}] \leq \epsilon \mathbf{E}[\xi_i^2 | F_{i-1}] \text{ a.s.}$, then

$$D(X_n) \leq C\left(\varepsilon \log n + ||⟨X⟩_n - 1||_{\infty}^{1/2}\right).$$

They also proved a result with item $||⟨X⟩_n - 1||_1^{1/3}$.

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

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

where $C_p$ is a constant depending only on $p$. Notice that Mourrat [13] has extended the item $\min\{||⟨X⟩_n - 1||_1^{1/3}, ||⟨X⟩_n - 1||_{\infty}^{1/2}\}$ of Bolthausen [4] to the more general item $\mathbf{E}[||⟨X⟩_n - 1|^p|^{1/(2p+1)}] + \epsilon^{2p/(2p+1)}$. Moreover, he also has justified the optimality of the item $\mathbf{E}[||⟨X⟩_n - 1|^p|^{1/(2p+1)}]$.

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 Hauesler [5], we prove that if there exist two positive numbers $\rho$ and $\epsilon$, such that

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

then

$$D(X_n) \leq C_\rho \left(\gamma + ||⟨X⟩_n - 1||_{\infty}^{1/2}\right),$$

where $C_\rho$ is a constant depending only on $\rho$ and

$$\gamma = \begin{cases} \epsilon^p, & \text{if } \rho \in (0, 1), \\ \epsilon |\log \epsilon|, & \text{if } \rho \geq 1. \end{cases}$$
We also justify the optimality of the item $\gamma$. Then with the method of Bolthausen \cite{17}, we obtain a significant improvement of Mourrat’s inequality (10) by dropping the item $\epsilon^3 n \log n$: If $|\xi_i| \leq \epsilon$ a.s., then for any $p \geq 1$,

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

(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. The application is given in Section 3. Proofs are deferred to Section 4.

Throughout the paper, $c$ and $c_\alpha$ probably supplied with some indices, denote respectively a generic positive absolute constant and a generic positive constant depending only on $\alpha$.

2. Main Results

In the sequel we shall use the following conditions:

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

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

(A2) There exists a number $\delta \in [0, \frac{1}{2}]$, such that $|\langle X \rangle_n - 1| \leq \delta^2$ 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 $\epsilon = \frac{1}{\sqrt{n}}$ and $\delta = 0$.

2. In the case of martingales, $\epsilon$ and $\delta$ 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

$$\mathbb{E}[|Y_i|^{2+\rho}|\mathcal{F}_{i-1}] \leq C^\rho \mathbb{E}[Y_i^2|\mathcal{F}_{i-1}]$$

for a positive absolute constant $C$ and all $i \geq 1$. Let $S_n = \sum_{i=1}^n Y_i$ and $s_n = \sqrt{\mathbb{E}[S_n^2]}$. 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 $f: X^n \mapsto \mathbb{R}$ be separately Lipschitz, such that

$$|f(x_1, x_2, \ldots, x_n) - f(x_1', x_2', \ldots, x_n')| \leq d(x_1, x_1') + d(x_2, x_2') + \cdots + d(x_n, x_n').$$

(14)

Let then

$$X_n := f(\eta_1, \ldots, \eta_n) - \mathbb{E}[f(\eta_1, \ldots, \eta_n)],$$

(15)

where $\eta_1, \ldots, \eta_n$ 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 $k \in \mathbb{N}$, $\mathcal{F}_k = \sigma(\eta_1, \eta_2, \ldots, \eta_k)$. Define then

$$g_k(\eta_1, \ldots, \eta_k) = \mathbb{E}[f(\eta_1, \ldots, \eta_n)|\mathcal{F}_k],$$

(16)
Assume conditions (A1) and (A2). Theorem 2.1. where the constant $c > 0$ does not depend on $\epsilon$.

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
  \[ D(X_n) \leq c_{\rho}(\epsilon^\rho + \delta). \]

- If $\rho \in [1, \infty)$, then
  \[ D(X_n) \leq c(\epsilon|\log \epsilon| + \delta). \]

We justify the optimality of the item $\epsilon^\rho$ of (19). Let $n = [\epsilon^{-2}]$ be the integer part of $\epsilon^{-2}$ and $\rho \in (0, 1)$. Renz’s inequality (5) shows that there exists a sequence of martingale differences satisfying condition (A1) and $(X)_n = 1$ a.s., such that for all $\epsilon$ small enough,

\[ \epsilon^{-\rho}D(X_n) \geq n^{\rho/2}D(X_n) \geq c, \]

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

Notice that the bound differences satisfy condition (A1) 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 $(X)_n = 1$ a.s., such that for all $\epsilon$ small enough,

\[ (3\epsilon|\log 3\epsilon|)^{-1}D(X_n) \geq 1/4 D(X_n)\sqrt{n}/\log n \geq c, \]

where the constant $c > 0$ does not depend on $\epsilon$. Thus the item $\epsilon|\log \epsilon|$ of (20) is exact even for bounded martingale differences.
Under the conditions (A1) and (A2), the order of the item $\epsilon |\log \epsilon|$ in (20) is less than the order of the item $\epsilon^3 n \log n$ in Bolthausen’s inequality (21). Indeed, by condition (A2), we have $3/4 \leq \langle X \rangle_n \leq n \epsilon^2$ a.s. (see Lemma [12]) 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 $\epsilon$ 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 item $\epsilon |\log \epsilon|$ 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

