On Wasserstein-1 distance in the central limit theorem for elephant random walk

Xiaohui Ma[a],   Mohamed El Machkouri[b],   Xiequan Fan[a,*]

[a]*Center for Applied Mathematics, Tianjin University, 300072 Tianjin, China*
[b]*Laboratoire de Mathématiques Raphaël Salem, UMR CNRS 6085, Université de Rouen, France.*

Abstract

Recently, the elephant random walk has attracted a lot of attentions. A wide range of literature is available for the asymptotic behavior of the process, such as the central limit theorems, functional limit theorems and the law of iterated logarithm. However, there is not result concerning Wasserstein-1 distance for the normal approximations. In this paper, we show that the Wasserstein-1 distance in the central limit theorem is totally different when a memory parameter p belongs to one of the three cases $0 < p < 1/2$, $1/2 < p < 3/4$ and $p = 3/4$.

Keywords: Elephant random walk; Central limit theorem; The Wasserstein distance

2000 MSC: Primary 60G42; 60F05; 60G50

1. Introduction and main results

Random walks, are widely used in theoretical physics to describe phenomena of traveling motion and mass transport. As a kind of interesting random walk, the elephant random walk was introduced by Schütz and Trimper [13] in 2004 in order to study the memory effects in the non-Markovian processes. The name being inspired by the fact that elephants can remember where they have been. It is a one-dimensional discrete-time random walk on integers, which has a complete memory of its whole history. The elephant random walk model can be described as follows. It starts at time zero on 0. At time $n = 1$, the elephant η_1 has a Rademacher $\mathcal{R}(q)$ distribution, where q lies between $[0, 1]$. For $n \geq 1$, we define

$$\eta_{n+1} = \alpha_n \eta_{\beta_n},$$

where α_n has a Rademacher $\mathcal{R}(p)$ distribution, p lies between $[0, 1]$ and β_n is uniformly distributed over $\{1, 2, \ldots, n\}$. Moreover, α_n is independent of η_{β_n}. The position at time $n + 1$ is

$$S_{n+1} = \sum_{i=1}^{n+1} \eta_i.$$

In particular, when $p = 0$ or $1/2$, the elephant random walk reduces to the classical symmetric random walk. Since the latter random walk has been well studied, we skip the cases $p = 0$ or $1/2$.

---

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

The elephant random walk model has drawn a lot of attention in the past few years. One of the interesting questions concerns the influence of the memory effect on the long time behavior. Quite recently, via a method of connection to Pólya-type urns, Baur and Bertoin [1] derived the functional limit theorems and showed that the limiting process of elephant random walk turns out to be Gaussian in the diffusive regime ($0 < p < 3/4$) and the critical regime ($p = 3/4$). But, in the superdiffusive regime ($3/4 < p \leqslant 1$), the limiting process of elephant random walk is non-Gaussian. See also Coletti, Gava and Schütz [5, 6] for central limit theorem (CLT) and strong invariance principle of elephant random walk. With a martingale approach, Bercu [2] recovered the CLT for the elephant random walk as $0 \leq p \leq 3/4$. For $3/4 < p \leq 1$, he also showed that the elephant random walk converges to a non-normal random variable. The strong law of large numbers and the law of iterated logarithm are also discussed in [2]. For the extensions of elephant random walk, we refer to Bercu and Laulin [3], where Bercu and Laulin have established the CLT for the multi-dimensional elephant random walk. Though the asymptotic behavior of the elephant random walk, such as CLT, functional limit theorems and the law of iterated logarithm, has been well studied, there is no result concerning Wasserstein-1 distance for the normal approximations. Set $a_1 = 1$ and denote, for $n \geq 2$,

$$a_n = \frac{\Gamma(n)\Gamma(2p)}{\Gamma(n+2p-1)} \quad \text{and} \quad v_n = \sum_{i=1}^{n} a_i^2,$$

where $\Gamma(x)$ is the Gamma function. Notice that the exact values of a_n and v_n can be easily calculated via computer. In this paper, our main interest is Wasserstein-1 distance for CLT of elephant random walk. Recall that the Wasserstein-1 distance for CLT is defined as follows:

$$\|F_{X_n} - \Phi\|_1 = \int_{-\infty}^{+\infty} |F_{X_n}(x) - \Phi(x)| \mathrm{d}x,$$

where F_{X_n} and Φ are the distribution function of the random variable X_n and the standard normal random variable, respectively. The main result of this paper is the following theorem which gives some Wasserstein-1 distance results for the CLT of elephant random walk. In particular, we get three different convergence rates depending on the value of the memory parameter p.

Theorem 1.1. *Let $n \geq 2$. The following inequalities hold.*

[i] *If $0 < p < 1/2$, then*

$$\|F_{a_n S_n/\sqrt{v_n}} - \Phi\|_1 \leq C \frac{\log n}{\sqrt{n}}.$$

[ii] *If $1/2 < p < 3/4$, then*

$$\|F_{a_n S_n/\sqrt{v_n}} - \Phi\|_1 \leq C_p \frac{\log n}{\sqrt{n^{3-4p}}}.$$

[iii] *If $p = 3/4$, then*

$$\|F_{a_n S_n/\sqrt{v_n}} - \Phi\|_1 \leq C \frac{\log \log n}{\sqrt{\log n}}.$$

Here, C is an absolute constant and C_p depends only on p.

The main idea in the proof of Theorem 1.1 is to introduce a martingale associated to the elephant random walk and to take advantage of powerfull limit theorems from martingale theory. In particular, we are going to apply the CLT established in Lemma 2.4 below.

2. Proof of Theorem 1.1

Define the filtration $\mathcal{F}_n = \sigma(\eta_i,\ 1 \le i \le n)$ and

$$M_n = a_n S_n - 2q + 1. \tag{1}$$

Then $(M_n, \mathcal{F}_n)_{n \ge 1}$ is a martingale. Indeed, it is easy to see that for

$$
\begin{aligned}
\mathbf{E}[M_{n+1}|\mathcal{F}_n] &= \mathbf{E}[a_{n+1}(S_n + \alpha_n \eta_{\beta_n}) - 2q + 1|\mathcal{F}_n] \\
&= a_{n+1} S_n + a_{n+1}\mathbf{E}[\alpha_n]\mathbf{E}[\eta_{\beta_n}|\mathcal{F}_n] - 2q + 1.
\end{aligned}
$$

Taking into account that

$$\mathbf{E}[\alpha_n] = p + (-1)(1 - p) \quad \text{and} \quad \mathbf{E}[\eta_{\beta_n}|\mathcal{F}_n] = \sum_{i=1}^{n} \frac{1}{n}\eta_i, \tag{2}$$

we deduce that

$$
\begin{aligned}
\mathbf{E}[M_{n+1}|\mathcal{F}_n] &= a_{n+1}\left(S_n + (p + (-1)(1 - p))\sum_{i=1}^{n} \frac{1}{n}\eta_i\right) - 2q + 1 \\
&= a_{n+1}\left(S_n + (2p - 1)\frac{S_n}{n}\right) - 2q + 1 \\
&= a_n S_n - 2q + 1 \\
&= M_n. \tag{3}
\end{aligned}
$$

Let $(\Delta M_n)_{n \ge 1}$ be the martingale differences defined by $\Delta M_1 = M_1$ and, for $n \ge 2$,

$$\Delta M_n = M_n - M_{n-1}.$$

Denote $\langle M \rangle_n$ be the quadratic variation of M_n, that is $\langle M \rangle_n = \sum_{i=1}^{n} \mathbf{E}[\Delta M_i^2|\mathcal{F}_{i-1}]$. The normalized martingale $(M_n/\sqrt{v_n}, \mathcal{F}_n)_{n \ge 1}$ satisfies the following property.

