Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications 119 (2009) 3699–3718

stochastic processes and their applications

www.elsevier.com/locate/spa

New techniques for empirical processes of dependent data

Herold Dehling[a], Olivier Durieu[b,*], Dalibor Volny[b]

[a] *Fakultät für Mathematik, Ruhr-Universität Bochum, Universitätsstraße 150, 44780 Bochum, Germany*
[b] *Laboratoire de Mathematique Raphaël Salem, UMR 6085 CNRS-Université de Rouen, France*

Received 2 October 2008; received in revised form 29 June 2009; accepted 4 July 2009
Available online 15 July 2009

Abstract

We present a new technique for proving the empirical process invariance principle for stationary processes $(X_n)_{n\geq 0}$. The main novelty of our approach lies in the fact that we only require the central limit theorem and a moment bound for a restricted class of functions $(f(X_n))_{n\geq 0}$, not containing the indicator functions. Our approach can be applied to Markov chains and dynamical systems, using spectral properties of the transfer operator. Our proof consists of a novel application of chaining techniques.

MSC: 60F17; 60G10; 62G30

Keywords: Empirical process; Invariance principle; Markov chain; Dynamical system; Chaining

1. Introduction

Let $(X_n)_{n\geq 0}$ be a stationary ergodic process of $\mathbb{R}$-valued random variables with marginal distribution function $F(t) = P(X_0 \leq t)$. Define the empirical distribution function $(F_n(t))_{t\in\mathbb{R}}$ and the empirical process $(U_n(t))_{t\in\mathbb{R}}$ by

$$F_n(t) := \frac{1}{n}\sum_{i=1}^{n} 1_{(-\infty,t]}(X_i), \quad t \in \mathbb{R},$$

* Corresponding address: Laboratoire de Mathematique Raphaël Salem, UMR 6085 CNRS-Université de Rouen, Avenue de l'Universite, BP. 12, Technopole du Madrillet, F76801, Saint-Etienne-du-Rouvray, France. Tel.: +33 2 32 95 52 72; fax: +33 2 32 95 52 86.

E-mail addresses: herold.dehling@rub.de (H. Dehling), olivier.durieu@univ-rouen.fr (O. Durieu), dalibor.volny@univ-rouen.fr (D. Volny).

0304-4149/$ - see front matter
doi:10.1016/j.spa.2009.07.003

$$U_n(t) := \sqrt{n}(F_n(t) - F(t)), \quad t \in \mathbb{R}.$$

The empirical process plays a prominent role in non-parametric statistical inference about the distribution function F. In all statistical applications, information about the distribution of the empirical process is needed.

For the case of i.i.d. observations, Donsker [8] proved in 1952 that the empirical process converges in distribution to a Brownian bridge process, thus confirming an earlier conjecture of Doob [9]. In 1968, Billingsley [2] extended Donsker's theorem to some weakly dependent processes, specifically to functionals of ϕ-mixing processes. One of the applications of Billingsley's theorem is to the empirical process of data generated by the continued fraction dynamical system $T : [0, 1] \to [0, 1]$, $T(x) := \frac{1}{x}$. Since 1968, many authors have studied the empirical process of weakly dependent data. Invariance principles for the empirical distribution of strong mixing random variables were proved in 1977 by Berkes and Philipp [1] and in 1980 for the multivariate case by Philipp and Pinzur [20]. Later, absolutely regular processes were studied by Doukhan et al. [10] and Borovkova et al. [3]. Many other weak dependence conditions have been studied, for examples by Doukhan and Louhichi [11], Prieur [21], Dedecker and Prieur [7], Wu and Shao [23] or Wu [22]. From the point of view of dynamical systems, an empirical process invariance principle for some uniformly expanding maps of the interval was proved by Collet et al. [5]. Recently, a similar result concerning intermittent maps was given by Dedecker [6]. Another one for ergodic torus automorphisms was proved by Durieu and Jouan [15].

Proofs of empirical process invariance principles usually consist of two parts, establishing finite dimensional convergence and tightness of the empirical process. Finite dimensional convergence, i.e. convergence in distribution of the sequence of vectors $(U_n(t_1), \ldots, U_n(t_k))_{n\geq 1}$, is an immediate consequence of the multivariate CLT for partial sums of the process

$$(1_{(-\infty,t_1]}(X_n), \ldots, 1_{(-\infty,t_k]}(X_n))_{n\geq 1}.$$

Tightness is far more difficult to establish. One ingredient is usually a probability bound on the increments of the empirical process

$$U_n(t) - U_n(s) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{1_{(s,t]}(X_i) - (F(t) - F(s))\},$$

for a fixed pair $s < t$. Such bounds can in the simplest approach be obtained from bounds on the fourth moments of $U_n(t) - U_n(s)$. Other results require higher order moment bounds or even exponential bounds.

The traditional approach to empirical process invariance principles, as outlined above, works well in situations when the sequence of indicator variables $(1_{(s,t]}(X_n))_{n\geq 0}$ inherits good properties from the original process $(X_n)_{n\geq 0}$. This holds, for example, when $(X_n)_{n\geq 0}$ is strong (uniform, beta) mixing, because then $(1_{(s,t]}(X_n))_{n\geq 0}$ has the same property. There are, however, situations where this is not the case or at least not easy to establish. For some kinds of Markov processes and dynamical systems (see e.g. [18]), one has good control over the properties of $(f(X_n))_{n\geq 0}$ when f is a Lipschitz function, but not for indicator functions. For example, Gouëzel [17] gave a uniformly expanding map of the interval which has a spectral gap on the space of Lipschitz functions but not on the space of bounded variation functions. In this paper, we develop an approach that is strictly based on properties of Lipschitz functions $f(X_i)$ of the original data. We make two basic assumptions, namely that the partial sums of Lipschitz functions satisfy the CLT and that a suitable fourth-moment bound is satisfied.

For our proof we develop a variant of the classical chaining technique that uses only Lipschitz functions at all stages of the chaining argument. We replace the usual finite dimensional convergence plus tightness approach by a method of approximation by a sequence of finite dimensional processes, which are different from the coordinate projections $(U_n(t_1), \ldots, U_n(t_k))$. We show convergence in distribution of the finite dimensional processes and prove that the finite dimensional process approximates the empirical process. In the final step, we use an improved version of a theorem of Billingsley [2] (see our Theorem 2) to establish convergence in distribution of the empirical process.

In the present paper, we make two assumptions concerning the process $(X_i)_{i\geq 0}$:

1. For any Lipschitz function f, the CLT holds, i.e.

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{f(X_i) - Ef(X_i)\} \xrightarrow{\mathcal{D}} N(0, \sigma^2), \tag{1}$$

where $N(0, \sigma^2)$ denotes a normal law with mean zero and variance

$$\sigma^2 = E(f(X_0) - Ef(X_0))^2 + 2\sum_{i=1}^{\infty} \mathrm{Cov}(f(X_0), f(X_i)).$$

2. A bound on the fourth central moments of partial sums of $(f(X_i))_{i\geq 0}$, f bounded Lipschitz with $E(f(X_0)) = 0$, of the type

$$E\left\{\sum_{i=1}^{n} f(X_i)\right\}^4 \leq C m_f^3 \Bigg(n\|f(X_0)\|_1 \log^{\alpha}(1 + \|f\|) + n^2 \|f(X_0)\|_1^2 \log^{\beta}(1 + \|f\|)\Bigg), \tag{2}$$

where C is some universal constant, α and β are some nonnegative integers,

$$\|f\| = \sup_x |f(x)| + \sup_{x\neq y} \frac{|f(x) - f(y)|}{|x - y|}$$

and

$$m_f = \max\{1, \sup_x |f(x)|\}.$$

Remark 1.1. These assumptions can be verified for a large class of Markov chains and dynamical systems. Concerning the CLT for Lipschitz functions, many results can be found in the literature; see e.g. [18] for the spectral gap method and [4] for the mixing approach. Durieu [14] has established fourth-moment bounds of the type (2) for Markov chains and dynamical systems under spectral properties. For more details and concrete examples see Section 4 of the present paper.