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

instead of condition (A1), Fan, Grama and Liu [3] have obtained the Berry-Esseen bound (20). Note that the conditional Bernstein condition implies that $\xi_i$ has conditional exponential moment. Now we only assume that $\xi_i$ 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
  $$D(X_{v(n)}) \leq c_{\rho} \epsilon^\rho. \quad (23)$$

- If $\rho \in [1, \infty)$, then
  $$D(X_{v(n)}) \leq c \epsilon |\log \epsilon|. \quad (24)$$

Inequality (23) significantly improves an earlier result of Ouchti [12] under the following condition

$$E[|\xi_i|^3 | \mathcal{F}_{i-1}] \leq n^{-1/2} E[\xi_i^2 | \mathcal{F}_{i-1}] \quad \text{a.s. for all } i \geq 1.$$
where $i$ for all $\epsilon$ a.s. for all $i = 1, \ldots, n$, then $\mathbf{E}[\max_{1 \leq i \leq n} |\xi_i|^{2p}] \leq \epsilon^{2p}$, while $\sum_{i=1}^{n} \mathbf{E}[|\xi_i|^{2p}] \leq n\epsilon^{2p}$.

Note that condition (A1) implies that

$$\mathbf{E}[\max_{1 \leq i \leq n} |\xi_i|^{2p}]^{1/(2p+1)} \leq \left(\sum_{i=1}^{n} \mathbf{E}[|\xi_i|^{2p}] \right)^{1/(2p+1)} \leq \epsilon^{(2p-2)/(2p+1)} \left(\sum_{i=1}^{n} \mathbf{E}[|\xi_i|^{2p}] \right)^{1/(2p+1)};$$

see Lemma 4.1 for the last line. Thus

$$\mathbf{E}[\max_{1 \leq i \leq n} |\xi_i|^{2p}]^{1/(2p+1)} = O(\epsilon^{(2p-2)/(2p+1)}).$$

When $p \to \infty$, it follows that

$$\mathbf{E}[|\langle X \rangle_n - 1|^{p}]^{1/(2p+1)} \to ||\langle X \rangle_n - 1||_{\infty}^{1/2}, \quad \epsilon^{(2p-2)/(2p+1)} \to \epsilon, \quad \epsilon^{2p/(2p+1)} \to \epsilon.$$

Thus the bound (25) coincides with the optimal bound (19).

Similarly, the bound (26) coincides with the optimal bound (20). Notice that for any given $\alpha > 0$, it holds $\epsilon^\alpha |\log \epsilon| \to 0$ as $\epsilon \to 0$.

For martingales having bounded differences $|\xi_i| \leq \epsilon$ a.s. for all $i = 1, \ldots, n$, it is easy to see that $\mathbf{E}[\max_{1 \leq i \leq n} |\xi_i|^{2p}] \leq \epsilon^{2p}$. Thus Theorem 2.2 implies the following corollary.

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

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

It is obvious that 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 $\epsilon \geq \sqrt{1/n \log n}$, it holds $\epsilon^{2p/(2p+1)} \leq \epsilon^3 n \log n$ for any $p \geq 1$. Moreover, when $\epsilon \to 0$ and $\epsilon \geq 1/\sqrt{n}$, we have $\epsilon^3 n \log n \to \infty$ and $\epsilon^{2p/(2p+1)} \to 0$. Thus our bound (27) is significantly smaller the bound of Mourrat (10).

3. Application to Lipschitz functionals of random variables

Let $(\mathcal{X}, d_{l,i})$ be some complete separable metric spaces, where $l = 1, 2$ and $i = 1, \ldots, n$. Let $f : \mathcal{X}^n \to \mathbb{R}$ be separately Lipschitz, such that

$$d_{l,i}(x_i, x'_i) \leq |f(x_1, x_2, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x'_i, x_{i+1}, \ldots, x_n)| \leq d_{2,i}(x_i, x'_i)$$

for all $i = 1, \ldots, n$. Let

$$X_n = f(\eta_1, \ldots, \eta_n) - \mathbf{E}[f(\eta_1, \ldots, \eta_n)], \tag{28}$$

where $\eta_1, \ldots, \eta_n$ are independent random variables. Denote by $(\eta'_1, \ldots, \eta'_n)$ an independent copy of $(\eta_1, \ldots, \eta_n)$. Dedecker and Fan [2] have obtained some tight deviation inequalities on tail probabilities of $X_n$. In the following theorem, we give a Berry-Esseen bound for $X_n$. **
Theorem 3.1. Let $X_n$ be defined by (28), and denote by $\text{Var}(X_n)$ the variance of $X_n$. Assume that

$$\epsilon_n := \max_{1 \leq i \leq n} \frac{\left(\mathbb{E}[|d_{2,i}(\eta_i, \eta_i')|^{\rho}]\right)^{1/\rho}}{\sqrt{\sum_{i=1}^{n} \mathbb{E}[|d_{1,i}(\eta_i, \eta_i')|^{2}]}} \to 0$$

and

$$\delta_n := \left| \sum_{i=1}^{n} \frac{\mathbb{E}[|d_{2,i}(\eta_i, \eta_i')|^{2}]}{\sum_{i=1}^{n} \mathbb{E}[|d_{1,i}(\eta_i, \eta_i')|^{2}]} - 1 \right| \to 0, \quad n \to \infty.$$

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

$$D\left( \frac{X_n}{\sqrt{\text{Var}(X_n)}} \right) \leq c\rho \left( \epsilon_n + \delta_n \right). \quad (29)$$

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

$$D\left( \frac{X_n}{\sqrt{\text{Var}(X_n)}} \right) \leq c \left( \epsilon_n \ln \epsilon_n + \delta_n \right). \quad (30)$$

Note that if $d_{1,i}(\cdot, \cdot) = d_{2,i}(\cdot, \cdot), i = 1, \ldots, n$, then we have $\delta_n = 0$. The condition $d_{1,i}(\cdot, \cdot) = d_{2,i}(\cdot, \cdot), i = 1, \ldots, n$, is satisfied provided that $f(\eta_1, \ldots, \eta_n) = \sum_{i=1}^{n} a_i \eta_i, a_i \in \mathbb{R}$. Moreover, if $\eta_1, \ldots, \eta_n$ are i.i.d. random variables, then it holds $\epsilon_n = O(1/\sqrt{n}), n \to \infty$.