Lemma 2.1. *For $n \ge 2$, we have*

$$\left\|\frac{\Delta M_i}{\sqrt{v_n}}\right\|_{\infty} \le \frac{2}{\sqrt{v_n}}a_i$$

and for any $t \ge 1$, there exists a positive constant C_t such that

$$\left\|\frac{\langle M \rangle_n}{v_n} - 1\right\|_t \le \begin{cases} \dfrac{C_t}{3 - 4p}n^{-1}, & \text{if } 0 < p < 3/4, \\[2ex] C_t\big(\log n\big)^{-1}, & \text{if } p = 3/4. \end{cases}$$

Proof. Clearly, from Lemma 3.1 of Fan et al. [9], we have

$$\|\Delta M_i/\sqrt{v_n}\|_{\infty} \le 2a_i/\sqrt{v_n},$$

3

which gives the first desired inequality. Next we give an estimation of $\|\langle M \rangle_n / v_n - 1\|_t, 1 \leq t < \infty$, for $0 < p \leq 3/4$. From (1), we get $\Delta M_k = a_k \varepsilon_k = a_k(S_k - \gamma_{k-1}S_{k-1})$. Thus, it holds

$$
\begin{aligned}
\mathbf{E}[(\Delta M_k)^2|\mathcal{F}_{k-1}] &= a_k^2 \mathbf{E}[(S_k - \gamma_{k-1}S_{k-1})^2|\mathcal{F}_{k-1}] \\
&= a_k^2 \big(\mathbf{E}[S_k^2|\mathcal{F}_{k-1}] - 2\gamma_{k-1}S_{k-1}\mathbf{E}[S_k|\mathcal{F}_{k-1}] + \gamma_{k-1}^2 S_{k-1}^2\big).
\end{aligned}
$$

By (2) and $(\alpha_k X_{\beta_k})^2 = 1$, it is easy to see that

$$
\begin{aligned}
\mathbf{E}[S_k^2|\mathcal{F}_{k-1}] &= \mathbf{E}[(S_{k-1} + \alpha_k X_{\beta_k})^2|\mathcal{F}_{k-1}] \\
&= S_{k-1}^2 + 2S_{k-1}\mathbf{E}[\alpha_k X_{\beta_k}|\mathcal{F}_{k-1}] + 1 \\
&= S_{k-1}^2 + 2\frac{2p-1}{k-1}S_{k-1}^2 + 1 \\
&= (2\gamma_{k-1} - 1)S_{k-1}^2 + 1
\end{aligned}
$$

and that

$$
\mathbf{E}[S_k|\mathcal{F}_{k-1}] = \mathbf{E}[S_{k-1} + \alpha_k X_{\beta_k}|\mathcal{F}_{k-1}] = S_{k-1} + \frac{2p-1}{k-1}S_{k-1} = \gamma_{k-1}S_{k-1}.
$$

Thus, we have $\mathbf{E}[(\Delta M_1)^2] = 1 = a_1^2$ and for $k \geq 2$,

$$
\begin{aligned}
\mathbf{E}[(\Delta M_k)^2|\mathcal{F}_{k-1}] &= a_k^2\big((2\gamma_{k-1} - 1)S_{k-1}^2 + 1 - 2\gamma_{k-1}^2 S_{k-1}^2 + \gamma_{k-1}^2 S_{k-1}^2\big) \\
&= a_k^2\big(1 - (\gamma_{k-1} - 1)^2 S_{k-1}^2\big) \\
&= a_k^2 - (2p-1)^2 a_k^2 \big(\frac{S_{k-1}}{k-1}\big)^2.
\end{aligned}
$$

Hence, by the definition of v_n and M_k, we obtain

$$
\langle M \rangle_n = v_n - (2p-1)^2 \bigg(\sum_{k=1}^{n-1} \Big(\frac{a_{k+1}}{a_k}\Big)^2 \Big(\frac{M_k}{k}\Big)^2 \bigg).
$$

Since $\frac{a_{n+1}}{a_n} \sim 1$ as $n \to \infty$ (cf. equality (7)), by Minkowski's inequality, we have for $t \geq 1$,

$$
\|\langle M \rangle_n - v_n\|_t \leq C(2p-1)^2 \bigg\| \sum_{k=1}^{n-1} (\frac{M_k}{k})^2 \bigg\|_t \leq C \sum_{k=1}^{n-1} \frac{1}{k^2} \|M_k\|_{2t}^2, \tag{4}
$$

where C is a positive constant which values may change from line to line. Using Rio's inequality (cf. Theorem 2.1 of [12]) and $\|\Delta M_i\|_\infty \leq 2a_i$, we derive that for $t \geq 1$,

$$
\|M_k\|_{2t}^2 \leq (2t-1) \sum_{i=1}^k \|\Delta M_i\|_{2t}^2 \leq (2t-1)4v_k.
$$

In the diffusive regime $0 < p < 3/4$, by (8), we get for $t \geq 1$,

$$
\|\langle M \rangle_n - v_n\|_t \leq 4C(2t-1) \sum_{k=1}^{n-1} \frac{1}{k^2} v_k \leq C(2t-1)\frac{\Gamma(2p)^2}{3-4p} \sum_{k=1}^{n-1} k^{1-4p} \leq C\frac{2t-1}{3-4p} n^{2-4p}. \tag{5}
$$

4

In the critical regime $p = 3/4$, by (9), we have for $t \geq 1$,

$$\|\langle M \rangle_n - v_n\|_t \leq C(2t-1) \sum_{k=1}^{n-1} \frac{\log k}{k^2} \leq C(2t-1). \tag{6}$$

Consequently, again by (8) and (9), we obtain the desired inequality. This completes the proof of lemma 2.1. $\qquad\square$

For simplicity of notation, denote by

$$\xi_i = \frac{\Delta M_i}{\sqrt{v_n}}, \quad 1 \leq i \leq n, \quad X_n = \sum_{i=1}^{n} \xi_i \quad \text{and} \quad \langle X \rangle_n = \sum_{i=1}^{n} \mathbf{E}[\xi_i^2 | \mathcal{F}_{i-1}].$$

From the last lemma, we have:

Lemma 2.2. *For $n \geq 2$, we have*

$$\|\xi_i\|_\infty \leq \begin{cases} C\, n^{-1/2}, & \text{if } 0 < p < 1/2, \\ C_p\, n^{-(3-4p)/2}, & \text{if } 1/2 < p < 3/4, \\ C(\log n)^{-1/2}, & \text{if } p = 3/4, \end{cases}$$

and for any $t \geq 1$, there exists a positive constant C_t such that

$$\left\| \langle X \rangle_n - 1 \right\|_t \leq \begin{cases} \dfrac{C_t}{3-4p} n^{-1}, & \text{if } 0 < p < 3/4, \\ \\ C_t\big(\log n\big)^{-1}, & \text{if } p = 3/4, \end{cases}$$

where C is an absolute constant and C_p depends only on p.