We shall assume some regularity for the distribution function of X_0. We define the modulus of continuity of a function $f : \mathbb{R} \longrightarrow \mathbb{R}$ by

$$\omega_f(\delta) = \sup\{|f(s) - f(t)| : s, t \in \mathbb{R}, |s - t| < \delta\}.$$

We can now state our main result.

Theorem 1. *Let $(X_i)_{i\geq 0}$ be an $\mathbb{R}$-valued stationary ergodic random process such that the conditions (1) and (2) hold. Assume that X_0 has a distribution function F satisfying the following condition:*

$$\omega_F(\delta) \leq D|\log(\delta)|^{-\gamma} \quad \text{for some } D > 0 \text{ and } \gamma > \max\left\{\frac{\alpha}{2}, \beta\right\}. \tag{3}$$

Then

$$(U_n(t))_{t\in\mathbb{R}} \xrightarrow{\mathcal{D}} (W(t))_{t\in\mathbb{R}},$$

where $W(t)$ is a mean-zero Gaussian process with covariances

$$\begin{aligned} EW(s)\cdot W(t) &= \mathrm{Cov}(1_{(-\infty,s]}(X_0), 1_{(-\infty,t]}(X_0)) + \sum_{k=1}^{\infty}\mathrm{Cov}(1_{(-\infty,s]}(X_0), 1_{(-\infty,t]}(X_k)) \\ &\quad + \sum_{k=1}^{\infty}\mathrm{Cov}(1_{(-\infty,s]}(X_k), 1_{(-\infty,t]}(X_0)). \end{aligned}$$

Further, almost surely, $(W(t))_{t\in\mathbb{R}}$ has continuous sample paths.

Remark 1.2. In particular, if X_0 has a Hölder continuous distribution function then (3) holds.

Remark 1.3. If the X_i's are i.i.d., $(W(t))_{t\in\mathbb{R}}$ is a Brownian bridge, but this is not always the case for dependent variables, as in [2] or [5].

In order to prove Theorem 1, we apply the following theorem, which is a stronger version of Theorem 4.2 of Billingsley [2] for the complete case. We do not need to assume a priori that $X^{(m)}$ has a limit in distribution.

Theorem 2. *Let (S,ρ) be a complete separable metric space and let X_n, $X_n^{(m)}$ and $X^{(m)}$, $n, m \geq 1$, be S-valued random variables satisfying*

$$X_n^{(m)} \xrightarrow{\mathcal{D}} X^{(m)} \quad \text{as } n\to\infty, \forall m \tag{4}$$

$$\lim_{m\to\infty}\limsup_{n\to\infty} P(\rho(X_n, X_n^{(m)}) \geq \varepsilon) = 0, \quad \forall\varepsilon > 0. \tag{5}$$

Then there exists an S-valued random variable X such that

$$X_n \xrightarrow{\mathcal{D}} X \quad \text{as } n\to\infty.$$

Moreover $X^{(m)} \xrightarrow{\mathcal{D}} X$ as $m\to\infty$.

Both theorems are proved in Sections 2 and 3.

2. Proof of Theorem 1

2.1. The bounded case

We first prove the result for bounded variables. Let $(X_i)_{i\geq 0}$ be a $[0,1]$-valued stationary ergodic random process such that (1)–(3) hold.

In our approach we work with Lipschitz continuous approximations to the indicator functions $1_{(-\infty,t]}(x)$. Given a partition

$$0 = t'_0 < \cdots < t'_m = 1$$

we define

$$t_j = F^{-1}(t'_j)$$

where F^{-1} is given by

$$F^{-1}(t) = \sup\{s \in [0, 1] : F(s) \le t\}.$$

Thus, by continuity of F, we have a partition

$$0 \le t_0 < \cdots < t_m = 1.$$

We introduce the functions $\varphi_j : [0, 1] \to \mathbb{R}$ by

$$\varphi_j(x) = \varphi\left(\frac{x - t_{j-1}}{t_{j-1} - t_{j-2}}\right), \quad \text{for } j = 2, \ldots, m$$

where

$$\varphi(x) = 1_{(-\infty,-1]}(x) - x1_{(-1,0]}(x) \tag{6}$$

and $\varphi_1 \equiv 0$.

The function φ_j will serve as a Lipschitz continuous approximation to the indicator function $1_{(-\infty,t_{j-1}]}(x)$. Note that $\varphi_j(x)$ depends on the partition, not only on the point t_{j-1}. We now define the process

$$\begin{aligned} F_n^{(m)}(t) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} 1_{[t_{j-1},t_j)}(t)\varphi_j(X_i) \\ &= \sum_{j=1}^{m}\left(\frac{1}{n}\sum_{i=1}^{n}\varphi_j(X_i)\right)1_{[t_{j-1},t_j)}(t). \end{aligned}$$

Note that $F_n^{(m)}(t)$ is a piecewise constant approximation to the empirical distribution function $F_n(t)$. For $t \in [t_{j-1}, t_j]$, we have the inequality

$$F_n(t_{j-2}) \le F_n^{(m)}(t) \le F_n(t_{j-1}).$$

We define further

$$F^{(m)}(t) = E\left(F_n^{(m)}(t)\right) = \sum_{j=1}^{m} E\left(\varphi_j(X_0)\right)1_{[t_{j-1},t_j)}(t),$$

and finally the centered and normalized process

$$U_n^{(m)}(t) = \sqrt{n}\left(F_n^{(m)}(t) - F^{(m)}(t)\right). \tag{7}$$

Our proof of Theorem 1 now consists of two parts, each of which will be formulated separately as a proposition below. The theorem will follow by application of Theorem 2, where (S, ρ) is the space of cadlag functions $D[0,1]$ provided with the Skorohod topology and the metric d_0; see [2, p. 113]. Note that $(D[0,1], d_0)$ is a complete separable metric space.

Proposition 2.1. *For any partition* $0 = t'_0 < \cdots < t'_m = 1$, *there exists a piecewise constant Gaussian process* $\left(W^{(m)}(t)\right)_{0\le t\le 1}$ *such that*

$$\left(U_n^{(m)}(t)\right)_{0\le t\le 1} \xrightarrow{\mathcal{D}} \left(W^{(m)}(t)\right)_{0\le t\le 1}.$$

The sample paths of the processes $\left(W^{(m)}(t)\right)_{0\le t\le 1}$ *are constant on each of the intervals* $[t_{j-1}, t_j)$, $1 \le j \le m$, *and* $W^{(m)}(0) = 0$. *The vector* $(W^{(m)}(t_1), \ldots, W^{(m)}(t_m))$ *has a multivariate normal distribution with mean zero and covariances*

$$\mathrm{Cov}(W^{(m)}(t_{i-1}), W^{(m)}(t_{j-1})) = \mathrm{Cov}(\varphi_i(X_0), \varphi_j(X_0)) + \sum_{k=1}^{\infty} \mathrm{Cov}(\varphi_i(X_0), \varphi_j(X_k)) + \sum_{k=1}^{\infty} \mathrm{Cov}(\varphi_i(X_k), \varphi_j(X_0)).$$

Proof. Using (1) and the Cramér–Wold device, we can show that for any Lipschitz functions $f_1, \ldots, f_k$, the multivariate CLT holds, i.e.

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{(f_1(X_i), \ldots, f_k(X_i)) - E(f_1(X_0), \ldots, f_k(X_0))\} \xrightarrow{\mathcal{D}} N(0, \Sigma_{f_1,\ldots,f_k}),$$

where $N(0, \Sigma_{f_1,\ldots,f_k})$ denotes a multivariate normal law with mean zero and covariance matrix

$$\Sigma_{f_1,\ldots,f_k} = (\sigma_{f_i,f_j})_{1\le i,j\le k}$$

where for any Lipschitz functions f, g we define

$$\sigma_{f,g} = \mathrm{Cov}(f(X_0), g(X_0)) + \sum_{k=1}^{\infty} \mathrm{Cov}(f(X_0), g(X_k)) + \sum_{k=1}^{\infty} \mathrm{Cov}(f(X_k), g(X_0)).$$

This result proves the proposition. $\square$

Proposition 2.2. *For any* $\varepsilon, \eta > 0$ *there exists a partition* $0 = t'_0 < \cdots < t'_m = 1$ *such that*

$$\limsup_{n\to\infty} P\left(\sup_{0\le t\le 1} \left|U_n(t) - U_n^{(m)}(t)\right| > \varepsilon\right) \le \eta.$$

Proof. By a variant of the well known chaining technique we will control

$$P\left(\sup_{0\le t\le 1} \left|U_n(t) - U_n^{(m)}(t)\right| \ge \varepsilon\right),$$

and then show that this probability can be made arbitrarily small by choosing a partition $0 = t'_0 < \cdots < t'_m = 1$ that is fine enough. From here on we assume that the partition $0 = t'_0 < \cdots < t'_m = 1$ is uniformly distributed (i.e. $t'_j = \frac{j}{m}$). Let $h = \frac{1}{m} = t'_j - t'_{j-1}$, for $j = 1, \ldots, m$.