4. 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 4.1. If there exists an $s > 2$, such that

$$\mathbb{E}[|\xi_i|^s | \mathcal{F}_{i-1}] \leq \epsilon^{s-2} \mathbb{E}[\xi_i^2 | \mathcal{F}_{i-1}], \quad (31)$$

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

$$\mathbb{E}[|\xi_i|^t | \mathcal{F}_{i-1}] \leq \epsilon^{t-2} \mathbb{E}[\xi_i^2 | \mathcal{F}_{i-1}]. \quad (32)$$

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

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

Solving the last equations, we get

$$l = \frac{2(s-t)}{s-2}, \quad p = \frac{s-2}{s-t}, \quad q = \frac{s-2}{t-2}. \quad 8$$
By Hölder’s inequality and (31), it is easy to see that

\[
E[|\xi_i|^t |\mathcal{F}_{i-1}] = E[|\xi_i|^{t-l} |\mathcal{F}_{i-1}] \\
\leq \left( E[|\xi_i|^{(t-l)q} |\mathcal{F}_{i-1}] \right)^{1/q} \\
\leq \left( E[|\xi_i|^q |\mathcal{F}_{i-1}] \right)^{1/q} \\
\leq \left( E[|\xi_i|^2 |\mathcal{F}_{i-1}] \right)^{1/q} \\
\leq \epsilon^{(s-2)/q} E[|\xi_i|^s |\mathcal{F}_{i-1}] \\
= \epsilon^{t-2} E[|\xi_i|^t |\mathcal{F}_{i-1}].
\]

This completes the proof of lemma.

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

**Lemma 4.2.** If there exists an \( s > 2 \), such that

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

then

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

In particular, condition (A1) implies (34).

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

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

Thus

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

which implies (34).

### 4.1. Proof of Theorem 2.1

Theorem 2.1 is a refinement of Lemma 3.3 of Grama and Haeusler where it is assumed that \( \xi_i \)'s are bounded, which is a particular case of condition (A1). See also Lemma 3.1 of Fan, Grama and Liu. Compared to the proofs of Grama and Haeusler and Fan, Grama and Liu, the main challenge of our proof comes from the control of \( I_1 \) defined in (41).

By Lemma 4.1, 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:

\[
V_k = \langle X \rangle_k \mathbf{1}_{\{k<n\}} + T \mathbf{1}_{\{k=n\}}.
\]

It is obvious that \( V_0 = 0 \), \( V_n = T \), and that \( (V_k, \mathcal{F}_k)_{k=0,...,n} \) 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 $c_\ast$ be an absolute constant, whose exact value will be chosen later. Define the following non-increasing discrete time predictable process

$$A_k = c_\ast^2 \gamma^2 + T - V_k, \quad k = 1, \ldots, n.$$ 

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

$$\Phi_u(x, y) = \Phi\left(\frac{u - x}{\sqrt{y}}\right).$$

Let $N = N(0, 1)$ be a standard normal random variable, which is independent of $X_n$. Using a smoothing procedure, by Lemma 4.4 we get

$$\sup_u \left| P(X_n \leq u) - \Phi(u) \right| \leq c_1 \sup_u \left| P(X_n + c_\ast \gamma N \leq u) - \Phi(u) \right| + c_2 \gamma$$

$$= c_1 \sup_u \left| E[\Phi_u(X_n, A_n)] - \Phi(u) \right| + c_2 \gamma$$

$$\leq c_1 \sup_u \left| E[\Phi_u(X_n, A_n)] - E[\Phi_u(X_0, A_0)] \right|$$

$$+ c_1 \sup_u \left| E[\Phi_u(X_0, A_0)] - \Phi(u) \right| + c_2 \gamma$$

$$= c_1 \sup_u \left| E[\Phi_u(X_n, A_n)] - E[\Phi_u(X_0, A_0)] \right|$$

$$+ c_1 \sup_u \left| \Phi\left(\frac{u}{\sqrt{c_\ast^2 \gamma^2 + T}}\right) - \Phi(u) \right| + c_2 \gamma. \quad (37)$$

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

$$\left| \Phi\left(\frac{u}{\sqrt{c_\ast^2 \gamma^2 + T}}\right) - \Phi(u) \right| \leq c_3 \left| \frac{1}{\sqrt{c_\ast^2 \gamma^2 + T}} - 1 \right| \leq c_4 \gamma. \quad (38)$$

Returning to (37), we obtain

$$\sup_u \left| P(X_n \leq u) - \Phi(u) \right| \leq c_1 \sup_u \left| E[\Phi_u(X_n, A_n)] - E[\Phi_u(X_0, A_0)] \right| + c_5 \gamma. \quad (39)$$