Proof. We only need to prove the first inequality. The second one holds obviously. By Stirling's formula

$$\log \Gamma(x) = \left(x - \frac{1}{2}\right)\log x - x + \frac{1}{2}\log 2\pi + O\left(\frac{1}{x}\right) \quad \text{as } x \to \infty,$$

we deduce that

$$\lim_{n \to \infty} a_n n^{2p-1} = \Gamma(2p). \tag{7}$$

Moreover, in the diffusive regime $(0 \leq p < 3/4)$, we have

$$\lim_{n \to \infty} \frac{v_n}{n^{3-4p}} = \frac{\Gamma(2p)^2}{3-4p}, \tag{8}$$

and, in the critical regime $(p = 3/4)$, it holds

$$\lim_{n \to \infty} \frac{v_n}{\log n} = \frac{3}{4}. \tag{9}$$

5

See also Bercu [2] for the equalities (7)-(9). Hence, by Lemma 2.1 and the equalities (7)-(9), we have

$$\|\xi_i\|_\infty \leq \begin{cases} C\, n^{-1/2}, & \text{if } 0 < p < 1/2, \\ C_p\, n^{-(3-4p)/2}, & \text{if } 1/2 < p < 3/4, \\ C(\log n)^{-1/2}, & \text{if } p = 3/4, \end{cases}$$

which gives the desired inequality. $\qquad\square$

In the sequel we shall use the following conditions:

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

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

(A2) There exists a number $\delta_n \in [0, \frac{1}{2}]$, such that

$$\|\langle X\rangle_n - 1\|_\infty \leq \delta_n^2 \quad \text{ for all } i = 1, 2, \dots, n.$$

With conditions (A1) and (A2), we have the following lemma for the mean martingale CLT, which is of independent interest. The proof of the lemma is given in the next section. Notice that the convergence rate of the Wasserstein-1 distance in the following lemma is same to the best possible Berry-Esseen bound for martingales, see Fan [8] and El Machkouri and Ouchti [7].

Lemma 2.3. *Assume the conditions (A1) and (A2). Then*

$$\|F_{X_n} - \Phi\|_1 \leq C_\rho(\hat{\varepsilon}_n + \delta_n), \tag{10}$$

where

$$\hat{\varepsilon}_n = \begin{cases} \varepsilon_n^\rho & \text{if } \rho \in (0, 1), \\ \varepsilon_n|\log \varepsilon_n| & \text{if } \rho \geq 1. \end{cases} \tag{11}$$

The condition (A2) is very restrictive, while we have the following lemma when condition (A2) is dropped. The proof of this lemma is present in Section 4.

Lemma 2.4. *Assume condition (A1). Then for any $t \geq 1$,*

$$\|F_{X_n} - \Phi\|_1 \leq C_{t,\rho}\Big(\hat{\varepsilon}_n + \Big(\mathbf{E}|\langle X\rangle_n - 1|^t + \mathbf{E}\max_{1\leq i\leq n}|\xi_i|^{2t}\Big)^{1/2t}\Big), \tag{12}$$

where $\hat{\varepsilon}_n$ is defined by (11).

Now, we are in position to prove Theorem 1.1. Recall that $X_n = (a_n S_n - 2q + 1)/\sqrt{v_n}$. By Lemma 2.2, it is easy to see that

$$\mathbf{E}|\langle X\rangle_n - 1| \leq \begin{cases} \dfrac{C}{3-4p} n^{-1}, & \text{if } 0 < p < 3/4, \\[2mm] C\big(\log n\big)^{-1}, & \text{if } p = 3/4. \end{cases}$$

6

Applying Lemmas 2.2 and 2.4 to X_n with $t = 1$ and

$$\varepsilon_n = \begin{cases} C\ n^{-1/2}, & \text{if } 0 < p < 1/2, \\ C_p\ n^{-(3-4p)/2}, & \text{if } 1/2 < p < 3/4, \\ C(\log n)^{-1/2}, & \text{if } p = 3/4, \end{cases}$$

we obtain the desired inequalities of Theorem 1.1. Note that

$$\|F_{a_n S_n / \sqrt{v_n}} - \Phi\|_1 \leq \|F_{(a_n S_n - 2q+1)/\sqrt{v_n}} - \Phi\|_1 + 1/\sqrt{v_n}.$$

This completes the proof of Theorem 1.1.

$\square$

3. Proof of Lemma 2.3

In the sequel, constants C and C_p are always numerical constants that may change between appearances. In the proof of Lemma 2.3, we shall use the following technical lemma of Van Dung et al. [14].

Lemma 3.1. *Let X and Y be two random variables. Then for $r > 1/2$,*

$$\|F_X - \Phi\|_1 \leq \|F_{X+Y} - \Phi\|_1 + 2(2r+1)\|\mathbf{E}[Y^{2r}|X]\|_\infty^{1/2r}.$$

The following lemma can be found in Fan [8].

Lemma 3.2. *Under condition (A1), ξ_i has a bounded conditional variance, that is,*

$$\mathbf{E}[\xi_i^2|\mathcal{F}_{i-1}] \leq \varepsilon_n^2 \quad a.s. \tag{13}$$

for all $i = 1, 2, \ldots, n$.

In the proof of Lemma 2.3, we also need the following lemma.

