Rate of convergence for computing expectations of stopping functionals of an α-mixing process

Mohamed Ben ALAYA* Gilles PAGÈS†

Abstract

The *shift* method consists in computing the expectation of an integrable functional F defined on the probability space $((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^d)^{\otimes\mathbb{N}}, \mu^{\otimes\mathbb{N}})$ (μ is a probability measure on $\mathbb{R}^d$) using the Birkhoff's Pointwise Ergodic Theorem : $\frac{1}{n}\sum_{k=0}^{n-1} F\circ\theta^k \to \mathbb{E}(F)$ *a.s.* as $n\to+\infty$, where θ denotes the canonical shift operator. When F lies in $L^2(\mathcal{F}_T, \mu^{\otimes\mathbb{N}})$ for some integrable enough stopping time T, several weak (CLT) or strong (Gàl-Koksma Theorem or LIL) converging rates hold. The method successfully competes with Monte Carlo. The aim of this paper is to extend these results to more general probability distributions $\mathbb{P}$ on $((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^d)^{\otimes\mathbb{N}})$, namely when the canonical process $(X_n)_{n\in\mathbb{N}}$ is $\mathbb{P}-$stationary, α-mixing and fulfills Ibragimov's assumption $\sum_{n\geq 0}\alpha^{\frac{\delta}{2+\delta}}(n) < +\infty$ for some $\delta > 0$. One application is the computation of the expectation of functionals of an α-mixing Markov Chain under its stationary distribution $\mathbb{P}_\nu$. It may provide both a better accuracy and save the random number generator compared to the usual Monte Carlo or to the shift method on independent innovations.

1991 MSC classification : 60F05, 60F15, 60F17.
Key words : $\alpha-$mixing process, Monte Carlo method, Rate of convergence.

1 Introduction and mathematical framework

The origin of the problem is motivated by the computation of the expectation of a functional F defined on the canonical space $((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^d)^{\otimes\mathbb{N}}, \mu^{\otimes\mathbb{N}})$ using the Birkhoff's Pointwise Ergodic Theorem. Several contributions (Bouleau [4], [5] and Ben Alaya [1],[2]) have established some rates of convergence for a wide class of functionals, namely $\mathcal{F}_T^X$-measurable for some integrable enough stopping time T and square integrable. As a matter-of-fact, both strong (Gàl & Koksma Theorem, Law of the Iterated Logarithm) and weak (Central Limit Theorem) convergence rates hold in the Birkhoff's Theorem.

To be more specific, one considers the canonical space $((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^d)^{\otimes\mathbb{N}})$ endowed with the product measure $\mu^{\otimes\mathbb{N}}$, the canonical projections $X_k,\ k\geq 0$, defined for every $\omega := (\omega_k)_{k\geq 0} \in$

*LAGA URA 742, Institut Galilée Université Paris 13. Avenue J.B. Clément 93430 Villetaneuse & CERMICS, Ecole Nationale des Ponts et Chaussées. La courtine 93160 Noisy-le-Grand. Mail: mba@erato.enpc.fr.
†Univ. Paris 12, Fac. Sciences, 61 av. Général de Gaulle, F-94010 Créteil Cedex & Labo. Probabilités, URA 224, Univ. P.& M. Curie, Tour 56, F-75252 Cedex 05. Mail: gpa@ccr.jussieu.fr.

$(\mathbb{R}^d)^{\mathbb{N}}$ by $X_k(\omega) := \omega_k$ and the (left) shift operator on $(\mathbb{R}^d)^{\mathbb{N}}$ $\theta(X_0, X_1, \cdots) := (X_1, X_2, \cdots)$. It is widely known (see *e.g.* [13]) that the dynamical system $((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^d)^{\mathbb{N}}, \mu^{\otimes\mathbb{N}}, \theta)$ being ergodic, the Birkhoff Pointwise Ergodic Theorem implies that

$$\forall\, F \in L^1((\mathbb{R}^d)^{\mathbb{N}}, \mu^{\otimes\mathbb{N}}), \qquad \mu^{\otimes\mathbb{N}}\text{-}a.s. \quad \frac{1}{n}\sum_{k=0}^{n-1} F\circ\theta^k \longrightarrow \mathbb{E}(F) = \int F d\mu^{\otimes\mathbb{N}}.$$

Similarly, one can define the right shift (this time on $(\mathbb{R}^d)^{\mathbb{Z}}$) by setting $X_n \circ \theta^* = X_{n-1}$. Then identifying $L^1((\mathbb{R}^d)^{\mathbb{N}}, \mu^{\otimes\mathbb{N}})$ to a subspace of $L^1((\mathbb{R}^d)^{\mathbb{Z}}, \mu^{\otimes\mathbb{Z}})$

$$\forall\, F \in L^1((\mathbb{R}^d)^{\mathbb{N}}, \mu^{\otimes\mathbb{N}}), \qquad \mu^{\otimes\mathbb{N}}\text{-}a.s. \quad \frac{1}{n}\sum_{k=0}^{n-1} F\circ(\theta^*)^k \longrightarrow \mathbb{E}(F) = \int F d\mu^{\otimes\mathbb{N}}.$$

The Shift on Independent Innovations Method(s) (SIIM) simply is/are the data-processing of these convergence results. The expectation $\mathbb{E}(F)$ is then computed by averaging some dependent paths while the usual Monte Carlo method (MCM) requires some independent paths.

