Theory of Stochastic Processes
Vol. 21 (37), no. 1, 2016, pp. 53–63

G. M. MOLYBOGA

AN ANALOGUE OF THE BERRY-ESSEEN THEOREM FOR FUNCTIONALS OF WEAKLY ERGODIC MARKOV PROCESSES

An upper bound is obtained for the rate of convergence in central limit theorem for functionals of weakly ergodic Markov processes that has the rate $O\left(\frac{\ln^{3/2}(n)}{n^{1/4}}\right)$. The approach is based on the one proposed in [1, 2].

1. Introduction

In this paper we study the rate of convergence in the Central Limit Theorem (CLT) for particular class of dependent random variables. This problem naturally arises in various areas; just to mention one we refer to the analysis of asymptotic properties of statistical estimators in [3], [4], [5], and [6]. In what follows, we consider the random sequence of the form

$$Y_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \xi_k, \tag{1.1}$$

where the sequence $\{\xi_k\}$ has the form $\xi_k = A(X_k)$ with a given function A and some (time-homogeneous) Markov process X.

By analogy with the classical Berry-Esseen theorem for i.i.d. sequences, one naturally expects to have the rate of convergence in a CLT equal to $O(1/\sqrt{n})$ in models with dependence, as well. For stationary sequences with α-mixing properties, such a rate (up to some slowly increasing function) was obtained in [7] by means of a proper version of the Stein method. The Stein method was also used, combined with the Malliavin calculus, in [8] where the rate $O(1/\sqrt{n})$ was obtained for normally distributed Markov processes. The same rate was obtained in [9] for general stationary Markov processes which possess L_2-spectral gap property.

In our studies, we are mainly interested in the Markov processes which has the *intrinsic memory* property, which mean that for any $x \neq y$ the transition probabilities $P_k(x, \cdot)$ and $P_k(y, \cdot)$ are mutually singular for any $k \geq 1$. The excellent example of a process with such an effect is given in [10] in the framework of stochastic differential delay equations. This effect occurs quite typically when the Markov process has a complicated state space, e.g. in models which involve SDDEs, SPDEs, equations with fractional noises, etc., hence the analysis of the related limit theorems is a natural problem. We note that, for the Markov processes with intrinsic memory, the *stabilization* (i.e. convergence of the transition probabilities to the invariant distribution) is not possible in the total variation distance, but is still possible in a weaker sense; we refer to the detailed exposition of this effect and further references to [11]. Markov processes which exhibit stabilization in a weaker sense than the convergence in the total variation distance, are called *weakly ergodic*.

For Markov processes with intrinsic memory the methods developed in [7] – [9] are hardly applicable. The Stein method actually exploits the property of the law of the

2010 *Mathematics Subject Classification.* 60F05; 60J05.
Key words and phrases. Markov process, corrector, weak ergodicity.

bridge of the process X to be stabilized when the starting point and the end point tend to $\mp\infty$. The "intrinsic memory" feature actually means that the entire previous trajectory can be uniquely determined given a unique value of the process, and in this case the "weak stabilization" the law of the bridge is not possible. This makes a large difference with the weak stabilization of the law of the process with a fixed starting point, and makes it apparently impossible to use the Stein method in such a framework. The weak spectral method, developed in [9], is based on Keller-Liverani perturbation theorem and exploits essentially the spectral gap property of the L_2-semigroup associated with the process. Such a property is slightly stronger than the α-mixing property, and fails for the processes with intrinsic memory. Vaguely speaking, for such a process the semigroup may exhibit a convergence to an equilibrium on some functions g which has an additional regularity properties (e.g. Lemma 4.1 below), but such a convergence can not be uniform in $g \in L_2$. Such a uniformity is crucial in the approach developed in [9], thus we can not apply it here.

For weakly ergodic Markov processes one possible way to provide the CLT is to use the "individual" (or "weak") stabilization property explained above and use the standard martingale CLT [12]; we refer to [13] for details. The rate of convergence was not studied in [13], but the general bound for such a rate in the framework of the martingale CLT was established in [14], and one can check that in the case under the consideration this bound will have the form $O(1/n^{1/8})$. A more direct approach, which provides the CLT for weakly ergodic processes, which is a modification of the "corrector term" approach from in [15], was developed in [1]; see also [11]. In the current paper we analyze the bound for the rate of convergence which can be obtained within the corrector term approach. We will obtain the rate $O\left(\frac{\ln^{3/2}(n)}{n^{1/4}}\right)$, which yet do not achieve the optimal rate $O(n^{-1/2})$ available in simpler settings ([7, 9]), but at the same time is better than the general martingale CLT rate $O(n^{-1/8})$ from [14].

The structure of the paper is as follows. Preliminaries are presented in Section 2. In Section 3 we formulate and prove the main result. The proof is essentially based on an auxiliary inequality (3.4). This inequality is proved in Section 4.