Lemma 3.3. *Let G be a function $\mathbb{R} \to \mathbb{R}$, which has derivative G' which together with G belong to $L^1(\mathbb{R})$. For a random variable X and a constant $a > 0$, it holds*

$$\mathbf{E}\left[\int_{-\infty}^{+\infty} G\left(\frac{X+t}{a}\right)dt\right] \leq \|G'\|_1\|F_X - \Phi\|_1 + a\|G\|_1.$$

Proof. Let N be a standard normally distributed random variable. By the definition of expectation, it is easy to see that

$$\left|\mathbf{E}\left[\int_{-\infty}^{+\infty} G\left(\frac{X+t}{a}\right)dt\right] - \mathbf{E}\left[\int_{-\infty}^{+\infty} G\left(\frac{N+t}{a}\right)dt\right]\right|$$
$$= \frac{1}{a}\left|\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(F_X(x) - \Phi(x)\right)G'\left(\frac{x+t}{a}\right)dtdx\right|. \tag{14}$$

7

Using integration by substitution with $\begin{cases} u = \frac{x+t}{a} \\ v = x \end{cases}$, we deduce that the r.h.s. of equality (14) equals to $\left| \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (F_X(v) - \Phi(v))G'(u)\mathrm{d}u\mathrm{d}v \right|$. Hence, from (14), we have

$$\left| \mathbf{E}\left[\int_{-\infty}^{+\infty} G\left(\frac{X+t}{a}\right)\mathrm{d}t \right] - \mathbf{E}\left[\int_{-\infty}^{+\infty} G\left(\frac{N+t}{a}\right)\mathrm{d}t \right] \right| \le \|G'\|_1 \|F_X - \Phi\|_1. \tag{15}$$

On the other hand, it holds

$$\mathbf{E}\left[\int_{-\infty}^{+\infty} G\left(\frac{N+t}{a}\right)\mathrm{d}t \right] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} G\left(\frac{x+t}{a}\right)\Phi'(x)\mathrm{d}t\mathrm{d}x \le a\|G\|_1. \tag{16}$$

Combining (15) and (16) together, we obtain the desired inequality. $\qquad\square$

Now, we are in position to prove Lemma 2.3. We follow the argument of Grama and Haeusler [10], where Grama and Haeusler proved a Berry-Esseen's bound for CLT instead of Wasserstein-1 distance for CLT. For simplicity of notations, denote

$$\gamma_n = \begin{cases} \varepsilon_n + \delta_n & \text{if } \rho \in (0,1), \\ \varepsilon_n|\log\varepsilon_n| + \delta_n & \text{if } \rho \ge 1. \end{cases} \tag{17}$$

Set $T = 1 + \delta_n$, and introduce a modification of the conditional variance $\langle X \rangle$ as follows:

$$V_k = \langle X \rangle_k \mathbf{1}_{\{k<n\}} + T\mathbf{1}_{\{k=n\}}. \tag{18}$$

It is obvious that $V_k = \langle X \rangle_k$, $1 \le k < n$, $V_n = T \ge \langle X \rangle_n$. Let c be a large constant depending on ρ and satisfies $c \ge 2$, whose value will be chosen later. We introduce a discrete time predictable process

$$A_k = c^2\gamma_n^2 + T - V_k, \qquad k = 1,\ldots,n.$$

Then A_k is $\mathcal{F}_{k-1}$-measurable and non-increasing, which is easily to verify. Moreover, it is obvious that $A_0 = c^2\gamma_n^2 + T$, $A_n = c^2\gamma_n^2$. For $y > 0$ and $u, x \in \mathbb{R}$, denote

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

Let N be a standard normal random variable, which is independent of X_n. By Lemma 3.1 we deduce that

$$\|F_{X_n} - \Phi\|_1 \le \|F_{X_n + c\gamma_n N} - \Phi\|_1 + C_1\gamma_n$$
$$= \int_{-\infty}^{+\infty} \left| \mathbf{E}\Phi_u(X_n, A_n) - \Phi(u) \right| \mathrm{d}u + C_1\gamma_n.$$

Applying triangle inequality to the first term in the last line, we have

$$\|F_{X_n} - \Phi\|_1 \le \int_{-\infty}^{+\infty} \left| \mathbf{E}\Phi_u(X_n, A_n) - \mathbf{E}\Phi_u(X_0, A_0) \right| \mathrm{d}u + \int_{-\infty}^{+\infty} \left| \mathbf{E}\Phi_u(X_0, A_0) - \Phi(u) \right| \mathrm{d}u + C_1\gamma_n.$$

For the second term in the r.h.s. of the last inequality, by the definition of A_0, it is easy to see that

$$\int_{-\infty}^{+\infty} \Big|\mathbf{E}\Phi_u(X_0, A_0) - \Phi(u)\Big|\mathrm{d}u = \int_{-\infty}^{+\infty} \Big|\Phi(\frac{u}{\sqrt{c^2\gamma_n^2 + T}}) - \Phi(u)\Big|\mathrm{d}u \le C_2\gamma_n, \tag{19}$$

then it follows that

$$\|F_{X_n} - \Phi\|_1 \le \int_{-\infty}^{+\infty} \Big|\mathbf{E}\Phi_u(X_n, A_n) - \mathbf{E}\Phi_u(X_0, A_0)\Big|\mathrm{d}u + C_3\gamma_n. \tag{20}$$

Next, we give an estimation for $\int_{-\infty}^{\infty} |\mathbf{E}\Phi_u(X_n, A_n) - \mathbf{E}\Phi_u(X_0, A_0)|\mathrm{d}u$. By a simple telescope, we derive 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 deduce that

$$\mathbf{E}\Phi_u(X_n, A_n) - \mathbf{E}\Phi_u(X_0, A_0) = I_1 - I_2 + I_3, \tag{21}$$

where

$$I_1 = \mathbf{E}\Big[\sum_{k=1}^{n} \Phi_u(X_k, A_k) - \Phi_u(X_{k-1}, A_k) - \frac{\partial}{\partial x}\Phi_u(X_{k-1}, A_k)\xi_k - \frac{1}{2}\frac{\partial^2}{\partial x^2}\Phi_u(X_{k-1}, A_k)\xi_k^2\Big], \tag{22}$$

$$I_2 = \mathbf{E}\Big[\sum_{k=1}^{n} \Phi_u(X_{k-1}, A_{k-1}) - \Phi_u(X_{k-1}, A_k) - \frac{\partial}{\partial y}\Phi_u(X_{k-1}, A_k)\Delta V_k\Big], \tag{23}$$

$$I_3 = \frac{1}{2}\mathbf{E}\Big[\sum_{k=1}^{n} \frac{\partial^2}{\partial x^2}\Phi_u(X_{k-1}, A_k)(\Delta\langle X\rangle_k - \Delta V_k)\Big], \tag{24}$$

with $\Delta\langle X\rangle_k = \langle X\rangle_k - \langle X\rangle_{k-1}$ and $\Delta V_k = V_k - V_{k-1}, k = 1, 2, \ldots, n$. Notice that, as $\frac{\partial^2}{\partial x^2}\Phi_u(X_{k-1}, A_k)$ is $\mathcal{F}_{k-1}$-measurable, it holds

$$\mathbf{E}\Big[\frac{\partial^2}{\partial x^2}\Phi_u(X_{k-1}, A_k)\xi_k^2\Big] = \mathbf{E}\Big[\frac{\partial^2}{\partial x^2}\Phi_u(X_{k-1}, A_k)\Delta\langle X\rangle_k\Big].$$

Next, we give estimates for $\int_{-\infty}^{\infty} |I_1|\mathrm{d}u$, $\int_{-\infty}^{\infty} |I_2|\mathrm{d}u$ and $\int_{-\infty}^{\infty} |I_3|\mathrm{d}u$, respectively. Denote by φ the density function of the standard normal random variable. Moreover, θ_k's stand for random variables satisfying $0 \le \theta_k \le 1$, which may take different values in different places.