On the interval $[t'_{j-1}, t'_j]$ we introduce a sequence of refining partitions

$$t'_{j-1} = s_0'^{(k)} < s_1'^{(k)} < \cdots < s_{2^k}'^{(k)} = t'_j$$

by

$$s_l'^{(k)} = t'_{j-1} + l \cdot \frac{h}{2^k}, \quad 0 \le l \le 2^k.$$

Let us define

$$s_l^{(k)} = F^{-1}(s_l'^{(k)}), \quad 0 \le l \le 2^k.$$

We now have partitions of $[t_{j-1}, t_j]$,

$$t_{j-1} = s_0^{(k)} < s_1^{(k)} < \cdots < s_{2^k}^{(k)} = t_j.$$

For convenience, we also consider the points

$$s_{-1}^{(k)} = F^{-1}\left(t'_{j-1} - \frac{h}{2^k}\right)$$

and the points

$$s_{2^k+1}^{(k)} = F^{-1}\left(t'_{j-1} + (2^k+1)\frac{h}{2^k}\right).$$

For any $t \in [t_{j-1}, t_j)$ and $k \ge 0$ we define the index

$$l(k,t) = \max\left\{l : s_l^{(k)} \le t\right\}.$$

In this way we obtain a chain

$$t_{j-1} = s_{l(0,t)}^{(0)} \le s_{l(1,t)}^{(1)} \le \cdots \le s_{l(k,t)}^{(k)} \le t \le s_{l(k,t)+1}^{(k)},$$

linking the left endpoint t_{j-1} to t. Note that for $t \in [t_{j-1}, t_j)$ we have by definition $U_n^{(m)}(t) = U_n^{(m)}(t_{j-1})$. We define the functions $\psi_l^{(k)}$, $k \ge 0$, $0 \le l \le 2^k$, by

$$\psi_l^{(k)}(x) = \varphi\left(\frac{x}{s_l^{(k)} - s_{l-1}^{(k)}}\right),$$

where φ is defined as in (6). Note that $\psi_{l(0,t)}^{(0)}(x - s_{l(0,t)}^{(0)}) = \varphi_j(x)$. To be consistent, in the case $j = 1$, we have to fix $\psi_0^{(k)} \equiv 0$, for all $k \ge 0$. We build a chain bridging the gap between

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} 1_{(-\infty,t]}(X_i)$$

and

$$F_n^{(m)}(t) = \frac{1}{n}\sum_{i=1}^{n} \varphi_j(X_i)$$

by the functions

$$\begin{aligned}\varphi_j(x) &= \psi^{(0)}_{l(0,t)}(x - s^{(0)}_{l(0,t)})\\ &\le \psi^{(1)}_{l(1,t)}(x - s^{(1)}_{l(1,t)})\\ &\le \cdots\\ &\le \psi^{(K)}_{l(K,t)}(x - s^{(K)}_{l(K,t)})\\ &\le 1_{(-\infty,t]}(x)\\ &\le \psi^{(K)}_{l(K,t)+2}(x - s^{(K)}_{l(K,t)+2}),\end{aligned}$$

where K is some integer to be chosen later. In this way we get

$$\begin{aligned}F_n(t) - F_n^{(m)}(t) &= \sum_{k=1}^{K} \frac{1}{n} \sum_{i=1}^{n} \left(\psi^{(k)}_{l(k,t)}(X_i - s^{(k)}_{l(k,t)}) - \psi^{(k-1)}_{l(k-1,t)}(X_i - s^{(k-1)}_{l(k-1,t)})\right)\\ &\quad + \frac{1}{n} \sum_{i=1}^{n} \left(1_{(-\infty,t]}(X_i) - \psi^{(K)}_{l(K,t)}(X_i - s^{(K)}_{l(K,t)})\right).\end{aligned} \tag{8}$$

Observe that by definition of $s^{(k)}_{l(k,t)}$ and of $\psi^{(K)}$,

$$\begin{aligned}0 &\le 1_{(-\infty,t]}(X_i) - \psi^{(K)}_{l(K,t)}(X_i - s^{(K)}_{l(K,t)})\\ &\le \psi^{(K)}_{l(K,t)+2}(X_i - s^{(K)}_{l(K,t)+2}) - \psi^{(K)}_{l(K,t)}(X_i - s^{(K)}_{l(K,t)}).\end{aligned}$$

From (8) we get by centering and normalization

$$\begin{aligned}U_n(t) - U_n^{(m)}(t) &= \sum_{k=1}^{K} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{\left(\psi^{(k)}_{l(k,t)}(X_i - s^{(k)}_{l(k,t)}) - E\psi^{(k)}_{l(k,t)}(X_i - s^{(k)}_{l(k,t)})\right)\right.\\ &\quad \left. - \left(\psi^{(k-1)}_{l(k-1,t)}(X_i - s^{(k-1)}_{l(k-1,t)}) - E\psi^{(k-1)}_{l(k-1,t)}(X_i - s^{(k-1)}_{l(k-1,t)})\right)\right\}\\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \left(1_{(-\infty,t]}(X_i) - F(t)\right)\right.\\ &\quad \left. - \left(\psi^{(K)}_{l(K,t)}(X_i - s^{(K)}_{l(K,t)}) - E\psi^{(K)}_{l(K,t)}(X_i - s^{(K)}_{l(K,t)})\right)\right\}.\end{aligned}$$

For the last term on the r.h.s. we have the following upper and lower bounds:

$$\begin{aligned}&\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{\left(1_{(-\infty,t]}(X_i) - F(t)\right) - \left(\psi^{(K)}_{l(K,t)}(X_i - s^{(K)}_{l(K,t)}) - E\psi^{(K)}_{l(K,t)}(X_i - s^{(K)}_{l(K,t)})\right)\right\}\\ &\le \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{\left(\psi^{(K)}_{l(K,t)+2}(X_i - s^{(K)}_{l(K,t)+2}) - E\psi^{(K)}_{l(K,t)+2}(X_i - s^{(K)}_{l(K,t)+2})\right)\right.\\ &\quad \left. - \left(\psi^{(K)}_{l(K,t)}(X_i - s^{(K)}_{l(K,t)}) - E\psi^{(K)}_{l(K,t)}(X_i - s^{(K)}_{l(K,t)})\right)\right\}\\ &\quad + \sqrt{n}\left(E\psi^{(K)}_{l(K,t)+2}(X_i - s^{(K)}_{l(K,t)+2}) - F(t)\right)\end{aligned}$$

and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\left(1_{(-\infty,t]}(X_i)-F(t)\right)-\left(\psi^{(K)}_{l(K,t)}(X_i-s^{(K)}_{l(K,t)})-E\psi^{(K)}_{l(K,t)}(X_i-s^{(K)}_{l(K,t)})\right)\right\}$$
$$\geq -\sqrt{n}\left(F(t)-E\psi^{(K)}_{l(K,t)}(X_i-s^{(K)}_{l(K,t)})\right).$$

Now choose $K=4+\left\lfloor \log\left(\frac{\sqrt{n}h}{\varepsilon}\right)\log^{-1}(2)\right\rfloor$ and note that

$$\frac{\varepsilon}{2^4}\leq\sqrt{n}\frac{h}{2^K}\leq\frac{\varepsilon}{2^3}$$

and thus

$$\sqrt{n}\left|E\psi^{(K)}_{l(K,t)+2}(X_i-s^{(K)}_{l(K,t)+2})-E\psi^{(K)}_{l(K,t)}(X_i-s^{(K)}_{l(K,t)})\right|$$
$$\leq\sqrt{n}\left|F(s^{(K)}_{l(K,t)+2})-F(s^{(K)}_{l(K,t)-1})\right|$$
$$\leq\frac{\varepsilon}{2}.$$