2. Preliminaries

Let $\mathbf{X}$ be a Polish space, and d be a distance-like function on it; that is, a nonnegative, symmetric, lower semicontinuous function on $\mathbf{X} \times \mathbf{X}$ such that $d(x, x) = 0, x \in \mathbf{X}$, $d(x, y) > 0, x \neq y, x \in \mathbf{X}, y \in \mathbf{X}$. For an arbitrary probability measures μ, ν on $\mathbf{X}$ we define

$$d(\mu, \nu) = \inf_{(\xi,\eta)\in C(\mu,\nu)} Ed(\xi, \eta), \tag{2.1}$$

where $C(\mu, \nu)$ denotes the set of pairs (ξ, η) of random elements valued in X such as $\text{Law}(\xi) = \mu$, $\text{Law}(\eta) = \nu$.

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be some probability space. Consider a time-homogeneous stationary Markov process $X = \{X_n, n \geq 0\}$ taking values in $\mathbf{X}$. Denote its transition probability by $P_k(x, dy)$. Let $\mathcal{F}^X$ be the natural filtration of process X.

For some function $V : \mathbf{X} \to \mathbb{R}_+$, distance-like function d and real $\rho \in (0, 1)$, we say that the process X satisfies condition $M(V, d, \rho)$ if X admits a unique invariant measure π and the following holds:

$$d(P_k(x, \cdot), \pi) \leq V(x)\rho^k, \tag{2.2}$$

$$\int_{\mathbb{X}} V^2 d\pi < \infty. \tag{2.3}$$

In what follows, by C we denote an arbitrary constant whose value is not specified and may vary from place to place.

For a real $p > 1$ and distance-like function d, we denote by $H_{d,p}$ the class of all functions $f : \mathbf{X} \to \mathbb{R}$ such that:

$$||f||_{d,p} = \sup_{x,y \in \mathbf{X}} \frac{|f(x) - f(y)|}{(d(x,y))^{1/p}} < \infty. \tag{2.4}$$

For process X that satisfies $M(V, d, \rho)$ and function $f \in H_{d,p}$ such that $\int_{\mathbf{X}} f d\pi = 0$, we denote by

$$\mathcal{R}f(x) = \sum_{k=0}^{\infty} E_x^X f(X_k) \tag{2.5}$$

the *potential* of f. The series (2.5) converges in the mean-square sense with respect to π. The proof of this fact can be found in Section 3.3 of [11]. Let us also consider the following function that is correctly defined under the same assumptions:

$$\hat{\mathcal{R}}f(x) = \sum_{k=1}^{\infty} E_x^X f(X_k) \tag{2.6}$$

3. The main result

Theorem 3.1. *Suppose the following conditions hold:*

(1) *For some $V : \mathbf{X} \to \mathbb{R}_+$, distance-like function d and real $\rho \in (0,1)$ process X satisfies $M(V, d, \rho)$;*
(2) $\int_{\mathbf{X}} A d\pi = 0$, $\int_{\mathbf{X}} |A|^4 d\pi < \infty$;
(3) *For some real $p > 1$: A, A^2, $\hat{\mathcal{R}}A$ and $A\hat{\mathcal{R}}A$ belong to $H_{d,p}$;*
(4) $\sigma^2 := \int_{\mathbf{X}} \left((A(y))^2 + 2A(y)\hat{\mathcal{R}}A(y) \right) d\pi > 0$.

Then

$$\sup_{z \in \mathbf{R}} \left| \mathbf{P}\left(\frac{1}{\sqrt{n}} \sum_{k=1}^{n} A(X_k) < z\right) - \Phi\left(\frac{z}{\sigma}\right) \right| \le C \frac{\ln^{3/2}(n)}{n^{1/4}}, \tag{3.1}$$

where Φ denotes the distribution function of standard normal distribution.

To illustrate the theorem, we examine the example from [11], Section 2.3:
Consider an SDDE in $\mathbb{R}$ of the form:

$$dX_t = -cX_t dt + g(X_{t-1}) dW_t, \tag{3.2}$$

where $c > 0$, W is a one-dimensional Brownian motion and g is a strictly increasing positive bounded Lipschitz continuous function such that:

$$c > ||g||^2_{Lip}. \tag{3.3}$$

Following the [11], we state that *segment process* $X(t) = \{X_{t+s}, s \in [-1, 0]\} \in C(-1, 0), t \ge 0$ possesses the Markov property. It was mentioned that:

$$||X(t) - X'(t)||_{TV} = 2,$$

where the pair of processes $X(t)$ and $X'(t)$ started from two different initial points $x, x' \in C(-1, 0)$. Let $d(x, y) = ||x - y||^2_{C(-1,0)}$. Then:

$$Ed(X, X') \le Ce^{-(c - ||g||^2_{Lip})t} d(x, x')$$

By the Itô formula, because g is bounded one can get:

$$EX_t^4 - X_0^4 = \int_0^t \Big[-4cEX_s^4 + 6g^2(X_{s-1})EX_s^2 \Big] ds \le \int_0^t \Big[-3cEX_s^4 + C \Big] ds.$$