By a simple telescoping, we deduce that

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

Using the fact

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

we obtain

$$E[\Phi_u(X_n, A_n)] - E[\Phi_u(X_0, A_0)] = I_1 + I_2 - I_3, \quad (40)$$

10
where

\[ I_1 = \mathbb{E} \left[ \sum_{k=1}^{n} \left( \Phi_u(X_k, A_k) - \Phi_u(X_{k-1}, A_k) - \frac{\partial}{\partial x} \Phi_u(X_{k-1}, A_k) \xi_k - \frac{1}{2} \frac{\partial^2}{\partial x^2} \Phi_u(X_{k-1}, A_k) \xi_k^2 \right) \right], \]  \hspace{1cm} (41)

\[ I_2 = \frac{1}{2} \mathbb{E} \left[ \sum_{k=1}^{n} \frac{\partial^2}{\partial x^2} \Phi_u(X_{k-1}, A_k) \left( \Delta \langle X \rangle_k - \Delta V_k \right) \right], \]  \hspace{1cm} (42)

\[ I_3 = \mathbb{E} \left[ \sum_{k=1}^{n} \left( \Phi_u(X_{k-1}, A_{k-1}) - \Phi_u(X_{k-1}, A_k) - \frac{\partial}{\partial y} \Phi_u(X_{k-1}, A_k) \Delta V_k \right) \right], \]  \hspace{1cm} (43)

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 \( \varphi \) the density function of the standard normal random variable. Moreover, \( \vartheta_i \)'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

\[ R_k =: \Phi_u(X_k, A_k) - \Phi_u(X_{k-1}, A_k) - \frac{\partial}{\partial x} \Phi_u(X_{k-1}, A_k) \xi_k - \frac{1}{2} \frac{\partial^2}{\partial x^2} \Phi_u(X_{k-1}, A_k) \xi_k^2 \]

\[ = \Phi \left( T_{k-1} + \frac{\xi_k}{\sqrt{A_k}} \right) - \Phi(T_{k-1}) - \Phi'(T_{k-1}) \frac{\xi_k}{\sqrt{A_k}} - \frac{1}{2} \Phi''(T_{k-1}) \left( \frac{\xi_k}{\sqrt{A_k}} \right)^2. \]

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

\[ |R_k| = \left| \frac{1}{6} \Phi''' \left( T_{k-1} + \vartheta \frac{\xi_k}{\sqrt{A_k}} \right) \left( \frac{\xi_k}{\sqrt{A_k}} \right)^3 \right| \]

\[ \leq \left| \Phi''' \left( T_{k-1} + \vartheta \frac{\xi_k}{\sqrt{A_k}} \right) \right| \left( \frac{\xi_k}{\sqrt{A_k}} \right)^{2+\vartheta}. \]

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

\[ |R_k| \leq \frac{1}{2} \left( |\Phi'' \left( T_{k-1} + \vartheta \frac{\xi_k}{\sqrt{A_k}} \right)| + |\Phi''(T_{k-1})| \right) \left( \frac{\xi_k}{\sqrt{A_k}} \right)^2 \]

\[ \leq \left| \Phi'' \left( T_{k-1} + \vartheta \frac{\xi_k}{\sqrt{A_k}} \right) \right| \left( \frac{\xi_k}{\sqrt{A_k}} \right)^2 \]

\[ \leq \left| \Phi'' \left( T_{k-1} + \vartheta \frac{\xi_k}{\sqrt{A_k}} \right) \right| \left( \frac{\xi_k}{\sqrt{A_k}} \right)^{2+\vartheta}, \]

where

\[ \vartheta = \begin{cases} \vartheta, & \text{if } \left| \Phi'' \left( T_{k-1} + \vartheta \frac{\xi_k}{\sqrt{A_k}} \right) \right| \geq \left| \Phi''(T_{k-1}) \right|, \\
0, & \text{if } \left| \Phi'' \left( T_{k-1} + \vartheta \frac{\xi_k}{\sqrt{A_k}} \right) \right| < \left| \Phi''(T_{k-1}) \right|. \end{cases} \]
By the inequality \( \max \{|\Phi''(t)|, |\Phi'''(t)|\} \leq \varphi(t)(1 + t^2) \), it follows that

\[
\left| R_k 1_{\{\xi_k/\sqrt{A_k} \leq 1 + |T_{k-1}|/2\}} \right| \leq \varphi \left( T_{k-1} + \vartheta_1 \frac{\xi_k}{\sqrt{A_k}} \right) \left( 1 + \left( T_{k-1} + \vartheta_1 \frac{\xi_k}{\sqrt{A_k}} \right)^2 \right) 
\leq g_1(T_{k-1}),
\]

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

\[
\left| R_k 1_{\{\xi_k/\sqrt{A_k} \leq 1 + |T_{k-1}|/2\}} \right| \leq g_1(T_{k-1}) \left| \frac{\xi_k}{\sqrt{A_k}} \right|^{2+\rho} 1_{\{\xi_k/\sqrt{A_k} \leq 1 + |T_{k-1}|/2\}}. \tag{44}
\]

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

\[
\left| \Phi(x + \Delta x) - \Phi(x) - \Phi'(x)\Delta x - \frac{1}{2} \Phi''(x)(\Delta x)^2 \right| \\
= \left( \left| \frac{\Phi(x + \Delta x) - \Phi(x)}{|\Delta x|^{2+\rho}} \right| + |\Phi'(x)| + |\Phi''(x)| \right) |\Delta x|^{2+\rho} \\
\leq \left( 4 \left| \frac{\Phi(x + \Delta x) - \Phi(x)}{(2 + |x|)^2} \right| + |\Phi'(x)| + |\Phi''(x)| \right) |\Delta x|^{2+\rho} \\
\leq \left( \frac{c_1}{(2 + |x|)^2} + |\Phi'(x)| + |\Phi''(x)| \right) |\Delta x|^{2+\rho} \\
\leq \frac{c_2}{(2 + |x|)^2} |\Delta x|^{2+\rho}.
\]