Thus we get for all $t\in[t_{j-1},t_j]$,

$$\left|U_n(t)-U_n^{(m)}(t)\right|\leq\sum_{k=1}^{K}\frac{1}{\sqrt{n}}\left|\sum_{i=1}^{n}\left\{\left(\psi^{(k)}_{l(k,t)}(X_i-s^{(k)}_{l(k,t)})-E\psi^{(k)}_{l(k,t)}(X_i-s^{(k)}_{l(k,t)})\right)\right.\right.$$
$$\left.\left.-\left(\psi^{(k-1)}_{l(k-1,t)}(X_i-s^{(k-1)}_{l(k-1,t)})-E\psi^{(k-1)}_{l(k-1,t)}(X_i-s^{(k-1)}_{l(k-1,t)})\right)\right\}\right|$$
$$+\frac{1}{\sqrt{n}}\left|\sum_{i=1}^{n}\left\{\left(\psi^{(K)}_{l(K,t)+2}(X_i-s^{(K)}_{l(K,t)+2})-E\psi^{(K)}_{l(K,t)+2}(X_i-s^{(K)}_{l(K,t)+2})\right)\right.\right.$$
$$\left.\left.-\left(\psi^{(K)}_{l(K,t)}(X_i-s^{(K)}_{l(K,t)})-E\psi^{(K)}_{l(K,t)}(X_i-s^{(K)}_{l(K,t)})\right)\right\}\right|+\frac{\varepsilon}{2}.$$

Note that by definition of $l(k,t)$ and of $s_l^{(k)}$, we have $s^{(k-1)}_{l(k-1,t)}\in\{s^{(k)}_{l(k,t)},s^{(k)}_{l(k,t)-1}\}$ and thus

$$l(k-1,t)=\left\lfloor\frac{l(k,t)}{2}\right\rfloor.$$

Therefore

$$\sup_{t_{j-1}\leq t\leq t_j}\left|U_n(t)-U_n^{(m)}(t)\right|\leq\sum_{k=1}^{K}\frac{1}{\sqrt{n}}\max_{0\leq l\leq 2^k-1}\left|\sum_{i=1}^{n}\left((\psi_l^{(k)}(X_i-s_l^{(k)})\right.\right.$$
$$\left.\left.-\ E\psi_l^{(k)}(X_i-s_l^{(k)}))-(\psi^{(k-1)}_{\lfloor\frac{l}{2}\rfloor}(X_i-s^{(k-1)}_{\lfloor\frac{l}{2}\rfloor})-E\psi^{(k-1)}_{\lfloor\frac{l}{2}\rfloor}(X_i-s^{(k-1)}_{\lfloor\frac{l}{2}\rfloor}))\right)\right|$$
$$+\frac{1}{\sqrt{n}}\max_{0\leq l\leq 2^K-1}\left|\sum_{i=1}^{n}\left((\psi^{(K)}_{l+2}(X_i-s^{(K)}_{l+2})-E\psi^{(K)}_{l+2}(X_i-s^{(K)}_{l+2}))\right.\right.$$
$$\left.\left.-(\psi_l^{(K)}(X_i-s_l^{(K)})-E\psi_l^{(K)}(X_i-s_l^{(K)}))\right)\right|+\frac{\varepsilon}{2}.$$

Now take $\varepsilon_k := \frac{\varepsilon}{4k(k+1)}$ and note that $\sum_{k=1}^{K} \varepsilon_k \leq \frac{\varepsilon}{4}$.
Then we obtain

$$\begin{aligned}
&P\left(\sup_{t_{j-1}\le t\le t_j}\left|U_n(t)-U_n^{(m)}(t)\right|\ge\varepsilon\right)\\
&\le\sum_{k=1}^{K}\sum_{l=0}^{2^k-1}P\left(\frac{1}{\sqrt{n}}\left|\sum_{i=1}^{n}\left\{\left(\psi_l^{(k)}(X_i-s_l^{(k)})-E\psi_l^{(k)}(X_i-s_l^{(k)})\right)\right.\right.\right.\\
&\quad\left.\left.\left.-\left(\psi_{\lfloor\frac{l}{2}\rfloor}^{(k-1)}(X_i-s_{\lfloor\frac{l}{2}\rfloor}^{(k-1)})-E\psi_{\lfloor\frac{l}{2}\rfloor}^{(k-1)}(X_i-s_{\lfloor\frac{l}{2}\rfloor}^{(k-1)})\right)\right\}\right|\ge\varepsilon_k\right)\\
&\quad+\sum_{l=0}^{2^K-1}P\left(\frac{1}{\sqrt{n}}\left|\sum_{i=1}^{n}\left\{\left(\psi_{l+2}^{(K)}(X_i-s_{l+2}^{(K)})-E\psi_{l+2}^{(K)}(X_i-s_{l+2}^{(K)})\right)\right.\right.\right.\\
&\quad\left.\left.\left.-\left(\psi_l^{(K)}(X_i-s_l^{(K)})-E\psi_l^{(K)}(X_i-s_l^{(K)})\right)\right\}\right|\ge\frac{\varepsilon}{4}\right).
\end{aligned}$$

At this point we use Markov's inequality together with the fourth-moment bound (2):

$$\begin{aligned}
&P\left(\sup_{t_{j-1}\le t\le t_j}\left|U_n(t)-U_n^{(m)}(t)\right|\ge\varepsilon\right)\\
&\le C\sum_{k=1}^{K}\sum_{l=0}^{2^k-1}\left\{\frac{1}{n\varepsilon_k^4}\left\|\psi_l^{(k)}(X_0-s_l^{(k)})-\psi_{\lfloor\frac{l}{2}\rfloor}^{(k-1)}(X_0-s_{\lfloor\frac{l}{2}\rfloor}^{(k-1)})\right\|_1\right.\\
&\times\log^{\alpha}\left(1+\left\|\psi_l^{(k)}-\psi_{\lfloor\frac{l}{2}\rfloor}^{(k-1)}\right\|\right)+\frac{1}{\varepsilon_k^4}\left\|\psi_l^{(k)}(X_0-s_l^{(k)})-\psi_{\lfloor\frac{l}{2}\rfloor}^{(k-1)}(X_0-s_{\lfloor\frac{l}{2}\rfloor}^{(k-1)})\right\|_1^2\\
&\left.\times\log^{\beta}\left(1+\left\|\psi_l^{(k)}-\psi_{\lfloor\frac{l}{2}\rfloor}^{(k-1)}\right\|\right)\right\}\\
&+C\sum_{l=0}^{2^k-1}\left\{\frac{4^4}{n\varepsilon^4}\left\|\psi_{l+2}^{(K)}(X_0-s_{l+2}^{(K)})-\psi_l^{(K)}(X_0-s_l^{(K)})\right\|_1\log^{\alpha}\left(1+\left\|\psi_{l+2}^{(K)}-\psi_l^{(K)}\right\|\right)\right.\\
&\left.+\frac{4^4}{\varepsilon^4}\left\|\psi_{l+2}^{(K)}(X_0-s_{l+2}^{(K)})-\psi_l^{(K)}(X_0-s_l^{(K)})\right\|_1^2\log^{\beta}\left(1+\left\|\psi_{l+2}^{(K)}-\psi_l^{(K)}\right\|\right)\right\}.
\end{aligned}$$

Note that

$$\begin{aligned}
\left\|\psi_l^{(k)}(X_0-s_l^{(k)})-\psi_{\lfloor\frac{l}{2}\rfloor}^{(k-1)}(X_0-s_{\lfloor\frac{l}{2}\rfloor}^{(k-1)})\right\|_1 &\le\left|F(s_l^{(k)})-F(s_{\lfloor\frac{l}{2}\rfloor-1}^{(k-1)})\right|\\
&\le\left|F(s_l^{(k)})-F(s_{l-3}^{(k)})\right|\\
&=\frac{3h}{2^k}
\end{aligned}$$

and

$$\begin{aligned}
\left\|\psi_{l+2}^{(K)}(X_0-s_{l+2}^{(K)})-\psi_l^{(K)}(X_0-s_l^{(K)})\right\|_1 &\le\left|F(s_{l+2}^{(K)})-F(s_{l-1}^{(K)})\right|\\
&=\frac{3h}{2^K}.
\end{aligned}$$