From the comparison principle we get the integrability of EX_t^4. Following the methodology presented in [11], Section 2.3, one can also show that for some $x \in C(-1,0)$: $\sup_{t\ge 0} E_x ||X(t)||^4_{C(-1,0)} < \infty$. Then, by Theorem 2.25 from [11] there exists a unique invariant measure π and the following rate of convergence holds true:

$$d(P_t(x,\cdot),\pi(\cdot)) \le Ce^{-(c-||g||^2_{Lip})t} \int_{C(-1,0)} ||x-x'||^2 \pi(dx'), t \ge 0.$$

Hence, in 2.2 we have $\rho = e^{-(c-||g||^2_{Lip})}$ and $V(x) = C \int_{C(-1,0)} ||x-x'||^2\pi(dx')$. The condition 2.3 can be get from the estimate on $\sup_{t\ge 0} E_x||X(t)||^4_{C(-1,0)}$.

The proof of Theorem 3.1 is based on two following lemmas.

Lemma 3.1. *Under the conditions of Theorem 3.1, the following inequality holds true:*

$$|Ee^{i\lambda \frac{1}{\sqrt{n}} \sum_{k=1}^{n} A(X_k)} - e^{-\frac{\lambda^2\sigma^2}{2}}| \le C\frac{|\lambda|\ln^3(n)}{\sqrt{n}}, \tag{3.4}$$

for all natural n and real λ such that $|\lambda| < \frac{n^{1/4}}{\ln^{3/2}(n)}$.

We prove Lemma 3.1 in Section 4.

Lemma 3.2. *Suppose F and G are some distribution functions with the same mean. Suppose ϕ and γ be respectively the characteristics functions of F and G. If also $\exists g = G', \exists m$ such that $|g| \le m$, then the following holds true:*

$$|F(x) - G(x)| \le \frac{1}{\pi}\int_{-T}^{T} |\frac{\phi(\lambda)-\gamma(\lambda)}{\lambda}| d\lambda + \frac{24m}{\pi T},$$

for all real x and positive T.

Proof of Lemma 3.2 can be found in Section 16.3 of [16].

Proof of Theorem 3.1. Apply Lemma 3.2 with

$$F = \mathbf{P}(\frac{1}{\sqrt{n}}\sum_{k=1}^{n} A(X_k) < z), G = \Phi(\frac{z}{\sigma}), T = \frac{n^{1/4}}{\ln^{3/2}(n)}.$$

Thus, we have

$$|\mathbf{P}(\frac{1}{\sqrt{n}}\sum_{k=1}^{n} A(X_k) < z) - \Phi(\frac{z}{\sigma})| \le \frac{2CT\ln^3(n)}{\pi\sqrt{n}} + \frac{4}{T} \le (\frac{2C}{\pi}+4)\frac{\ln^{3/2}(n)}{n^{1/4}}.$$

This completes the proof of Theorem 3.1. □

We note that in [14] an analogue of the inequality (3.4) is used with $\frac{C|\lambda|^3}{\sqrt{n}}$ in the right hand side. It is not difficult to show that in this case Lemma 3.2 provides the estimate on the rate of convergence of order $O(n^{-1/8})$.

So, the main goal is to get inequality of the form (3.4). The method of corrector term let us analyze the error more precisely.

4. The proof of Lemma 3.1

Let us consider not only $Y_n = \frac{1}{\sqrt{n}} \sum\limits_{k=1}^{n} A(X_k)$ but the series of random variables

$$Y_n(\frac{k}{n}) = \frac{1}{\sqrt{n}} \sum_{j=1}^{k} A(X_j). \tag{4.1}$$

It is clear that $Y_n(1) = Y_n$. Representation (4.1) allows us to get a recurrence relation:

$$e^{i\lambda Y_n(\frac{k}{n})} = e^{i\lambda Y_n(\frac{k-1}{n})} e^{\frac{i\lambda}{\sqrt{n}}A(X_k)}$$

From the Taylor expansion for e^x with $x = \frac{i\lambda}{\sqrt{n}} A(X_k)$, we have:

$$e^{i\lambda Y_n(\frac{k}{n})} = e^{i\lambda Y_n(\frac{k-1}{n})} + \frac{i\lambda}{\sqrt{n}} A(X_k) e^{i\lambda Y_n(\frac{k-1}{n})} - \frac{\lambda^2}{2n} A^2(X_k) e^{i\lambda Y_n(\frac{k-1}{n})} + r_n^1(k), \tag{4.2}$$

where

$$|r_n^1(k)| \le \frac{|\lambda|^3 |A(X_k)|^3}{6n\sqrt{n}}. \tag{4.3}$$

4.1. **Corrector term.** Following the methodology explained in [11], we introduce the *corrector* term:

$$U_n(k) = \hat{\mathcal{R}}A(X_k) i\lambda e^{i\lambda Y_n(\frac{k-1}{n})}, \quad k \ge 1. \tag{4.4}$$