(a) Estimate of $\int_{-\infty}^{\infty} |I_1|\mathrm{d}u$.

For brevity of notation, denote by $T_{k-1} = \frac{u - X_{k-1}}{\sqrt{A_k}}$. Clearly, it holds

$$I_1 = \sum_{k=1}^{n} \mathbf{E}R_k,$$

9

where

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

$$= \Phi(T_{k-1} - \frac{\xi_k}{\sqrt{A_k}}) - \Phi(T_{k-1}) - \Phi'(T_{k-1})(-\frac{\xi_k}{\sqrt{A_k}}) - \frac{1}{2}\Phi''(T_{k-1})(\frac{\xi_k}{\sqrt{A_k}})^2. \qquad (25)$$

To bound up the bound (25), we distinguish two cases as follows.

(1) Case of $|\frac{\xi_k}{\sqrt{A_k}}| \le 1 + \frac{|T_{k-1}|}{2}$.

When $|\frac{\xi_k}{\sqrt{A_k}}| \le 1$, by a three term Taylor's expansion, we have

$$|R_k| = \left|\frac{1}{3!}\Phi'''(T_{k-1} - \theta_k\frac{\xi_k}{\sqrt{A_k}})(\frac{\xi_k}{\sqrt{A_k}})^3\right|$$

$$\le \left|\Phi'''(T_{k-1} - \theta_k\frac{\xi_k}{\sqrt{A_k}})\right|\left|\frac{\xi_k}{\sqrt{A_k}}\right|^{2+\rho}.$$

When $|\frac{\xi_k}{\sqrt{A_k}}| > 1$, by a two term Taylor's expansion, we get

$$|R_k| \le \frac{1}{2}\left(|\Phi''(T_{k-1} - \widehat{\theta}_k\frac{\xi_k}{\sqrt{A_k}})| + |\Phi''(T_{k-1})|\right)(\frac{\xi_k}{\sqrt{A_k}})^2$$

$$\le \left|\Phi''(T_{k-1} - \theta'_k\frac{\xi_k}{\sqrt{A_k}})\right|\left|\frac{\xi_k}{\sqrt{A_k}}\right|^{2+\rho},$$

where

$$\theta'_k = \begin{cases} \widehat{\theta}_k, & \text{if } |\Phi''(T_{k-1} - \widehat{\theta}_k\frac{\xi_k}{\sqrt{A_k}})| \ge |\Phi''(T_{k-1})|, \\ 0, & \text{otherwise.} \end{cases}$$

Recall that $|\frac{\xi_k}{\sqrt{A_k}}| \le 1 + \frac{|T_{k-1}|}{2}$ and $0 \le \theta_k, \theta'_k \le 1$. By the inequality $\max\{|\Phi''(t)|, |\Phi'''(t)|\} \le \varphi(t)(1 + t^2)$, it follows that

$$|R_k| \le \varphi(T_{k-1} - \theta_1\frac{\xi_k}{\sqrt{A_k}})\left(1 + (T_{k-1} - \theta_1\frac{\xi_k}{\sqrt{A_k}})^2\right)\left|\frac{\xi_k}{\sqrt{A_k}}\right|^{2+\rho}$$

$$\le g_1(T_{k-1})\left|\frac{\xi_k}{\sqrt{A_k}}\right|^{2+\rho}, \qquad (26)$$

where

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

(2) Case of $|\frac{\xi_k}{\sqrt{A_k}}| > 1 + \frac{|T_{k-1}|}{2}$.

10

It is easy to see that for $|\Delta x| > 1 + \frac{|x|}{2}$, it holds

$$\left|\Phi(x + \Delta x) - \Phi(x) - \Phi'(x) - \frac{1}{2}\Phi''(x)(\Delta x)^2\right| \leq \left(\frac{|\Phi(x + \Delta x) - \Phi(x)|}{|\Delta x|^{2+\rho}} + |\Phi'(x)| + |\Phi''(x)|\right)|\Delta x|^{2+\rho}$$

$$\leq \left(8\frac{|\Phi(x + \Delta x) - \Phi(x)|}{(2 + |x|)^{2+\rho}} + |\Phi'(x)| + |\Phi''(x)|\right)|\Delta x|^{2+\rho}$$

$$\leq \left(\frac{C}{(4 + x^2)^{1+\rho/2}} + |\Phi'(x)| + |\Phi''(x)|\right)|\Delta x|^{2+\rho}$$

$$\leq \frac{C_4}{(4 + x^2)^{1+\rho/2}}|\Delta x|^{2+\rho}. \tag{27}$$

Then in this case we have

$$|R_k| \leq g_2(T_{k-1})|\frac{\xi_k}{\sqrt{A_k}}|^{2+\rho}, \tag{28}$$

where

$$g_2(z) = \frac{C_4}{(4 + z^2)^{1+\rho/2}}.$$

Now, we return to the estimation of bound (25). Set

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

then it is easy to verify that $G \in L^1$, $G' \in L^1$ and $G(z), z > 0$, is non-increasing in z. By (26) and (28), it holds

$$|R_k| \leq G(T_{k-1})\left|\frac{\xi_k}{\sqrt{A_k}}\right|^{2+\rho}. \tag{29}$$

By (29) and condition (A1), that is $\mathbf{E}[|\xi_k|^{2+\rho}|\mathcal{F}_{k-1}] \leq \varepsilon_n^\rho \Delta\langle X\rangle_k$, we derive that

$$|I_1| \leq \sum_{k=1}^{n} \mathbf{E}|R_k| \leq \mathbf{E}\left[\sum_{k=1}^{n} G(T_{k-1})\frac{1}{A_k^{1+\rho/2}}\mathbf{E}[|\xi_k|^{2+\rho}|\mathcal{F}_{k-1}]\right]$$

$$\leq \varepsilon_n^\rho \mathbf{E}\left[\sum_{k=1}^{n} G(T_{k-1})\frac{1}{A_k^{1+\rho/2}}\Delta\langle X\rangle_k\right]$$

$$\leq \varepsilon_n^\rho \mathbf{E}\left[\sum_{k=1}^{n} G(T_{k-1})\frac{1}{A_k^{1+\rho/2}}\Delta V_k\right]. \tag{30}$$

To bound up (30), we introduce the stopping time τ_t as follows: for any $t \in [0, T]$,

$$\tau_t = \min\{k \leq n : \langle X\rangle_k > t\}, \qquad \text{where } \min \emptyset = n. \tag{31}$$

Denote $(\sigma_k)_{k=1,\ldots,n}$ be the increasing sequence of moments while the stopping time has jumps, then $\Delta V_k = \int_{[\sigma_k, \sigma_{k+1})} \mathrm{d}t$ and $k = \tau_t$ for $t \in [\sigma_k, \sigma_{k+1})$. This follows from the definition of τ_t. Therefore, we have