If (3) is satisfied,

$$\begin{aligned}\left\|\psi_l^{(k)}\right\| &\le 1+\left[\inf\left\{s>0:\forall t,\, F(t+s)-F(t)\ge\frac{h}{2^k}\right\}\right]^{-1}\\ &\le 1+\left[\inf\left\{s>0:D|\log(s)|^{-\gamma}\ge\frac{h}{2^k}\right\}\right]^{-1}\\ &= 1+\exp\left(\left(\frac{D2^k}{h}\right)^{\frac{1}{\gamma}}\right).\end{aligned}$$

Thus we have

$$\begin{aligned}&P\left(\sup_{t_{j-1}\le t\le t_j}\left|U_n(t)-U_n(t_j)\right|\ge\varepsilon\right)\\ &\le 4^4C\sum_{k=1}^{K}2^k\frac{(k(k+1))^4}{\varepsilon^4}\frac{1}{n}\frac{3h}{2^k}\log^{\alpha}\left(2+\exp\left(\left(\frac{D2^k}{h}\right)^{\frac{1}{\gamma}}\right)\right)\\ &\quad+4^4C\sum_{k=1}^{K}2^k\frac{(k(k+1))^4}{\varepsilon^4}\frac{(3h)^2}{2^{2k}}\log^{\beta}\left(2+\exp\left(\left(\frac{D2^k}{h}\right)^{\frac{1}{\gamma}}\right)\right)\\ &\quad+4^4C2^K\frac{1}{\varepsilon^4}\frac{1}{n}\frac{3h}{2^K}\log^{\alpha}\left(2+\exp\left(\left(\frac{D2^k}{h}\right)^{\frac{1}{\gamma}}\right)\right)\\ &\quad+4^4C2^K\frac{1}{\varepsilon^4}\frac{(3h)^2}{2^{2K}}\log^{\beta}\left(2+\exp\left(\left(\frac{D2^k}{h}\right)^{\frac{1}{\gamma}}\right)\right)\\ &\le\frac{1}{n}\frac{C'}{\varepsilon^4}\sum_{k=1}^{K}k^8h\left(\frac{D2^k}{h}\right)^{\frac{\alpha}{\gamma}}+\frac{C'}{\varepsilon^4}\sum_{k=1}^{K}\frac{k^8}{2^k}h^2\left(\frac{D2^k}{h}\right)^{\frac{\beta}{\gamma}}\\ &\le D^{\frac{\alpha}{\gamma}}\frac{1}{n}\frac{C'}{\varepsilon^4}h\left(\frac{2^K}{h}\right)^{\frac{\alpha}{\gamma}}\sum_{k=1}^{K}k^8+D^{\frac{\beta}{\gamma}}\frac{C'}{\varepsilon^4}h^{2-\frac{\beta}{\gamma}}\sum_{k=1}^{\infty}k^82^{k(\frac{\beta}{\gamma}-1)}\\ &\le\frac{h}{n}\frac{C''}{\varepsilon^4}\left(\frac{\sqrt{n}}{\varepsilon}\right)^{\frac{\alpha}{\gamma}}K^9+\frac{C''}{\varepsilon^4}h^{2-\frac{\beta}{\gamma}}\end{aligned}$$

where C' and C'' are some constants. In addition, we have used convergence of the series $\sum_{k=1}^{\infty}k^82^{k(\frac{\beta}{\gamma}-1)}$.

Finally, using $mh=1$,

$$\begin{aligned}P\left(\sup_{0\le t\le 1}\left|U_n(t)-U_n^{(m)}(t)\right|\ge\varepsilon\right)&\le\sum_{j=1}^{m}P\left(\sup_{t_{j-1}\le t\le t_j}\left|U_n(t)-U_n^{(m)}(t)\right|\ge\varepsilon\right)\\ &\le mhn^{\frac{\alpha}{2\gamma}-1}\frac{C''}{\varepsilon^{4+\frac{\alpha}{\gamma}}}K^9+m\frac{C''}{\varepsilon^4}h^{2-\frac{\beta}{\gamma}}\\ &\le n^{\frac{\alpha}{2\gamma}-1}\frac{C''}{\varepsilon^{4+\frac{\alpha}{\gamma}}}\left(4+\log\frac{\sqrt{n}h}{\varepsilon}\right)^9+\frac{C''}{\varepsilon^4}h^{1-\frac{\beta}{\gamma}}.\end{aligned}$$

Now, the first of the two final summands converges to zero as $n \to \infty$. The second can be made arbitrarily small by choosing a partition that is fine enough (i.e. h small). $\square$

We used a technique different from the usual finite dimensional convergence plus tightness. Of course, since the weak convergence implies the finite dimensional convergence and the tightness, these two properties are satisfied. Nevertheless, we can also deduce a tightness criterion implying that, almost surely, the limit process has continuous sample paths (see [2, Theorem 15.5]).

Proposition 2.3. *For all $\varepsilon, \eta > 0$, there exist $\delta > 0$ and $N \geq 0$ such that for all $n \geq N$,*

$$P\left(\sup_{|t-s|<\delta} |U_n(t) - U_n(s)| \geq \varepsilon\right) \leq \eta.$$

In particular, $P(W \in C(\mathbb{R})) = 1$.

Proof. Let $\varepsilon > 0$ and $\eta > 0$. Let m be an integer such that

$$\frac{C}{\varepsilon^4} \frac{D^{\frac{\beta}{\gamma}}}{m^{1+\frac{\beta}{\gamma}}} < \frac{\eta}{4} \tag{9}$$

and consider the regular partition of [0, 1] with mesh $\frac{1}{m}$.

By Proposition 2.2, there exists $N \geq 0$ such that for all $n \geq N$,

$$P\left(\sup_{0\leq t\leq 1} \left|U_n(t) - U_n^{(m)}(t)\right| \geq \frac{\varepsilon}{3}\right) \leq \frac{\eta}{4}.$$

Let us have $\delta > 0$ such that $\delta < \frac{1}{m}$. Then, for all $n \geq N$,

$$\begin{aligned} P\left(\sup_{|t-s|<\delta} |U_n(t) - U_n(s)| \geq \varepsilon\right) &\leq 2P\left(\sup_{0\leq t\leq 1} |U_n(t) - U_n^m(t)| \geq \frac{\varepsilon}{3}\right) \\ &\quad + P\left(\sup_{|t-s|<\delta} |U_n^m(t) - U_n^m(s)| \geq \frac{\varepsilon}{3}\right) \\ &\leq \frac{\eta}{2} + P\left(\sup_{|t-s|<\delta} |U_n^m(t) - U_n^m(s)| \geq \frac{\varepsilon}{3}\right). \end{aligned}$$

We recall, as $t_j = F^{-1}(t'_j) = F^{-1}(\frac{j}{m})$, that

$$\|\varphi_j(X_0) - \varphi_{j+1}(X_0)\|_1 \leq P\left(t_{j-2} \leq X_0 \leq t_j\right) \leq \frac{2}{m},$$

$$\|\varphi_j\| \leq 1 + \exp\left(\left(\frac{D}{m}\right)^{\frac{1}{\gamma}}\right).$$

Thus, by the fourth-moment bound (2),

$$P\left(\sup_{|t-s|<\delta} |U_n^m(t) - U_n^m(s)| \geq \frac{\varepsilon}{3}\right) \leq \frac{C}{n\varepsilon^4}\left(\frac{D}{m}\right)^{\frac{\alpha}{\gamma}} + \frac{C}{\varepsilon^4}\frac{D^{\frac{\beta}{\gamma}}}{m^{1+\frac{\beta}{\gamma}}}.$$

Now there exists $N' \geq N$ such that

$$\frac{C}{n\varepsilon^4}\left(\frac{D}{m}\right)^{\frac{\alpha}{\gamma}} \leq \frac{\eta}{4}.$$

Finally, by (9),

$$P\left(\sup_{|t-s|<\delta} |U_n(t) - U_n(s)| \geq \varepsilon\right) \leq \eta. \quad \square$$

2.2. *The unbounded case*

Let $(X_i)_{i\geq 0}$ be an $\mathbb{R}$-valued stationary ergodic random process such that (1)–(3) hold. We will show that it can be reduced to the case of bounded variables.