Denote $M_k = \hat{\mathcal{R}}A(X_k) + \sum\limits_{j=0}^{k} A(X_j)$. It is proved in Section 3.3 of [11] that M_k is a martingale. Therefore, $E\big[\Delta_k M | \mathcal{F}_{k-1}^X\big] = 0$. By definition, $\Delta_k \hat{\mathcal{R}}A(x) = -A(X_k) + \Delta_k M$. Let us consider an increment of the corrector term:

$$\begin{aligned} U_n(k) - U_n(k-1) = {} & (\hat{\mathcal{R}}A(X_k) - \hat{\mathcal{R}}A(X_{k-1})) i\lambda e^{i\lambda Y_n(\frac{k-1}{n})} \\ & + \hat{\mathcal{R}}A(X_{k-1}) i\lambda (e^{i\lambda Y_n(\frac{k-1}{n})} - e^{i\lambda Y_n(\frac{k-2}{n})}). \end{aligned} \tag{4.5}$$

Here we can use (4.2) and get:

$$\begin{aligned} & U_n(k) - U_n(k-1) = i\lambda e^{i\lambda Y_n(\frac{k-1}{n})} (-A(X_k) + \Delta_k M) \\ & + \hat{\mathcal{R}}A(X_{k-1}) i\lambda (\frac{i\lambda}{\sqrt{n}} A(X_{k-1}) e^{i\lambda Y_n(\frac{k-2}{n})} - \frac{\lambda^2}{2n} A^2(X_{k-1}) e^{i\lambda Y_n(\frac{k-2}{n})} + r_n^1(k-1)). \end{aligned} \tag{4.6}$$

Let us "correct" the decomposition (4.2) by adding the equation (4.6) divided on $\sqrt{n}$:

$$\begin{aligned} e^{i\lambda Y_n(\frac{k}{n})} + \frac{1}{\sqrt{n}} U_n(k) = {} & e^{i\lambda Y_n(\frac{k-1}{n})} + \frac{1}{\sqrt{n}} U_n(k-1) \\ & + \frac{1}{\sqrt{n}} i\lambda e^{i\lambda Y_n(\frac{k-1}{n})} (-A(X_k) + \Delta_k M) \\ & + \frac{i\lambda}{\sqrt{n}} A(X_k) e^{i\lambda Y_n(\frac{k-1}{n})} - \frac{\lambda^2}{2n} A^2(X_k) e^{i\lambda Y_n(\frac{k-1}{n})} \\ & + \hat{\mathcal{R}}A(X_{k-1}) i\lambda (\frac{i\lambda}{n} A(X_{k-1}) e^{i\lambda Y_n(\frac{k-2}{n})}) - r_n^1(k) \\ & + \hat{\mathcal{R}}A(X_{k-1}) \frac{i\lambda}{\sqrt{n}} (-\frac{\lambda^2}{2n} A^2(X_{k-1}) e^{i\lambda Y_n(\frac{k-2}{n})} + r_n^1(k-1)). \end{aligned} \tag{4.7}$$

Denote

$$r_n^2(k) = \hat{\mathcal{R}}A(X_{k-1})\frac{i\lambda}{\sqrt{n}}(-\frac{\lambda^2}{2n}A^2(X_{k-1})e^{i\lambda Y_n(\frac{k-2}{n})} + r_n^1(k-1)). \tag{4.8}$$

Using (4.3), we get:

$$|r_n^2(k)| \le \frac{|\lambda|^3}{n\sqrt{n}}\Big(\frac{|\hat{\mathcal{R}}A(X_{k-1})A(X_{k-1})^2|}{2} + \frac{|\hat{\mathcal{R}}A(X_{k-1})A(X_{k-1})^3|}{6}\Big). \tag{4.9}$$

From the equation (4.7) and the definition (4.8) it follows that

$$\begin{aligned} e^{i\lambda Y_n(\frac{k}{n})} + \frac{1}{\sqrt{n}}U_n(k) &= e^{i\lambda Y_n(\frac{k-1}{n})} + \frac{1}{\sqrt{n}}U_n(k-1) \\ &- \frac{\lambda^2}{2n}\left(A^2(X_k)e^{i\lambda Y_n(\frac{k-1}{n})} + 2\hat{\mathcal{R}}A(X_{k-1})A(X_{k-1})e^{i\lambda Y_n(\frac{k-2}{n})}\right) \\ &+ \frac{i\lambda}{\sqrt{n}}\Delta_k M e^{i\lambda Y_n(\frac{k-1}{n})} + r_n^1(k) + r_n^2(k). \end{aligned} \tag{4.10}$$