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

$$= \int_0^T \frac{1}{A_{\tau_t}^{1+\rho/2}}G(T_{\tau_t-1})\mathrm{d}t. \tag{32}$$

11

Set $a_t = c^2\gamma_n^2 + T - t$. By Lemma 3.2, we have

$$\Delta V_n = T - V_{n-1} = 1 + \delta_n^2 - \langle X \rangle_{n-1} = 1 + \delta_n^2 - \langle X \rangle_n + \Delta\langle X \rangle_n \le 2\delta_n^2 + \varepsilon_n^2.$$

Recall that $A_k = c^2\gamma_n^2 + T - V_k$, then we deduce that $A_{\tau_t} \le a_t$ and

$$A_{\tau_t} \ge c^2\gamma_n^2 + T - t - \Delta V_{\tau_t} \ge c^2\gamma_n^2 + T - t - (2\delta_n^2 + \varepsilon_n^2) \ge \frac{1}{2}a_t.$$

Thus, by (30) and (32) and the fact $\frac{1}{2}a_t \le A_{\tau_t} \le a_t$, it holds

$$
\begin{aligned}
|I_1| &\le \varepsilon_n^\rho \mathbf{E}\left[\int_0^T \frac{1}{A_{\tau_t}^{1+\rho/2}} G(T_{\tau_t-1}) \mathrm{d}t \right] \\
&\le 2^{1+\rho/2}\varepsilon_n^\rho \int_0^T \frac{1}{a_t^{1+\rho/2}} \mathbf{E}\left[G\left(\frac{u - X_{\tau_t-1}}{\sqrt{a_t}} \right) \right] \mathrm{d}t.
\end{aligned}
\tag{33}
$$

Applying Fubini's theorem to the last inequality, we have

$$\int_{-\infty}^{+\infty} |I_1| \mathrm{d}u \le 2^{1+\rho/2}\varepsilon_n^\rho \int_0^T \frac{1}{a_t^{1+\rho/2}} \mathbf{E}\left[\int_{-\infty}^{+\infty} G\left(\frac{u - X_{\tau_t-1}}{\sqrt{a_t}} \right) \mathrm{d}u \right] \mathrm{d}t. \tag{34}$$

By Lemma 3.3, we get

$$\mathbf{E}\left[\int_{-\infty}^{+\infty} G\left(\frac{u - X_{\tau_t-1}}{\sqrt{a_t}} \right) \mathrm{d}u \right] \le C_5 \|F_{X_{\tau_t-1}} - \Phi\|_1 + C_6\sqrt{a_t}. \tag{35}$$

By Lemma 3.1, we have

$$\|F_{X_{\tau_t-1}} - \Phi\|_1 \le \|F_{X_n} - \Phi\|_1 + C_7\|\mathbf{E}[(X_n - X_{\tau_t-1})^2 | X_{\tau_t-1}]\|_\infty^{1/2}. \tag{36}$$

Now we consider $\mathbf{E}[(X_n - X_{\tau_t-1})^2 | \mathcal{F}_{\tau_t-1}]$. It is easy to see that

$$
\begin{aligned}
\mathbf{E}[(X_n - X_{\tau_t-1})^2 | \mathcal{F}_{\tau_t-1}] &= \mathbf{E}\left[\sum_{k=\tau_t}^n \xi_k^2 \Big| \mathcal{F}_{\tau_t-1} \right] = \mathbf{E}\left[\sum_{k=\tau_t}^n \mathbf{E}[\xi_k^2 | \mathcal{F}_{k-1}] \Big| \mathcal{F}_{\tau_t-1} \right] \\
&= \mathbf{E}[\langle X \rangle_n - \langle X \rangle_{\tau_t-1} | \mathcal{F}_{\tau_t-1}] \\
&\le V_n - V_{\tau_t-1} \\
&\le a_t,
\end{aligned}
\tag{37}
$$

where the last line follows by the fact

$$V_n - V_{\tau_t-1} \le V_n - V_{\tau_t} + \Delta V_{\tau_t} \le T - t + \varepsilon_n^2 + 2\delta_n^2 \le a_t.$$

Combining (34)-(37) together, we obtain

$$\int_{-\infty}^{+\infty} |I_1| \mathrm{d}u \le C_8\, \varepsilon_n^\rho \int_0^T \frac{\mathrm{d}t}{a_t^{1+\rho/2}} \|F_{X_n} - \Phi\|_1 + C_9\, \varepsilon_n^\rho \int_0^T \frac{\mathrm{d}t}{a_t^{(1+\rho)/2}}. \tag{38}$$

After some simple calculations we deduce that

$$\int_0^T \frac{dt}{a_t^{1+\rho/2}} \le \frac{2}{\rho c^\rho \gamma_n^\rho}$$

and

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

Since $\varepsilon_n \le \gamma_n$, finally we get

$$\int_{-\infty}^{+\infty} |I_1| du \le \frac{C_{10}}{\rho c^\rho} \|F_{X_n} - \Phi\|_1 + C_{\rho,1} \hat{\varepsilon}_n, \tag{39}$$

where $\hat{\varepsilon}_n$ is defined in (11) and γ_n is defined in (17).

(b) Estimate of $\int_{-\infty}^{+\infty} |I_2| du$.

By a two-term Taylor expansion, we get

$$I_2 = \frac{1}{2} \mathbf{E} \Big[\sum_{k=1}^n \frac{\partial^2}{\partial y^2} \Phi_u(X_{k-1}, A_k - \theta_k \Delta A_k) \Delta A_k^2 \Big].$$

By the fact that $\frac{\partial^4}{\partial x^4} \Phi_u(x,y) = 4 \frac{\partial^2}{\partial y^2} \Phi_u(x,y)$, we deduce that

$$I_2 = \frac{1}{8} \mathbf{E} \Big[\sum_{k=1}^n \frac{\partial^4}{\partial x^4} \Phi_u(X_{k-1}, A_k - \theta_k \Delta A_k) \Delta A_k^2 \Big]$$

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

Since $|\Delta A_k| := A_k - A_{k-1} = \Delta V_k \le \varepsilon_n^2 + 2\delta_n^2$, we have

$$A_k \le A_k - \theta_k \Delta A_k \le c^2 \gamma_n^2 + T - V_k + \varepsilon_n^2 + 2\delta_n^2 \le 2A_k.$$

Set $\bar{G}(z) = \sup_{|t-z| \le 2} |\varphi'''(t)|$, then $\bar{G}(xz)$ is decreasing in $x > 0$. It follows that

$$|I_2| \le \mathbf{E} \Big[\sum_{k=1}^n \frac{1}{A_k^2} \bar{G} \Big(\frac{u - X_{k-1}}{\sqrt{2A_k}} \Big) \Delta A_k^2 \Big]$$

$$\le (\varepsilon_n^2 + 2\delta_n^2) \mathbf{E} \Big[\sum_{k=1}^n \frac{1}{A_k^2} \bar{G} \Big(\frac{T_{k-1}}{\sqrt{2}} \Big) \Delta V_k \Big],$$

where the last line follows by the facts that $|\Delta A_k| \le \varepsilon_n^2 + 2\delta_n^2$ and $\Delta A_k = \Delta V_k$. It is easy to verify that $\bar{G}, \bar{G}' \in L^1$ and $|\varphi'''(z)| \le \bar{G}(z)$ for any $z \ge 0$. By an argument similar to the proof of (39), we deduce that

$$\int_{-\infty}^{+\infty} |I_2| du \le \frac{C_{11}}{c} \|F_{X_n} - \Phi\|_1 + C_{\rho,2} \hat{\varepsilon}_n. \tag{40}$$

13

(c) Estimate of $\int_{-\infty}^{+\infty} |I_3| \mathrm{d}u$.