For all $x < y \in \mathbb{R}$, we say that the closed interval $[x, y]$ is a 'bad' interval (for F) if

$$F(y) - F(x) \geq y - x.$$

We say that $[x, y]$ is a maximal 'bad' interval (for F) if for all 'bad' intervals $[a, b]$, we have $[a, b] \subset [x, y]$ or $[a, b] \cap [x, y] = \emptyset$.

We denote by $I^{\max}$ the set of all maximal 'bad' intervals.

Lemma 2.4. (i) *The Lebesgue measure of*

$$I := \bigcup_{[x,y]\in I^{\max}} [x, y]$$

is smaller than 1.

(ii) *For all* $[x, y] \in I^{\max}$, *we have*

$$F(y) - F(x) = y - x.$$

Proof. Because F is non-decreasing and takes values in [0, 1], the first assertion is clear.

If for $x < y$, $F(y) - F(x) > y - x$, then there exists $\varepsilon > 0$ such that

$$F(y) - F(x) > y - x + \varepsilon.$$

Thus, for all $z > y$ such that $z - y \leq \varepsilon$, by monotonicity of F, we have

$$\begin{aligned} F(z) - F(x) &\geq F(y) - F(x) \\ &> y - x + \varepsilon \\ &\geq z - x \end{aligned}$$

and then $[x, y]$ is not maximal. $\square$

We define the function g from $\mathbb{R}$ to $]0, 1[$ by

$$\text{for all } [x, y] \in I^{\max},\ \text{for all } t \in [x, y],\ g(t) := F(x) + t - x$$

and

$$\text{for all } t \notin I, \quad g(t) := F(t).$$

Then g is a 1-Lipschitz function.

We define the [0, 1]-valued stationary ergodic random process $(Y_i)_{i\geq 0}$ by

$$Y_i = g(X_i), \quad i \geq 0.$$

Since g is Lipschitz, $(Y_i)_{i\geq 0}$ satisfies (1) and (2).

We also have

$$G(t) := P(Y_0 \leq t) = F \circ g^{-1}(t)$$

where

$$g^{-1}(t) = \sup\{s \in \mathbb{R} : g(s) \leq t\}.$$

Clearly, G is the identity on $g(\mathbb{R} \setminus I)$. Further, for all $[x, y] \in I^{\max}$, the graph of G on $g([x, y])$ is the graph of F on $[x, y]$ and the Lebesgue measure of $g([x, y])$ is equal to the Lebesgue measure of $[x, y]$. Then

$$\omega_G(\delta) \leq \max\{\omega_F(\delta), \delta\}$$

and (3) holds.

We define the associated distribution functions and empirical processes

$$\begin{aligned}
F_n(t) &:= \frac{1}{n}\sum_{i=1}^{n} 1_{(-\infty,t]}(X_i), \quad t \in \mathbb{R}, \\
U_n(t) &:= \sqrt{n}(F_n(t) - F(t)), \quad t \in \mathbb{R}, \\
G_n(t) &:= \frac{1}{n}\sum_{i=1}^{n} 1_{[0,t]}(Y_i), \quad 0 \leq t \leq 1, \\
V_n(t) &:= \sqrt{n}(G_n(t) - G(t)), \quad 0 \leq t \leq 1.
\end{aligned}$$

We have

$$U_n(t) = V_n(g(t)), \quad t \in \mathbb{R}.$$

By the theorem for bounded variables (Section 2.1),

$$(V_n(t))_{0\leq t\leq 1} \xrightarrow{\mathcal{D}} (V(t))_{0\leq t\leq 1},$$

where $V(t)$ is a mean-zero Gaussian process such that $P(V \in C[0, 1]) = 1$.

Applying Theorem 5.1 of Billingsley [2] with

$$\begin{aligned}
h : D[0, 1] &\longrightarrow D(\mathbb{R}) \\
x &\mapsto x \circ g,
\end{aligned}$$

we get the weak convergence of $(U_n(t))_{t\in\mathbb{R}}$ to a Gaussian process

$$(W(t))_{t\in\mathbb{R}} = (V \circ g(t))_{t\in\mathbb{R}}$$

such that $P(W \in C(\mathbb{R})) = 1$.

3. Proof of Theorem 2

Lemma 3.1. *Let (X, d) be a complete metric space and let $x_n, x_n^{(m)}, x^{(m)} \in X$, $n \geq 1, m \geq 1$, be given with the properties*

$$\lim_{n\to\infty} d(x_n^{(m)}, x^{(m)}) = 0 \quad \forall m \tag{10}$$

$$\lim_{m\to\infty} \limsup_{n\to\infty} d(x_n, x_n^{(m)}) = 0. \tag{11}$$

Then $x := \lim_{m\to\infty} x^{(m)}$ *exists and*

$$\lim_{n\to\infty} d(x_n, x) = 0.$$

The proof is left to the reader.

Proof of Theorem 2. Let $\mu_n, \mu_n^{(m)}$ and $\mu^{(m)}$ denote the distributions of the random variables $X_n, X_n^{(m)}$ and $X^{(m)}$ respectively. These are elements of $M_1(S)$, the space of probability measures on S. We consider the Prohorov metric d on $M_1(S)$, defined by

$$d(\mu, \upsilon) = \inf\left\{\varepsilon > 0 : \mu(A) \le \upsilon(A^{\varepsilon}) + \varepsilon \forall A \subset S \text{ measurable}\right\}.$$

Note that $(M_1(S), d)$ is a complete metric space. If Y, Z are two S-valued random variables with distributions P_Y, P_Z, satisfying

$$P(\rho(Y, Z) \ge \varepsilon) \le \varepsilon,$$

then $d(P_Y, P_Z) \le \varepsilon$. Moreover d metrizes the topology of weak convergence, i.e. $\mu_n \to \mu$ if and only if $d(\mu_n, \mu) \to 0$. We now apply Lemma 3.1 to $\mu_n, \mu_n^{(m)}, \mu^{(m)}$. Note that condition (10) is a direct consequence of (4) and that (11) can be deduced from (5). We get a probability distribution μ on S such that if X is an S-valued random variable with distribution μ then $X^{(m)} \xrightarrow{\mathcal{D}} X$ as $m \to \infty$ and $X_n \xrightarrow{\mathcal{D}} X$ as $n \to \infty$. $\square$

4. Examples

According to Durieu [14] the fourth-moment bound (2) holds for Markov chains and dynamical systems under some assumptions on the Markov transition operator or the Perron–Frobenius operator.

Let (E, d) be a separable metric space and $(X_k)_{k\ge0}$ be an E-valued Markov chain with transition operator Q and invariant measure ν. Denote by $\mathcal{L}$ the space of all bounded Lipschitz continuous functions from E to $\mathbb{R}$ equipped with the norm defined in (2). We say that the Markov chain $(X_k)_{k\ge0}$ is $\mathcal{L}$-geometrically ergodic if there exist $C > 0$ and $0 < \theta < 1$ such that for all $f \in \mathcal{L}$,

$$\|Q^k f - \Pi f\| \le C\theta^k \|f\|, \tag{12}$$

where $\Pi f = E_\nu f(X_0)$. This condition corresponds to the fact that the Markov operator is quasi-compact on the space $\mathcal{L}$ with 1 as the only eigenvalue of modulus 1 and simple (see [18]). Since $\mathcal{L} \hookrightarrow L^\infty$, the following result is a special case of Corollary 1 of [14].

Proposition 4.1. *If* $(X_n)_{n\ge0}$ *is an* $\mathcal{L}$*-geometrically ergodic Markov chain then* (2) *holds for all* $f \in \mathcal{L}$ *such that* $Ef(X_0) = 0$, *with* $\alpha = 3$ *and* $\beta = 2$.

The same is true for dynamical systems whose Perron–Frobenius operators satisfy (12). This gives a large class of examples where our result applies.