Using the same equations with k replaced by $k-1, k-2, ..., 1$, we get

$$\begin{aligned} e^{i\lambda Y_n(\frac{k}{n})} + \frac{1}{\sqrt{n}}U_n(k) = 1 &+ \frac{i\lambda}{\sqrt{n}}\sum_{j=1}^{k-1}\Delta_{j+1}Me^{i\lambda Y_n(\frac{j}{n})} \\ &+ \frac{\lambda^2}{2n}A^2(X_k)e^{i\lambda Y_n(\frac{k-1}{n})} - \frac{\lambda^2}{2n}A^2(X_1) \\ &- \frac{\lambda^2}{2n}\sum_{j=1}^{k-1}\left((A^2(X_j) + 2\hat{\mathcal{R}}A(X_j)A(X_j))e^{i\lambda Y_n(\frac{j-1}{n})}\right) \\ &+ \sum_{j=1}^{k-1}(r_n^1(j+1) + r_n^2(j+1)). \end{aligned} \tag{4.11}$$

Denote by

$$R_k^1 = \sum_{j=2}^{k}(r_n^1(j) + r_n^2(j)), \tag{4.12}$$

and

$$r_n^3(k) = \frac{\lambda^2}{2n}A^2(X_k)e^{i\lambda Y_n(\frac{k-1}{n})} - \frac{\lambda^2}{2n}A^2(X_1) - \frac{1}{\sqrt{n}}U_n(k) \tag{4.13}$$

the errors presented in (4.11). Since $|\lambda| < \sqrt{n}$, we have:

$$|r_n^3(k)| \le \frac{|\lambda|(|\hat{\mathcal{R}}A(X_k)| + 1)}{\sqrt{n}}, \tag{4.14}$$

for sufficiently large n.

Denote by

$$B(X_j) = A^2(X_j) + 2\hat{\mathcal{R}}A(X_j)A(X_j). \tag{4.15}$$

Taking the expectation at both sides of the equation in (4.11), we get:

$$Ee^{i\lambda Y_n(\frac{k}{n})} = 1 - \frac{\lambda^2}{2n}E\sum_{j=1}^{k-1}\left(B(X_j)e^{i\lambda Y_n(\frac{j-1}{n})}\right) + E\left(R_k^1 + r_n^3(k)\right). \tag{4.16}$$

4.2. **Time delay.** In order to analyze the sum in the identity (4.16), we apply the *time delay* trick similar to the one used in [11]. We represent the sum $\sum_{j=1}^{k-1} E\left[B(X_j)e^{i\lambda Y_n(\frac{j-1}{n})}\right]$ in the following way:

$$\sum_{j=1}^{k-1} E\left[B(X_j)e^{i\lambda Y_n(\frac{j-1}{n})}\right] = \sum_{j=1}^{l(n)} E\left[B(X_j)e^{i\lambda Y_n(\frac{j-1}{n})}\right] + \sum_{j=l(n)}^{k-1} E\left[B(X_j)(e^{i\lambda Y_n(\frac{j-1}{n})} - e^{i\lambda Y_n(\frac{j-l(n)}{n})})\right] + \sum_{j=l(n)}^{k-1} E\left[B(X_j)e^{i\lambda Y_n(\frac{j-l(n)}{n})}\right] =: \Sigma_1 + \Sigma_2 + \Sigma_3, \tag{4.17}$$

where $l(n)$ is some function that will be specified later. Let us estimate the expectation of each sum of the partition separately.

Since

$$|EB(X_j)e^{i\lambda Y_n(\frac{j-1}{n})}| \le E|B(X_j)| \le 2EA^2(X_j) + E(\hat{\mathcal{R}}A(X_j))^2, \tag{4.18}$$

we have

$$E|\sum_{j=1}^{l(n)} \left[B(X_j)e^{i\lambda Y_n(\frac{j-1}{n})}\right]| \le l(n)\left(2EA^2(X_1) + E(\hat{\mathcal{R}}A(X_1))^2\right). \tag{4.19}$$

Before analyzing Σ_2, we give two auxiliary results. The proofs of the subsequent statements can be found in Section 3.2 of [11].