Therefore,

\[
\left| R_k 1_{\{|\xi_k/\sqrt{A_k}| > 1 + |T_{k-1}|/2\}} \right| \leq g_2(T_{k-1}) \left| \frac{\xi_k}{\sqrt{A_k}} \right|^{2+\rho} 1_{\{|\xi_k/\sqrt{A_k}| > 1 + |T_{k-1}|/2\}}, \tag{45}
\]

where

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

Set

\[
G(z) = g_1(z) + g_2(z).
\]

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

\[
\left| R_k \right| \leq G(T_{k-1}) \left| \frac{\xi_k}{\sqrt{A_k}} \right|^{2+\rho}, \tag{46}
\]

and thus

\[
\left| I_1 \right| = \left| E \left[ \sum_{k=1}^{n} R_k \right] \right| \leq E \left[ \sum_{k=1}^{n} G(T_{k-1}) \left| \frac{\xi_k}{\sqrt{A_k}} \right|^{2+\rho} \right]. \tag{47}
\]
Now we consider the conditional expectation of $|\xi_k|^{2+\rho}$. Using condition (A1), we have

$$E[|\xi_k|^{2+\rho}|F_{k-1}] \leq e^\rho \Delta \langle X \rangle_k,$$

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

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

and that

$$E[|\xi_k|^{2+\rho}|F_{k-1}] \leq e^\rho \Delta V_k. \quad (48)$$

Combining (47) and (48) together, we obtain

$$\left| I_1 \right| \leq J_1 := e^\rho \left[ \sum_{k=1}^n \frac{1}{A_k^{1+\rho/2}} G(T_{k-1}) \Delta V_k \right]. \quad (49)$$

To estimate $J_1$, we introduce the time change $\tau_t$ as follows: for any real $t \in [0,T]$,\n
$$\tau_t = \min \{ k \leq n : V_k > t \}, \quad \text{where } \min \emptyset = n. \quad (50)$$

It is clear that, for any $t \in [0,T]$, the stopping time $\tau_t$ is predictable. Let $(\sigma_k)_{k=1,\ldots,n+1}$ be the increasing sequence of moments when the increasing stepwise function $\tau_t, t \in [0,T], \text{ has jumps.}$ It is clear 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

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

Set $a_t = c_s^2 \gamma^2 + T - t$. Since $\Delta V_{\tau_t} \leq \epsilon^2 + 2\delta^2$ (cf. Lemma 12), we see that

$$t \leq V_{\tau_t} \leq V_{\tau_t-1} + \Delta V_{\tau_t} \leq t + \epsilon^2 + 2\delta^2, \quad t \in [0,T]. \quad (51)$$

Assume that $c_s \geq 4$. Then we have

$$\frac{1}{2} a_t \leq A_{\tau_t} = c_s^2 \gamma^2 + T - V_{\tau_t} \leq a_t, \quad t \in [0,T]. \quad (52)$$

Since $G(z)$ is symmetric and is non-increasing in $z \geq 0$, the last bound implies that

$$J_1 \leq 2^{1+\rho/2} e^\rho \int_0^T \frac{1}{a_t^{1+\rho/2}} E \left[ G \left( \frac{u - X_{\tau_t-1}}{a_t^{1/2}} \right) \right] dt. \quad (53)$$

It is easy to see that $G(z)$ is a symmetric integrable function of bounded variation. By Lemma 1.5 it is easy to see that

$$E \left[ G \left( \frac{u - X_{\tau_t-1}}{a_t^{1/2}} \right) \right] \leq c_6 \sup_z \left| \Phi(X_{\tau_t-1} - z) - \Phi(z) \right| + c_7 \sqrt{a_t}. \quad (54)$$
Thus

\[ V_n - V_{\tau - 1} \leq V_n - V_{\tau} + \Delta V_{\tau} \leq 2(\epsilon^2 + \delta^2) + T - t \leq a_t. \] (55)

Thus

\[
\mathbb{E} \left[ (X_n - X_{\tau-1})^2 | \mathcal{F}_{\tau-1} \right] = \mathbb{E} \left[ \sum_{k=\tau}^{n} \mathbb{E} [\xi_k^2 | \mathcal{F}_k] | \mathcal{F}_{\tau-1} \right]
\]

\[
= \mathbb{E} [\langle X \rangle_n - \langle X \rangle_{\tau-1} | \mathcal{F}_{\tau-1}]
\]

\[
\leq \mathbb{E} [V_n - V_{\tau-1} | \mathcal{F}_{\tau-1}]
\]

\[
\leq a_t.
\]

Then, by Lemma 4.4, we deduce that for any \( t \in [0, T] \),

\[
\sup_z \left| P(X_{\tau-1} \leq z) - \Phi(z) \right| \leq c_8 \sup_z \left| P(X_n \leq z) - \Phi(z) \right| + c_9 \sqrt{a_t}.
\] (56)

Combining (49), (53), (54) and (56) together, we obtain

\[
|I_1| \leq c_{10} \epsilon^\rho \int_0^T \frac{dt}{a_t^{1+\rho/2}} \sup_z \left| P(X_n \leq z) - \Phi(z) \right| + c_{11} \epsilon^\rho \int_0^T \frac{dt}{a_t^{(1+\rho)/2}}.
\] (57)