Since $\Delta\langle X\rangle_k = \Delta V_k, 1 \le k < n, \Delta V_n - \Delta\langle X\rangle_n \le 2\delta_n^2$, we have

$$
\begin{aligned}
|I_3| &= \left| \frac{1}{2}\mathbf{E}[\frac{\partial^2}{\partial x^2}\Phi_u(X_{n-1}, A_n)(\Delta\langle X\rangle_n - \Delta V_n)] \right| \\
&\le \mathbf{E}[\frac{\delta_n^2}{A_n}|\varphi'(T_{n-1})|] \le \frac{1}{c^2}\mathbf{E}[|\varphi'(T_{n-1})|].
\end{aligned}
$$

Set $\tilde{G}(z) = \sup_{|z-t|\le 1} |\varphi'(t)|$, then

$$
|I_3| \le \frac{1}{c^2}\mathbf{E}[\tilde{G}(T_{n-1})].
$$

Notice that $\tilde{G}, \tilde{G}' \in L^1$, by Lemmas 3.3 and 3.1, we get

$$
\begin{aligned}
\int_{-\infty}^{+\infty} |I_3|\mathrm{d}u &\le \frac{1}{c^2}\mathbf{E}\left[\int_{-\infty}^{+\infty} \tilde{G}\left(\frac{u - X_{n-1}}{\sqrt{A_n}}\right)\mathrm{d}u \right] \\
&\le \frac{C_{12}}{c^2}\|F_{X_n} - \Phi\|_1 + C_{\rho,3}(\hat{\varepsilon}_n + \delta_n),
\end{aligned}
\tag{41}
$$

where the last line follows by the fact that $\sqrt{A_n} = c\gamma_n = c(\hat{\varepsilon}_n + \delta_n)$.

Combining (39)-(41) together, we get

$$
\int_{-\infty}^{+\infty} |I_1| + |I_2| + |I_3|\mathrm{d}u \le \frac{C_{13}}{\rho c^\rho}\|F_{X_n} - \Phi\|_1 + C_{\rho,4}(\hat{\varepsilon}_n + \delta_n).
\tag{42}
$$

Implementing the last bound in (20), by (21) we derive that

$$
\|F_{X_n} - \Phi\|_1 \le \frac{C_{14}}{\rho c^\rho}\|F_{X_n} - \Phi\|_1 + C_{\rho,5}(\hat{\varepsilon}_n + \delta_n).
\tag{43}
$$

Choosing the value c such that $\rho c^\rho = 2C_{14}$, it follows that

$$
\|F_{X_n} - \Phi\|_1 \le 2C_{\rho,5}(\hat{\varepsilon}_n + \delta_n).
\tag{44}
$$

This completes the proof of Lemma 2.3. $\qquad\square$

4. Proof of Lemma 2.4

We follow the approach of Bolthausen [4]. The main idea is to construct a new martingale difference sequence $(\hat{\xi}_i, \hat{\mathcal{F}}_i)_{1\le i\le N}$, based on a modification of $(\xi_i, \mathcal{F}_i)_{1\le i\le n}$, such that $\sum_{i=1}^N \mathbf{E}[\hat{\xi}_i^2|\hat{\mathcal{F}}_{i-1}] = 1$ a.s., and then apply Lemma 3.1 to the new martingale different sequence for obtaining Wasserstein-1 distance in the central limit theorem for X_n, that is

$$
\|F_{X_n} - \Phi\|_1 \le \|F_{\hat{X}_N} - \Phi\|_1 + 2(2t+1)\left\|\mathbf{E}[|\hat{X}_N - X_n|^{2t}|X_n]\right\|_\infty^{1/2t}.
\tag{45}
$$

The first term can be estimated via Lemma 2.3, and the second term can be dominated via Theorem 2.11 of Hall and Heyde [11].

14

We introduce the following stopping time

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

Let δ be a positive number such that $0 < \delta \le \varepsilon_n$. Let $r = \lfloor \frac{1 - \langle X \rangle_\tau}{\delta^2} \rfloor$, where $\lfloor x \rfloor$ stands for the largest integer not exceeding x. Clearly, it holds $r \le \lfloor \frac{1}{\delta^2} \rfloor$. Let $N = n + r + 1$. For $\tau + 1 \le i \le \tau + r$, let ζ_i be random variables such that

$$\mathbf{P}(\zeta_i = \pm \delta \mid \mathcal{F}_\tau) = \frac{1}{2};$$

and for $i = \tau + r + 1$, let $\zeta_{\tau+r+1}$ be such that

$$\mathbf{P}\left(\zeta_{\tau+r+1} = \pm(1 - \langle X \rangle_\tau - r\delta^2)^{1/2} \mid \mathcal{F}_\tau\right) = \frac{1}{2},$$

with the sign determined independent of everything else; and let $\zeta_i = 0$ if $\tau + r + 1 < i \le N$. Denote

$$\hat{\xi}_i = \xi_i \mathbf{1}_{\{i \le \tau\}} + \zeta_i \mathbf{1}_{\{\tau < i \le \tau + r\}} + \zeta_i \mathbf{1}_{\{i = \tau + r + 1\}}, \qquad i = 1, \dots, N, \tag{46}$$

$\hat{\mathcal{F}}_i = \mathcal{F}_i$ for $i \le \tau$ and $\hat{\mathcal{F}}_i = \sigma\{\mathcal{F}_\tau, \zeta_j, \tau + 1 \le j \le i\}$ for $\tau + 1 \le i \le N$. Then $(\hat{\xi}_i, \hat{\mathcal{F}}_i)_{1 \le i \le N}$ still forms a martingale difference sequence. Moreover, it holds that

$$\sum_{i=1}^{N} \mathbf{E}[\hat{\xi}_i^2 | \hat{\mathcal{F}}_{i-1}] = \sum_{i=1}^{\tau} \mathbf{E}[\xi_i^2 | \mathcal{F}_{i-1}] + \sum_{i=\tau+1}^{\tau+r+1} \delta^2 + (1 - \langle X \rangle_\tau - r\delta^2)$$

$$= 1 \qquad a.s.$$

and that, by the fact $\delta \le \varepsilon_n$,

$$\mathbf{E}[|\hat{\xi}_i|^{2+\rho} | \hat{\mathcal{F}}_{i-1}] \le \varepsilon_n^\rho \mathbf{E}[\hat{\xi}_i^2 | \hat{\mathcal{F}}_{i-1}], \qquad \text{for } i = 1, \dots, N.$$