Lemma 4.1. *Suppose $g \in H_{d,p}$. Then*

$$|E_x g(X_k) - \int_{\mathbf{X}} g d\pi| \le ||g||_{d,p}(V(x)\rho^k)^{1/p}. \tag{4.20}$$

Lemma 4.2. *Suppose $g \in H_{d,p}$ and $\int_{\mathbf{X}} g^2 d\pi < \infty$; Then*

$$|Cov\,(g(X_k), g(X_j))\,| \le C||g||_{d,p}\rho^{(k-j)/p}. \tag{4.21}$$

From the Cauchy inequality for the j-th term of Σ_2, we get:

$$E\left[|B(X_j)(e^{i\lambda Y_n(\frac{j-1}{n})} - e^{i\lambda Y_n(\frac{j-l(n)}{n})})|\right] \le \left(E[|B(X_j)|^2]E[|(e^{i\lambda Y_n(\frac{j-1}{n})} - e^{i\lambda Y_n(\frac{j-l(n)}{n})})|^2]\right)^{1/2}. \tag{4.22}$$

Using the recurrence relation (4.2) with k replaced by $m, m-1, ..., m-k$, we get:

$$e^{i\lambda Y_n(\frac{m}{n})} - e^{i\lambda Y_n(\frac{m-k}{n})} = \sum_{j=m-k}^{m-1}\left(\frac{i\lambda A(X_j)}{\sqrt{n}} - \frac{\lambda^2\big(A(X_j)\big)^2}{2n}\right) - \sum_{j=m-k}^{m-1} r_n^1(j), \tag{4.23}$$

From Lemma 4.2, we have:

$$|Cov(A(X_k), A(X_j))| \le C||A||_{d,p}\rho^{(k-j)/p} \tag{4.24}$$

From (4.23) and the last inequality (4.24), we obtain an upper bound for the second expectation:

$$E[|(e^{i\lambda Y_n(\frac{j-1}{n})}-e^{i\lambda Y_n(\frac{j-l(n)}{n})})|^2] \le C\Big(\frac{l(n)\lambda^2||A||_{d,p}}{n(1-\rho^{1/p})}+\frac{|\lambda|^3(l(n))^2(E|A(X_1)|^3+1)}{n\sqrt{n}}+\frac{|\lambda|^4(l(n))^2(E|A(X_1)|^4+1)}{n^2}\Big). \tag{4.25}$$

Finally, for the entire j-th term of Σ_2, we get:

$$E\left[|B(X_j)(e^{i\lambda Y_n(\frac{j-1}{n})}-e^{i\lambda Y_n(\frac{j-l(n)}{n})})|\right] \le C\left(\left(\frac{l(n)\lambda^2}{n}+\frac{|\lambda|^3(l(n))^2}{n\sqrt{n}}\right)^{1/2}\right), \tag{4.26}$$

where constant C does not depend on j. Consider the conditional expectation of the j-th term of Σ_3 with respect to $\mathcal{F}^X_{j-l(n)}$:

$$E\Big[E(B(X_j)e^{i\lambda Y_n(\frac{j-l(n)}{n})}|\mathcal{F}^X_{j-l(n)})\Big]=E\Big[e^{i\lambda Y_n(\frac{j-l(n)}{n})}E(B(X_j)|\mathcal{F}^X_{j-l(n)})\Big].$$

By the theorem assumptions, $A^2(x), A(x)\hat{\mathcal{R}}A(x)\in H_{d,p}$. Hence, it is not hard to prove that $B(x)\in H_{d,p}$.

So, from Lemma 4.1 with $g(x)=B(x)$, we have

$$|E(B(X_j)|F^X_{j-l(n)})-E(B(X_j))|\le ||B||_{d,p}(V(X_{j-l(n)}))^{1/p}\rho^{l(n)/p}. \tag{4.27}$$

Specify $l(n)=[\ln^2(n)]$. From the inequality (4.27), we get:

$$\begin{aligned}\Big|\sum_{j=[\ln^2(n)]}^{k-1}\Big(E\left[B(X_j)e^{i\lambda Y_n(\frac{j-[\ln^2(n)]}{n})}\right]-EB(X_j)E\left[e^{i\lambda Y_n(\frac{j-[\ln^2(n)]}{n})}\right]\Big)\Big| &\le \sum_{j=[\ln^2(n)]}^{k-1}||B||_{d,p}E(V(X_{j-[\ln^2(n)]}))^{1/p}n^{\ln(\rho)\ln(n)/p}\\ &\le ||B||_{d,p}E(V(X_0))^{1/p}n^{1+\ln(\rho)\ln(n)/p}.\end{aligned} \tag{4.28}$$

Since $\rho\in(0,1)$, $n^{1+\ln(\rho)\ln(n)/p}$ tends to zero faster than an arbitrary power function of n^{-1}. Recall that σ^2 is equal to $EB(X_0)$. This yields that

$$\big|\Sigma_3-\sigma^2\sum_{j=[\ln^2(n)]}^{k-1}Ee^{i\lambda Y_n(\frac{j}{n})}-\sigma^2\sum_{j=1}^{[\ln^2(n)]-1}Ee^{i\lambda Y_n(\frac{j}{n})}\big|=O([\ln^2(n)]). \tag{4.29}$$

Let us denote $\phi_k(\lambda)=Ee^{i\lambda Y_n(\frac{k}{n})}$.