By some elementary computations, it follows that

\[
\int_0^T \frac{dt}{a_t^{1+\rho/2}} \leq \int_0^T \frac{dt}{(c_7^2 \gamma^2 + T - t)^{1+\rho/2}} \leq \frac{1}{c_7^2 \gamma^\rho}
\] (58)

and

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

Thus

\[
|I_1| \leq \frac{c_{12}}{c_7^2} \sup_z \left| P(X_n \leq z) - \Phi(z) \right| + c_{\rho, 1} \tilde{\epsilon},
\] (59)

where

\[
\tilde{\epsilon} = \left\{ \begin{array}{ll}
 \epsilon^\rho + \delta, & \text{if } \rho \in (0, 1), \\
 \epsilon |\log \epsilon| + \delta, & \text{if } \rho = 1.
\end{array} \right.
\]

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

\[
|I_2| \leq \mathbb{E} \left[ \frac{1}{2A_n} |\varphi' (T_{n-1}) (\Delta V_n - \Delta \langle X \rangle_n)| \right].
\]

Set \( \tilde{G}(z) = \sup_{|z-t| \leq 1} |\varphi'(t)| \). Then \( |\varphi'(z)| \leq \tilde{G}(z) \) for any real \( z \). Note that \( A_n = c_7^2 \gamma^2 \). Then we get the following estimation:

\[
|I_2| \leq \frac{1}{c_7^2} \mathbb{E} [\tilde{G}(T_{n-1})].
\]
Notice that $\tilde{G}(z)$ is non-increasing in $z \geq 0$, and thus it has bounded variation on $\mathbb{R}$. By Lemmas 4.2 and 4.3, we obtain
\[
|I_2| \leq \frac{c_{13}}{c_s^2} \sup_z \left| \mathbb{P}(X_n \leq z) - \Phi(z) \right| + c_{\rho,2} \hat{c}.
\] (60)

\textbf{c) Control of } I_3. \text{ By a two-term Taylor expansion, it follows that} 
\[
I_3 = \frac{1}{8} \mathbb{E} \left[ \sum_{k=1}^{n} \frac{1}{(A_k - \vartheta_k \Delta A_k)^2} \varphi'' \left( \frac{u - X_{k-1}}{\sqrt{A_k - \vartheta_k \Delta A_k}} \right) \Delta A_k^2 \right].
\]
Since $c_s \geq 4$, $\Delta A_k \leq 0$ and $|\Delta A_k| = \Delta V_k \leq \epsilon^2 + 2 \delta^2$, we have
\[
A_k \leq A_k - \vartheta_k \Delta A_k \leq \epsilon^2 \gamma^2 + T - V_k + \epsilon^2 + 2 \delta^2 \leq 2A_k.
\] (61)
Set $\tilde{G}(z) = \sup_{|t-z| \leq 2} |\varphi''(t)|$. Then $\tilde{G}(z)$ is symmetric, and is non-increasing in $z \geq 0$. By (61), we obtain
\[
|I_3| \leq \epsilon^2 + 2 \delta^2 \mathbb{E} \left[ \sum_{k=1}^{n} \frac{1}{A_k^2} \tilde{G} \left( \frac{T_{k-1}}{\sqrt{2}} \right) \Delta V_k \right].
\]
By an argument similar to the proof of (59), we get
\[
|I_3| \leq \frac{\epsilon^2 + 2 \delta^2}{c_s \gamma} \sup_z \left| \mathbb{P}(X_n \leq z) - \Phi(z) \right| + c_{\rho,3} \hat{c}
\]
\[
\leq \frac{2}{c_s} \sup_z \left| \mathbb{P}(X_n \leq z) - \Phi(z) \right| + c_{\rho,3} \hat{c}.
\] (62)

From (60), using (59), (60) and (62), we have
\[
\left| \mathbb{E} [\Phi_u(X_n, A_n)] - \mathbb{E} [\Phi_u(X_0, A_0)] \right| \leq \frac{c_{14}}{c_s^2} \sup_z \left| \mathbb{P}(X_n \leq z) - \Phi(z) \right| + c_{\rho,4} \hat{c}.
\]
Implementing the last bound in (59), we deduce that
\[
\sup_z \left| \mathbb{P}(X_n \leq z) - \Phi(z) \right| \leq \frac{c_{15}}{c_s^2} \sup_z \left| \mathbb{P}(X_n \leq z) - \Phi(z) \right| + c_{\rho,5} \hat{c},
\]
from which, choosing $c_s^p = \max\{2c_{15}, 4^\rho\}$, we get
\[
\sup_z \left| \mathbb{P}(X_n \leq z) - \Phi(z) \right| \leq 2c_{\rho,5} \hat{c},
\] (63)
which completes the proof of theorem. \hfill \Box

4.2. \textbf{Proof of Corollary 2.1}

Define $\eta_i = \xi_i$ if $i \leq v(n)$, $\eta_i = 0$ if $i > v(n)$. Then $(\eta_i, \mathcal{F}_i)_{i=0, \ldots, n}$ is also a sequence of martingale differences. It is easy to see that
\[
\mathbb{E} [\eta_i^{2+\rho} | \mathcal{F}_{i-1}] \leq \epsilon^\rho \mathbb{E} [\eta_i^2 | \mathcal{F}_{i-1}],
\]
If \( v(n) = \sup \{ k : \langle X \rangle_k \leq 1 \} \), then