Let $\hat{X}_N = \sum_{i=1}^{N} \hat{\xi}_i$, by Lemma 2.3, we have

$$\|F_{\hat{X}_N} - \Phi\|_1 \le C_\rho \hat{\varepsilon}_n. \tag{47}$$

Using Lemma 3.1, we obtain for $t > 1/2$,

$$\|F_{X_n} - \Phi\|_1 \le \|F_{\hat{X}_N} - \Phi\|_1 + 2(2t+1)\left\|\mathbf{E}\left[|\hat{X}_N - X_n|^{2t} \Big| X_n\right]\right\|_\infty^{1/2t}$$

$$\le C_\rho \hat{\varepsilon}_n + C_t \left(\mathbf{E}|\hat{X}_N - X_n|^{2t}\right)^{1/2t}. \tag{48}$$

For the estimation of $\mathbf{E}|\hat{X}_N - X_n|^{2t}$, we first note that

$$\hat{X}_N - X_n = \sum_{i=\tau+1}^{N} (\hat{\xi}_i - \xi_i),$$

where we put $\xi_i = 0$ for $i > n$. As τ is a stopping time, $(\hat{\xi}_i - \xi_i)_{\tau+1 \le i \le N}$ still forms a martingale difference sequence. By Theorem 2.11 of Hall and Heyde [11], it holds for $t > 1/2$,

$$\mathbf{E}|\hat{X}_N - X_n|^{2t} \le C_t \left(\mathbf{E}\left|\sum_{i=\tau+1}^{N} \mathbf{E}[(\hat{\xi}_i - \xi_i)^2 | \hat{\mathcal{F}}_{i-1}]\right|^t + \mathbf{E}[\max_{\tau+1 \le i \le N} |\hat{\xi}_i - \xi_i|^{2t}]\right). \tag{49}$$

15

As $\mathbf{E}[\xi_i\hat{\xi}_i|\hat{\mathcal{F}}_{i-1}] = 0$ for $\tau + 1 \le i \le N$, we get

$$\sum_{i=\tau+1}^{N} \mathbf{E}[(\hat{\xi}_i - \xi_i)^2|\hat{\mathcal{F}}_{i-1}] = \sum_{i=\tau+1}^{N} \mathbf{E}[\hat{\xi}_i^2|\hat{\mathcal{F}}_{i-1}] + \sum_{i=\tau+1}^{N} \mathbf{E}[\xi_i^2|\hat{\mathcal{F}}_{i-1}] = 1 - 2\langle X\rangle_\tau + \langle X\rangle_n.$$

Notice that $1 - \mathbf{E}[\xi_{\tau+1}^2|\mathcal{F}_\tau] < \langle X\rangle_\tau$. Hence, by the inequality

$$|a + b|^t \le \max\{2^{t-1}, 1\}\left(|a|^t + |b|^t\right), \quad t > 1/2, \tag{50}$$

and Jensen's inequality, we deduce that

$$\left|\sum_{i=\tau+1}^{N} \mathbf{E}[(\hat{\xi}_i - \xi_i)^2|\hat{\mathcal{F}}_{i-1}]\right|^t \le \left|\langle X\rangle_n - 1 + 2\mathbf{E}[\xi_{\tau+1}^2|\mathcal{F}_\tau]\right|^t$$

$$\le C_t\left(|\langle X\rangle_n - 1|^t + (\mathbf{E}[\xi_{\tau+1}^2|\mathcal{F}_\tau])^t\right)$$

$$\le C_t\left(|\langle X\rangle_n - 1|^t + \mathbf{E}[|\xi_{\tau+1}|^{2t}|\mathcal{F}_\tau]\right). \tag{51}$$

Taking expectations on both sides of the last inequality, we have

$$\mathbf{E}\left|\sum_{i=\tau+1}^{N} \mathbf{E}[(\hat{\xi}_i - \xi_i)^2|\hat{\mathcal{F}}_{i-1}]\right|^t \le C_t\left(\mathbf{E}|\langle X\rangle_n - 1|^t + \mathbf{E}|\xi_{\tau+1}|^{2t}\right)$$

$$\le C_t\left(\mathbf{E}|\langle X\rangle_n - 1|^t + \mathbf{E}[\max_{1\le i\le n}|\xi_i|^{2t}]\right). \tag{52}$$

Similarly, by inequality (50), it holds

$$\mathbf{E}[\max_{\tau+1\le i\le N}|\hat{\xi}_i - \xi_i|^{2t}] \le C_t\mathbf{E}[\max_{1\le i\le n}|\xi_i|^{2t} + \delta^{2t}]. \tag{53}$$

Applying the last inequality and (52) to (49), we get

$$\mathbf{E}|\hat{X}_N - X_n|^{2t} \le C_t\left(\mathbf{E}|\langle X\rangle_n - 1|^t + \mathbf{E}[\max_{1\le i\le n}|\xi_i|^{2t}] + \delta^{2t}\right). \tag{54}$$

Combining the last inequality with (48) and letting $\delta \to 0$, we obtain the desired inequality. $\quad\square$

References

[1] E. Baur, J. Bertoin, Elephant random walks and their connection to Pólya-type urns, Phys. Rev. E. 94 (5) (2016) 052134.

[2] B. Bercu, A martingale approach for the elephant random walk, J. Phys. A. 51 (1) (2017) 015201.

[3] B. Bercu, L. Laulin, On the center of mass of the elephant random walk, Stochastic Process. Appl. 133 (2021) 111-128.

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

[5] C.F. Coletti, R. Gava, G.M. Schütz, Central limit theorem and related results for the elephant random walk, J. Math. Phys. 58 (5) (2017) 053303.

[6] C.F. Coletti, R. Gava, G.M. Schütz, A strong invariance principle for the elephant random walk J. Stat. Mech. Theory E. 12 (2017) 123207

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

[8] X. Fan, Exact rates of convergence in some martingale central limit theorems, J. Math. Anal. Appl. 469 (2019) 1028-1044.

[9] X. Fan, H. Hu, X. Ma, Cramér moderate deviations for the elephant random walk, J. Stat. Mech. Theory E. (2021) 023402.

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

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

[12] E. Rio, Moment inequalities for sums of dependent random variables under projective conditions, J. Theoret. Probab. 22 (1) (2009) 146-163.

[13] G.M. Schütz, S. Trimper, Elephants can always remember: Exact long-range memory effects in a non-Markovian random walk, Phys. Rev. E. 70 (4) (2004) 045101.

[14] L. Van Dung, T.C. Son, N.D. Tien, L_1-bounds for some martingale central limit theorems, Lith. Math. J. 54 (1) (2014) 48-60.