4.1. Linear processes

Let $(A, \|.\|_A)$ be a separable Banach space and $\mathcal{A}$ its Borel sigma algebra. Denote by C some positive constant which may vary from line to line. Let $(a_i)_{i\geq 0}$ be a sequence of linear forms on A such that there exists $0 < \theta < 1$ such that

$$|a_i| \leq C\theta^i, \tag{13}$$

where $|a_i| = \sup_{\|x\|_A \leq 1} |a_i(x)|$. Let $(e_i)_{i\in\mathbb{Z}}$ be an i.i.d. bounded random sequence with values in a compact subset $B \subset A$ and marginal distribution μ. We define the real-valued linear process $(X_k)_{k\geq 0}$ by

$$X_k = \sum_{i\geq 0} a_i(e_{k-i}), \quad k \geq 0.$$

Several results have already been established for empirical processes of linear processes (see [12,22,7]). Here, the assumption on the $(a_i)_{i\geq 0}$ is stronger than in the aforementioned papers, but there will be no assumption on the distribution of the e_i's and the assumption on the distribution function of X_0 will be weaker. Note that $(X_k)_{k\geq 0}$ can be viewed as a functional of a Markov chain.

Let $Y_k = (e_k, e_{k-1}, \ldots)$; then $(Y_k)_{k\geq 0}$ is a stationary Markov chain on $B^{\mathbb{N}}$ (with stationary measure $\mu^{\otimes\mathbb{N}}$) and $X_k = \Phi(Y_k)$ where

$$\Phi : B^{\mathbb{N}} \longrightarrow \mathbb{R}, \quad \Phi(x_0, x_1, \ldots) = \sum_{i\geq 0} a_i(x_i).$$

Let Q be the Markov transition operator of the chain. On $B^{\mathbb{N}}$, we define a metric d by

$$d(x, y) = \sum_{i\geq 0} \theta^i \|x_i - y_i\|_A$$

where $x = (x_i)_{i\geq 0}$ and $y = (y_i)_{i\geq 0}$. As B is compact, then $(B^{\mathbb{N}}, d)$ is also compact. Let us denote by $\mathcal{L}$ the space of all Lipschitz functions from $B^{\mathbb{N}}$ to $\mathbb{R}$ provided with the norm $\|.\|$ defined by

$$\|f\| = \sup_{x\in B^{\mathbb{N}}} |f(x)| + \sup_{x\neq y} \frac{|f(x) - f(y)|}{d(x, y)}.$$

For all $f \in \mathcal{L}$ and for all $x = (x_i)_{i\geq 0}$ and $y = (y_i)_{i\geq 0} \in B^{\mathbb{N}}$, we have

$$\begin{aligned} |Q^k f(x) - Q^k f(y)| &= |E(f(Y_k)|Y_0 = x) - E(f(Y_k)|Y_0 = y)| \\ &= |E(f(e_k, \ldots, e_1, x_0, \ldots)) - E(f(e_k, \ldots, e_1, y_0, \ldots))| \\ &\leq \|f\| E\{d((e_k, \ldots, e_1, x_0, \ldots), (e_k, \ldots, e_1, y_0, \ldots))\} \\ &= C\theta^k \|f\| d(x, y), \end{aligned}$$

and

$$\begin{aligned} |Q^k f(x) - Ef(Y_0)| &= |E(f(Y_k)|Y_0 = x) - Ef(Y_k)| \\ &\leq E|f(e_k, e_{k-1}, \ldots, e_1, x_0, \ldots) - f(e_k, e_{k-1}, \ldots)| \\ &\leq C\theta^k \|f\| E\{d(x, Y_0)\}. \end{aligned}$$

Then, we have for all $f \in \mathcal{L}$,

$$\|Q^k f - E(f(Y_0))\| \leq C\theta^k \|f\|.$$

Since $(\mathcal{L}, \|.\|) \subset (L^\infty(\mu^{\otimes \mathbb{N}}), \|.\|_\infty)$, by Proposition 4.1, $(Y_k)_{k\geq 0}$ satisfies the fourth-moment bound (2) with $\alpha = 3$ and $\beta = 2$ for all bounded Lipschitz functions such that $Ef(Y_0) = 0$. And then the sequence $\sum_{i=0}^n Q^i f(Y_0)$ converges in $L^2(\mu^{\otimes \mathbb{N}})$ for all $f \in \mathcal{L}$ such that $Ef(Y_0) = 0$. By Gordin's theorem (see [16]), the CLT (1) is satisfied. Clearly, the function Φ is a Lipschitz continuous function on $B^{\mathbb{N}}$, and for all Lipschitz functions $g : \mathbb{R} \longrightarrow \mathbb{R}$, $g \circ \Phi$ is also a Lipschitz continuous function on $B^{\mathbb{N}}$. Thus conditions (1) and (2) hold for the process $(X_k)_{k\geq 0}$, for all Lipschitz functions on $\mathbb{R}$. Then Theorem 1 applies and we have:

Corollary 4.2. *Let $(X_k)_{k\geq 0}$ be a real linear process defined by a sequence of linear forms $(a_i)_{i\geq 0}$ and a sequence of i.i.d. bounded random variables $(e_i)_{i\in\mathbb{Z}}$, both on a measurable Banach space A. Assume (a_i) satisfies (13) and the distribution function F of X_0 satisfies*

$$\omega_F(\delta) \leq D|\log(\delta)|^{-\gamma} \quad \textit{for some } D > 0 \textit{ and } \gamma > 2.$$

Then $(U_n(t))_{t\in\mathbb{R}}$ converges in distribution to a mean-zero Gaussian process.

In the paper by Dedecker and Prieur [7], Corollary 1, X_0 has a bounded density. Here, the existence of a density is not needed. Our result is comparable to a result of Wu and Shao [23].

For a concrete example, consider $A = \{0, 1\}$, $a_i = \frac{2}{3^i}$, $i \geq 0$, and $e_k = 0$ or 1 with probability $\frac{1}{2}$, $k \in \mathbb{Z}$. Then

$$X_k = 2\sum_{i\geq 0} \frac{e_{k-i}}{3^i}, \quad k \geq 0$$

is a stationary process with values in $[0, 1]$ and the common distribution function of all the X_k is the Cantor function, which is not absolutely continuous but $\frac{\log 2}{\log 3}$-Hölder continuous (see [13]).

4.2. *Expanding maps*

In the setting of expanding maps of the interval, empirical process invariance principles have been established in [5,7] for classes of Lasota–Yorke transformations. For these maps, the transfer operator has a spectral gap on the space BV of bounded variation functions. According to Gouëzel [17], there exist some uniformly expanding maps of the interval for which the transfer operator does not act continuously on the space BV, but admits a spectral gap on the space of Lipschitz functions. The example given by Gouëzel is a transformation of the interval $[0, 1)$. Let $(a_n)_{n\geq 1}$ be a sequence of positive numbers with $\sum a_n < \frac{1}{4}$ and let $N > 0$ be an integer. Denote by I_n the subintervals $[4\sum_{i=1}^{n-1} a_i, 4\sum_{i=1}^{n} a_i)$. We decompose I_n into two subintervals of length $2a_n$ denoted by $I_n^{(1)}$ and $I_n^{(2)}$. We can find a map v_n (resp. w_n) on $[0, 1)$ with image $I_n^{(1)}$ (resp. $I_n^{(2)}$) such that the derivative at a point x is equal to $a_n(1 + 2\cos^2(2\pi n^4 x))$ (resp. $a_n(1+2\sin^2(2\pi n^4 x))$). The map T is defined on I_n in such a way that v_n and w_n are two inverse branches of it. There remains the interval $[4\sum_{i=1}^{\infty} a_i, 1)$ that we subdivide into N subintervals of equal length. T is defined as an affine transformation on each of these subintervals onto $[0, 1)$.

Theorem 3 (*Gouëzel [17]*). *If $a_n = \frac{1}{100n^3}$ and $N = 4$, then T is a Lebesgue measure preserving transformation and its associated transfer operator has a spectral gap on the space of Lipschitz* *functions with a simple eigenvalue at* 1 *and no other eigenvalue of modulus* 1. *Further, the transfer operator does not act continuously on BV.*

Here, we cannot directly apply the above mentioned results which require the spectral property of the transfer operator on BV. But using its spectral property on $\mathcal{L}$, we can deduce the following result.