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

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

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

Since \( \mathbf{E}[\xi_i^2 | \mathcal{F}_{i-1}] \leq \epsilon^2 \) for all \( i \) (cf. Lemma 4.2), we always have

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

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. \( \square \)

4.3. 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 4.3.** Let \( X \) and \( Y \) be random variables. Then for any \( p \geq 1 \),

\[
D(X + Y) \leq 2D(X) + 3 \left\| \mathbf{E}[|Y|^{2p} | X] \right\|_1^{1/(2p+1)}.
\]  

(64)

Following Bolthausen [1], consider the stopping time

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

Let \( r = \lfloor (1 - \langle X \rangle_\tau)/\epsilon^2 \rfloor \), where \( [x] \) is the largest integer less than \( x \). Then \( r \leq \lfloor 1/\epsilon^2 \rfloor \). Let \( N = n + \lfloor 1/\epsilon^2 \rfloor + 1 \). Consider a sequence of independent Rademacher random variables \( (\eta_i) \) (taking values +1 and −1 with equal probabilities) which is also independent of the martingale differences \( (\xi_i) \). For each \( i = 1, \ldots, N \) define \( \xi_i = \xi \) if \( i \leq \tau \), \( \xi_i = \epsilon \eta_i \) if \( \tau < i \leq \tau + r \), \( \xi_i = (1 - \langle X \rangle_\tau - r \epsilon^2)^{1/2} \eta_i \) 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 a martingale sequence w.r.t. the enlarged probability space and the enlarged filtration. Moreover \( \langle X' \rangle_N = 1 \) a.s. and condition (A1) is satisfied for \( (\xi'_k)_{k=1,\ldots,N} \). Denote by

\[
\gamma = \left\{ \begin{array}{ll}
\epsilon^\rho, & \text{if } \rho \in (0,1), \\
\epsilon |\log \epsilon|, & \text{if } \rho \geq 1.
\end{array} \right.
\]

By Theorem 2.2 it holds, for all \( x \in \mathbb{R} \),

\[
\left| \mathbf{P}(X'_N \leq x) - \Phi(x) \right| \leq c_\rho \gamma. \tag{65}
\]
Using Lemma 4.3 we get
\[
D(X_n) \leq 2D(X'_N) + 3 \left| \mathbb{E} \left[ \left| X_n - X'_N \right|^{2p} \right] \right|^{1/(2p+1)}_1 \\
\leq 2c_p + 3 \left( \mathbb{E} \left[ \left| X_n - X'_N \right|^{2p} \right] \right)^{1/(2p+1)}.
\]
(66)

As \( \tau \) is a stopping time, conditionally on \( \tau \), 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
\[
\mathbb{E} \left[ X'_N - X_n \right]^{2p} \leq c_p \left( \mathbb{E} \left[ \sum_{i=\tau+1}^N (\xi_i - \xi'_i)^2 \right] \right)^p + c_p \max_{\tau+1 \leq i \leq N} (\xi_i - \xi'_i)^2.
\]
(67)

It is easy to see that
\[
\sum_{i=\tau+1}^N \mathbb{E} \left[ (\xi_i - \xi'_i)^2 \right] = \sum_{i=\tau+1}^n \mathbb{E} \left[ (\xi_{\tau+1} - \xi_i)^2 \right] + \sum_{i=\tau+1}^N \mathbb{E} \left[ (\xi_i - \xi'_i)^2 \right] = (X)_n - 2\langle X \rangle_{\tau}.
\]

Since \( \mathbb{E} \left[ (\xi_i - \xi'_i)^2 \right] \leq \epsilon^2 \) for all \( i \) (cf. Lemma 4.2), it holds
\[
1 - \epsilon^2 \leq 1 - \mathbb{E} \left[ (\xi_{\tau+1} - \xi_{\tau})^2 \right] \leq \langle X \rangle_{\tau} \leq 1.
\]

Hence
\[
\sum_{i=\tau+1}^N \mathbb{E} \left[ (\xi_i - \xi'_i)^2 \right] \leq (X)_n - 1 + 2\epsilon^2.
\]
(68)

Using the inequality \( |a + b|^k \leq 2^{k-1}(|a|^k + |b|^k), k \geq 1 \), we get
\[
\mathbb{E} \left[ \max_{\tau+1 \leq i \leq N} \left| \xi_i - \xi'_i \right|^{2p} \right] \leq 2^{2p-1} \left( \mathbb{E} \left[ \max_{\tau+1 \leq i \leq n} \left| \xi_i \right|^{2p} \right] + \epsilon^2 \right)
\]
(69)

Combining (67), (68) and (69) together, we deduce that
\[
\mathbb{E} \left[ X'_N - X_n \right]^{2p} \leq c_p \left( \mathbb{E} \left[ (X)_n - 1 \right]^p \right) + c_p \max_{1 \leq i \leq n} \left| \xi_i \right|^{2p} + \epsilon^2.
\]

Returning to (66), we obtain
\[
D(X_n) \leq c_p \left( \gamma + \left( \mathbb{E} \left[ (X)_n - 1 \right]^p \right) + c_p \max_{1 \leq i \leq n} \left| \xi_i \right|^{2p} + \epsilon^2 \right)^{1/(2p+1)}.
\]
which gives (25). Moreover, for \( \rho \geq 1 \), it holds that \( \gamma \leq c_p \epsilon^{2p/(2p+1)} \) for \( \epsilon \in (0, 1) \). Thus for \( \rho \geq 1 \),
\[
D(X_n) \leq c_p \left( \mathbb{E} \left[ (X)_n - 1 \right]^p \right) + c_p \left( \max_{1 \leq i \leq n} \left| \xi_i \right|^{2p} + \epsilon^2 \right)^{1/(2p+1)}.
\]
which gives (26). This completes the proof of Theorem 2.2.
\[\square\]
4.4. Proof of Theorem 3.1