Denote by R^2_n the sum of all errors that we get under the "time delay" trick, multiplied by $\frac{\lambda^2}{2n}$. By the above

$$R^2_n=v_n+\frac{\lambda^2}{2n}\sum_{k=1}^{n}w_k\frac{\ln(n)|\lambda|}{\sqrt{n}}. \tag{4.30}$$

We stress that there exist constants w, v such as $|w_k|\le w, |v_n|\le\frac{v[\ln^2(n)]\lambda^2}{2n}$.

Let $R_{total}(k)=E\left(R^1_k+R^2_k+r^3_k\right)$; Then

$$\phi_n(\lambda)=1-\frac{\lambda^2\sigma^2}{2n}\sum_{j=1}^{n-1}\phi_j(\lambda)+R_{total}(n). \tag{4.31}$$

It can be shown in the usual way that the solution of (4.31) is:

$$\phi_n(\lambda) = (1-\frac{\lambda^2\sigma^2}{2n})^n + \sum_{k=1}^{n}(R_{total}(k)-R_{total}(k-1))(1-\frac{\lambda^2\sigma^2}{2n})^{n-k}. \tag{4.32}$$

4.3. **The inequality (3.4).** Consider the right hand side of the equation (4.32).

$$\left|e^{n\ln(1-\frac{\lambda^2\sigma^2}{2n})} - e^{-\frac{\lambda^2\sigma^2}{2}}\right| \le e^{-\frac{\lambda^2\sigma^2}{2}}\left(e^{\frac{\lambda^4\sigma^4}{8n}}-1\right) \le e^{-\frac{\lambda^2\sigma^2}{2}} \tag{4.33}$$

$$\sum_{k=1}^{n}(R_{total}(k)-R_{total}(k-1))(1-\frac{\lambda^2\sigma^2}{2n})^{n-k}. \tag{4.34}$$

By the definition,

$$R_{total}(k) = P_k + Q_k, \tag{4.35}$$

where

$$P_k = E\sum_{j=1}^{k}\left(r_j^1 + r_j^2 + \frac{\lambda^3\ln(n)}{2n\sqrt{n}}w_j\right), \tag{4.36}$$

$$Q_k = E\left(\frac{\lambda^2}{2n}A^2(X_k)e^{i\lambda Y_n(\frac{k-1}{n})} - \frac{\lambda^2}{2n}A^2(X_1) - \frac{1}{\sqrt{n}}U_n(k) + v_k\right). \tag{4.37}$$

Let us remark that $|Q_k| \le C\frac{\lambda}{\sqrt{n}}$ for sufficiently large n. Note also that

$$P_k = \sum_{j=1}^{k} p_j \tag{4.38}$$

and

$$\exists p:\ \forall j: |p_j| \le p\left(\frac{|\lambda|^3\ln(n)}{n\sqrt{n}}\right).$$

Using decomposition (4.35) and the Abel transformation, we get:

$$\begin{aligned}\sum_{k=1}^{n}(\Delta_k R_{total})(1-\frac{\lambda^2\sigma^2}{2n})^{n-k} &= \sum_{k=1}^{n}(\Delta_k P)(1-\frac{\lambda^2\sigma^2}{2n})^{n-k} \\ &- \sum_{k=1}^{n-1}(Q_k\frac{\lambda^2\sigma^2}{2n})(1-\frac{\lambda^2\sigma^2}{2n})^{n-k-1} + Q_n - Q_0(\frac{\lambda^2\sigma^2}{2n})^{n-1}.\end{aligned} \tag{4.39}$$

It is clear that for all k: $\Delta_k P$ and $(Q_k\frac{\lambda^2\sigma^2}{2n})$ are of order $O(\frac{\ln(n)|\lambda|^3}{n\sqrt{n}})$.

We claim the following Lemma.

Lemma 4.3. *Suppose a sequence z_k satisfies $\max_{k=1..n}|z_k| \le \frac{z\ln(n)|\lambda|^3}{n\sqrt{n}}$, for some positive z; Then*

$$\left|\sum_{k=1}^{n} z_k(1-\frac{\lambda^2\sigma^2}{2n})^{n-k}\right| \le (z+1)\,\frac{\ln^3(n)|\lambda|}{\sqrt{n}}$$

Proof. First note that $\sum\limits_{k=1}^{n} z_k(1-\frac{\lambda^2\sigma^2}{2n})^{n-k}$ can be represented in the following way

$$\sum_{k=1}^{n-k(n)} z_k(1-\frac{\lambda^2\sigma^2}{2n})^{n-k}, \tag{4.40}$$

$$\sum_{k=n-k(n)+1}^{n} z_k(1-\frac{\lambda^2\sigma^2}{2n})^{n-k}, \tag{4.41}$$

where $k(n)$ is some function that will be specified later.