Corollary 4.3. *The empirical process associated with Gouëzel's example satisfies the invariance principle of Theorem* 1.

Proof. The CLT (1) is given by the spectral gap of the transfer operator on $\mathcal{L}$. We obtain the fourth-moment bound (2) as a consequence of Theorem 2 of Durieu [14]. Indeed, Gouëzel's theorem entails that the required assumptions for this theorem (conditions (i) and (ii), page 1093 of [14]) hold for the space $\mathcal{L}$. Finally, the Lebesgue measure clearly satisfies (3). Hence, Theorem 1 applies. □

4.3. Further applications

Durieu [14] has also given fourth-moment bounds for subshifts of finite type, using the Ruelle–Perron–Frobenius theorem, as in [19]. Our result thus also applies here.

Another application concerns random iterative Lipschitz models, as considered in [14]. Let $g : E \times \mathbb{R} \longrightarrow E$ be a measurable function where $E \subset \mathbb{R}$ is some compact interval and let $(Y_n)_{n\geq 1}$ be an $\mathbb{R}$-valued i.i.d. process. Let X_0 be an E-valued random variable independent of $(Y_n)_{n\geq 1}$. Define the Markov chain $(X_n)_{n\in\mathbb{N}}$ by

$$X_n = g(X_{n-1}, Y_n), \quad n \geq 1.$$

Such models are studied, e.g., in nonlinear time series analysis.

Assume that for all $y \in \mathbb{R}$, $g(., y)$ is Lipschitz. Define the Lipschitz constant

$$K(y) := \sup_{x,x'\in E, x\neq x'} \frac{|g(x, y) - g(x', y)|}{|x - x'|}$$

and suppose that $EK(Y_1) < 1$. Let us have $x, x' \in E$ and let $(X_n)_{n\in\mathbb{N}}$ and $(X'_n)_{n\in\mathbb{N}}$ be defined by $X_0 = x$, $X'_0 = x'$, $X_n = g(X_{n-1}, Y_n)$ and $X'_n = g(X'_{n-1}, Y_n)$. Then, by induction, $|X_n - X'_n| \leq |x - x'| \prod_{i=1}^n K(Y_i)$ and thus

$$E\frac{|X_n - X'_n|}{|x - x'|} < (EK(Y_1))^n.$$

Hence the random map $g(., Y_1)$ is contracting in the sense of Definition X.1 of [18]. By Theorem X.3 of [18], there exists a unique invariant probability measure μ and moreover the Markov chain $(X_n)_{n\in\mathbb{N}}$ is $\mathcal{L}$-geometrically ergodic as defined in (12). By Proposition 4.1, if X_0 is μ-distributed, the fourth-moment bound (2) holds. The CLT (1) also follows from the result of [18]. If μ satisfies (3), the empirical process invariance principle holds.

This example has been investigated before by Dedecker and Prieur [7], with different techniques and under different conditions.

Remark 4.4. Condition (3) can be derived from suitable uniform conditions on the modulus of continuity of $g(x, Y)$, $x \in E$. Define

$$F^x(u) := P(g(x, Y) \leq u)$$

and assume that

$$\omega(\delta) := \sup_{x\in E}\ \sup_{u,v\in E, |u-v|\le\delta} |F^x(v) - F^x(u)| = O(|\log\delta|^{-\eta}).$$

If μ denotes the invariant measure, we have for $u, v \in E$

$$\mu((u,v]) = \int P(u < g(x,Y) \le v)\mathrm{d}\mu(x) = \int (F^x(v) - F^x(u))\mathrm{d}\mu(x).$$

Let X_0 be distributed according to μ, and denote by $F(u) := P(X_0 \le u)$ the distribution function. Then we get for $u, v \in E$, $|v - u| \le \delta$,

$$|F(v) - F(u)| \le \int |F^x(v) - F^x(u)|\mathrm{d}\mu(x) = O(|\log\delta|^{-\eta}).$$

Acknowledgements

We are grateful to Loïc Hervé for several lectures introducing us to the spectral gap method, and to Jérôme Dedecker for his critical comments on an earlier version of this paper. We also thank the referees and the associate editor for their helpful comments.

References

[1] István Berkes, Walter Philipp, An almost sure invariance principle for the empirical distribution function of mixing random variables, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 41 (2) (1977/78) 115–137.

[2] Patrick Billingsley, Convergence of Probability Measures, John Wiley & Sons Inc., New York, 1968.

[3] Svetlana Borovkova, Robert Burton, Herold Dehling, Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation, Trans Amer. Math. Soc. 353 (11) (2001) 4261–4318. (electronic).

[4] Richard C. Bradley, Introduction to Strong Mixing Conditions, vols. 1–3, Kendrick Press, Heber City, UT, 2007.

[5] P. Collet, S. Martinez, B. Schmitt, Asymptotic distribution of tests for expanding maps of the interval, Ergodic Theory Dynam. Systems 24 (3) (2004) 707–722.

[6] Jérôme Dedecker, An empirical central limit theorem for intermittent maps. Probab. Theory Related Fields (in press).

[7] Jérôme Dedecker, Clémentine Prieur, An empirical central limit theorem for dependent sequences, Stoch. Process Appl. 117 (1) (2007) 121–142.

[8] Monroe D. Donsker, Justification and extension of Doob's heuristic approach to the Komogorov–Smirnov theorems, Ann. Math. Statist. 23 (1952) 277–281.

[9] J.L. Doob, Heuristic approach to the Kolmogorov–Smirnov theorems, Ann. Math. Statist. 20 (1949) 393–403.

[10] P. Doukhan, P. Massart, E. Rio, Invariance principles for absolutely regular empirical processes, Ann. Inst. H. Poincaré Probab. Statist. 31 (2) (1995) 393–427.

[11] Paul Doukhan, Sana Louhichi, A new weak dependence condition and applications to moment inequalities, Stoch. Process Appl. 84 (2) (1999) 313–342.

[12] Paul Doukhan, Donatas Surgailis, Functional central limit theorem for the empirical process of short memory linear processes, C. R. Acad. Sci. Paris Sér. I Math. 326 (1) (1998) 87–92.

[13] O. Dovgoshey, O. Martio, V. Ryazanov, M. Vuorinen, The Cantor function, Expo. Math. 24 (1) (2006) 1–37.

[14] Olivier Durieu, A fourth moment inequality for functionals of stationary processes, J. Appl. Probab. 45 (4) (2008) 1086–1096.

[15] Olivier Durieu, Philippe Jouan, Empirical invariance principle for ergodic torus automorphisms; genericity, Stoch. Dyn. 8 (2) (2008) 173–195.

[16] M.I. Gordin, The central limit theorem for stationary processes, Dokl. Akad. Nauk SSSR 188 (1969) 739–741.

[17] Sébastien Gouëzel, An interval map with a spectral gap on Lipschitz functions, but not on bounded variation functions, Discrete Contin. Dyn. Syst. 24 (4) (2009) 1205–1208.

[18] Hubert Hennion, Loïc Hervé, Limit Theorems for Markov Chains and Stochastic Properties of Dynamical Systems by Quasi-compactness, in: Lecture Notes in Mathematics, vol. 1766, Springer-Verlag, Berlin, 2001.
[19] William Parry, Mark Pollicott, Zeta functions and the periodic orbit structure of hyperbolic dynamics, Astérisque 187–188 (1990) 268.
[20] Walter Philipp, Laurence Pinzur, Almost sure approximation theorems for the multivariate empirical process, Z. Wahrsch. Verw. Gebiete 54 (1) (1980) 1–13.
[21] Clémentine Prieur, An empirical functional central limit theorem for weakly dependent sequences, Probab. Math. Statist. 22 (2) (2002) 259–287. (Acta Univ. Wratislav. No. 2470).
[22] Wei Biao Wu, Empirical processes of stationary sequences, Statist. Sinica 18 (1) (2008) 313–333.
[23] Wei Biao Wu, Xiaofeng Shao, Limit theorems for iterated random functions, J. Appl. Probab. 41 (2) (2004) 425–436.