Let \( X_n = \xi_1 + \xi_2 + \cdots + \xi_n \) be Doob’s martingale decomposition of \( X_n \) (cf. (17)). Then it is easy to see that

\[
\langle X \rangle_n = \sum_{i=1}^{n} \mathbb{E}[\mathbb{E}[f(\eta_1, \ldots, \eta_n) | \mathcal{F}_i] - \mathbb{E}[f(\eta_1, \ldots, \eta_n) | \mathcal{F}_{i-1}]^2 | \mathcal{F}_{i-1}]
\]

\[
= \sum_{i=1}^{n} \mathbb{E}[\mathbb{E}[f(\eta_1, \ldots, \eta_n) | \mathcal{F}_i] - \mathbb{E}[f(\eta_1, \ldots, \eta'_1, \ldots, \eta_n) | \mathcal{F}_i]^2 | \mathcal{F}_{i-1}]
\]

\[
\geq \sum_{i=1}^{n} \mathbb{E}[(\mathbb{E}[d_{1,i}(\eta_i, \eta'_i) | \mathcal{F}_i)]^2 | \mathcal{F}_{i-1}]
\]

\[
= \sum_{i=1}^{n} \mathbb{E}[(\mathbb{E}[d_{1,i}(\eta_i, \eta'_i)] | \mathcal{F}_i)^2].
\]

Similarly, we have

\[
\langle X \rangle_n \leq \sum_{i=1}^{n} \mathbb{E}[(\mathbb{E}[d_{2,i}(\eta_i, \eta'_i)] | \mathcal{F}_i)^2],
\]

and then

\[
\sum_{i=1}^{n} \mathbb{E}[(\mathbb{E}[d_{1,i}(\eta_i, \eta'_i)] | \mathcal{F}_i)^2] \leq \text{Var}(X_n) \leq \sum_{i=1}^{n} \mathbb{E}[(\mathbb{E}[d_{2,i}(\eta_i, \eta'_i)] | \mathcal{F}_i)^2].
\]

Therefore,

\[
\left| \frac{\langle X \rangle_n}{\text{Var}(X_n)} - 1 \right| \leq \left| \frac{\sum_{i=1}^{n} \mathbb{E}[(\mathbb{E}[d_{2,i}(\eta_i, \eta'_i)] | \mathcal{F}_i)^2]}{\sum_{i=1}^{n} \mathbb{E}[(\mathbb{E}[d_{1,i}(\eta_i, \eta'_i)] | \mathcal{F}_i)^2]} - 1 \right| + \left| \frac{\sum_{i=1}^{n} \mathbb{E}[(\mathbb{E}[d_{2,i}(\eta_i, \eta'_i)] | \mathcal{F}_i)^2]}{\sum_{i=1}^{n} \mathbb{E}[(\mathbb{E}[d_{2,i}(\eta_i, \eta'_i)] | \mathcal{F}_i)^2]} - 1 \right|.
\]

Since the two items in the right hand side of the last inequality have the same order as \( \delta_n \to 0 \), we get

\[
\left| \frac{\langle X \rangle_n}{\text{Var}(X_n)} - 1 \right| \leq c \delta_n.
\]

By (18), we have

\[
\mathbb{E}\left[ \frac{\xi_i}{\sqrt{\text{Var}(X_n)}} \right]^{2+\rho} \mathbb{E}[\mathcal{F}_{i-1}] \leq \frac{\mathbb{E}[(\mathbb{E}[d_{2,i}(\eta_i, \eta'_i)] | \mathcal{F}_i)^{\rho}]}{\rho \sqrt{\text{Var}(X_n)}} \mathbb{E}\left[ \frac{\xi_i^2}{\sqrt{\text{Var}(X_n)}} \right] \mathbb{E}[\mathcal{F}_{i-1}]
\]

\[
\leq \epsilon_n^2 \mathbb{E}\left[ \frac{\xi_i^2}{\sqrt{\text{Var}(X_n)}} \right] \mathbb{E}[\mathcal{F}_{i-1}].
\]

Applying Theorem 2.1 to \((\xi_i/\sqrt{\text{Var}(X_n)}, \mathcal{F}_i)_{i=1,\ldots,n}\), we obtain the required inequalities.
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 4.4.** Let $X$ and $Y$ be random variables. Then

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

**Lemma 4.5.** Let $G(x)$ be an integrable function on $\mathbb{R}$ of bounded variation $||G||_V$. $X$ be a random variable and $a, b \neq 0$ are real numbers. Then

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

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

References

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

[2] Dedecker J., Fan X. (2015). Deviation inequalities for separately Lipschitz functionals of iterated random functions. *Stochastic Process. Appl.* 125(1): 60–90.

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

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

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

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

[7] Grama, I. and Haehsler, E. (2000). Large deviations for martingales via Cramer’s method. *Stochastic Process. Appl.* 85, 279–293.

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

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

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

[11] Joos, K. (1993). Nonuniform convergence rates in the central limit theorem for martingales. *Studia Sci. Math. Hungar.* 28(1-2), 145–158.
[12] Ouchti, L. (2005). On the rate of convergence in the central limit theorem for martingale difference sequences. *Ann. Inst. H. Poincaré Probab. Statist.* **41**(1): 35–43.

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

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