First we show the estimate for (4.40)

$$|\sum_{k=1}^{n-k(n)} z_k(1-\frac{\lambda^2\sigma^2}{2n})^{n-k}| \le (1-\frac{\lambda^2\sigma^2}{2n})^{k(n)}\frac{\ln(n)|\lambda|^3 z}{\sqrt{n}}. \tag{4.42}$$

Suppose $k(n) = \min\left(\frac{n\ln^2(n)}{\lambda^2}, n\right)$. From an analogue of the decomposition (4.33), we have

$$(1-\frac{\lambda^2\sigma^2}{2n})^{k(n)} \le e^{-\frac{\lambda^2 k(n)}{n}}, \tag{4.43}$$

As above, the right hand side of (4.43) tends to zero faster than an arbitrary power function of n^{-1} as $k(n) = \frac{n\ln^2(n)}{\lambda^2}$. Otherwise, if $k(n) = n$ then the sum (4.40) is equal to zero. So, it can be bounded by $\frac{\ln^3(n)|\lambda|}{\sqrt{n}}$.

To conclude the proof, it remains to note that $z\frac{\ln^3(n)|\lambda|}{\sqrt{n}}$ is the estimate for (4.41). □

So, using Lemma 4.3 with $z_k = \Delta_k P$ and $z_k = \frac{Q_k\lambda^2\sigma^2}{2n}$, we get

$$|\phi(n) - e^{-\frac{\lambda^2\sigma^2}{2}}| \le C\frac{|\lambda|\ln^3(n)}{\sqrt{n}}. \tag{4.44}$$

This completes the proof of Lemma 3.1. □

References

1. A. M. Kulik and A. Yu. Veretennikov, *Diffusion approximation of systems with weakly ergodic Markov perturbations I*, Theor. Probability and Math. Statist. **87** (2013), 13–29.
2. A. M. Kulik and A. Yu. Veretennikov, *Diffusion approximation of systems with weakly ergodic Markov perturbations II*, Theor. Probability and Math. Statist. **88** (2014), 1–17.
3. Y. Mishura and K. Ralchenko, *On Drift Parameter Estimation in Models with Fractional Brownian Motion by Discrete Observations*, Austrian Journal of Statistics **43** (2014), no. 3, 217–228.
4. P. Billingsley, *Statistical inference for Markov processes*, The University of Chicago Press, Chicago, 1961.
5. Y. Hu, D. Nualart, W. Xiao and W. Zhang, *Exact maximum likelihood estimator for drift fractional Brownian motion at discrete observation*, Acta Mathematica Scientia **31** (2011), no. 5, 1851–1859.
6. L. Hervé, J. Ledoux and V. Patilea, *A uniform Berry-Esseen theorem on M-estimators for geometrically ergodic Markov chains*, Bernoulli **18** (2012), no. 2, 703–734.
7. A. N. Tihomirov, *Convergence rate in the central limit theorem for weakly dependent random variables*, Theory of Probability and its Applications **25** (1980), no. 4, 790–809.
8. I. Nourdin and G. Peccati, *Stein's method and exact Berry-Esseen asymptotics for functionals of Gaussian fields*, The Annals of Probability **37** (2009), no. 6, 2231–2261.
9. D. Ferré, L. Hervé and J. Ledoux, *Limit theorems for stationary Markov processes with L^2-spectral gap*, Annales de l'Institut Henri Poincaré Probabilités et Statistiques **48** (2012), no. 2, 396–423.
10. M. Hairer, J. C. Mattingly and M. Scheutzow, *Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations*, Probability Theory and Related Fields **149** (2011), no. 1-2, 223–259.

11. A. M. Kulik, *Introduction to Ergodic rates for Markov chains and processes*, Potsdam University Press, Potsdam, 2015.
12. J. Jacod and A. N. Shiryaev, *Limit Theorems for Stochastic Processes*, Springer, 2003.
13. T. Komorowski and A. Walczuk, *Central limit theorem for Markov processes with spectral gap in the Wasserstein metric*, Stochastic Processes and their Applications **122** (2012), no. 5, 2155–2184.
14. R. Sh. Liptser and A. N. Shiryaev, *On the Rate of Convergence in the Central Limit Theorem for Semimartingales*, Theory of Probability and its Applications **27** (1982), no. 1, 1–13.
15. G. C. Papanicolaou, D. W. Stroock and S. R. S. Varadhan, *Martingale approach to some limit theorems*, Duke Univ. Series, 1977.
16. W. Feller, *An Introduction to Probability Theory and Its Applications, Volume 2*, Wiley, New York, 1971.

Ukraine, Kiev, 3, Tereschenkivska st
E-mail address: St.George.Molyboga@gmail.com