The main theoretical results concerning the θ-shift method are summed up below (see [2]). Let T be a $\mathcal{F}_n^X$-stopping time([1]) and $F \in L^2(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R})^{\otimes\mathbb{N}})$ be an $\mathcal{F}_T^X$-measurable functional([2]) where $\mathcal{F}_n^X := \sigma(X_0, \cdots, X_n)$ denotes the natural filtration of the canonical process $(X_n)_{n\in\mathbb{N}}$. Then if $T \in L^{2+\rho}$ for some $\rho > 0$,

• $\sigma^2(F) := \mathrm{Var}(F) + 2\sum_{k=1}^{+\infty} \mathrm{Cov}(F, F\circ\theta^k)$ is absolutely convergent, which in turn implies that :

• The Gàl-Koksma Theorem holds :

$$\forall\, \varepsilon > 0, \qquad \frac{1}{n}\sum_{k=0}^{n-1} F\circ\theta^k - \mathbb{E}(F) = o\left(n^{\frac{1}{2}}(\log(n))^{\frac{3}{2}+\varepsilon}\right) \quad \mu^{\otimes\mathbb{N}}\text{-}a.s.$$

• The CLT theorem holds, that is, whenever $\sigma(F) \neq 0$,

$$\frac{1}{\sigma(F)\sqrt{n}}\sum_{k=0}^{n-1}\left(F\circ\theta^k - \mathbb{E}(F)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0;1),$$

where $\mathcal{N}(0;1)$ denotes the standard normal distribution and $\xrightarrow{\mathcal{L}}$ the convergence in distribution.

• Moreover, if the stopping time T has finite polynomial moments (this assumption is slightly relaxable), the LIL holds

$$\limsup_{n\to\infty} \frac{\sum_{k=0}^{n-1}\left(F\circ\theta^k - \mathbb{E}(F)\right)}{\sqrt{2n\log\log n}} = \sigma(F) \qquad \text{and} \qquad \liminf_{n\to\infty} \frac{\sum_{k=0}^{n-1}\left(F\circ\theta^k - \mathbb{E}(F)\right)}{\sqrt{2n\log\log n}} = -\sigma(F).$$

[1] A $\mathbb{N}$-valued random variable is a $\mathcal{F}_n^X$-stopping time if $\{T \leq n\} \in \mathcal{F}_n^X$ for every $n \in \mathbb{N}$.
[2] $\mathcal{F}_T^X := \{A \in \mathcal{F}_\infty^X \,/\, A \cap \{T \leq n\} \in \mathcal{F}_n^X\}$ for every $n \in \mathbb{N}$

Similar results are obtained in the case of θ^*(see [5]). The computational performances of the shift method SIIM lie in the use of a storage box that partially avoids to uselessly re-simulate all the innovations X_i's when passing from a path to another while this is necessary in the usual MCM. Hence, for the same number of iterations, we observed on true simulations that the SIIM runs faster than the classical MCM (see [2]). The time savings are on the expenses of the data storage (dynamical or not) which is typical for the antagonism between time complexity and storage complexity. On the other hand, the SIIM also calls the random number generator less often than the MCM does. This may be crucial for large scale simulations. However, when $\sigma^2(F) > \mathrm{Var}(F)$ the required number of iterations is higher. Unfortunately no satisfactory estimate of $\frac{\sigma^2(F)}{Var(F)}$ is known to us and it is likely that, for most naturally encountered functionals F, this ratio is greater than 1. The balance between these two effects depends on the choice of F.

The aim of the paper is to extend these results to more general stationary probability distributions $\mathbb{P}$: whenever the dynamical $((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^d)^{\otimes\mathbb{N}}, \mathbb{P}, \theta)$ is ergodic, the Birkhoff's Theorem directly applied on the shifted paths of a $\mathbb{P}$-integrable functional F yields

$$\mathbb{P}\text{-}a.s. \qquad \frac{1}{n}\sum_{k=0}^{n-1} F\circ\theta^k \longrightarrow \mathbb{E}(F). \tag{1}$$

Of course, the plain ergodicity cannot provide a rate of convergence in the Birkhoff Pointwise Ergodic Theorem without any further assumption (see [13]). That is why we will assume from now on that the canonical process $(X_n)_{n\in\mathbb{N}}$ on $((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^d)^{\otimes\mathbb{N}})$ shares a strong mixing assumption, namely the Ibragimov α-mixing assumption, under the probability $\mathbb{P}$. This notion turns out to be the natural extension of the former of i.i.d. random variable setting in terms of Limit Theorems for our stopping functionals.

The α-mixing Markovian setting is a natural domain of application for the techniques. In fact, let $(X_n)_{n\in\mathbb{N}}$ be an homogeneous Markov chain on $\mathbb{R}^d$ with transition $P(x,dy)$ and a starting distribution μ_0. A commonly encountered problem of Numerical Probability is to compute (an approximate of) $\mathbb{E}_\nu(f(X_0,\cdots,X_{\ell-1}))$ where ν denotes the invariant probability measure – assumed to be unique – of the transition P. When the chain is positively recurrent (resp. stable), the natural method is to apply the Law of Large Numbers along the available paths of the chain that is, for every $x\in\mathbb{R}^d$, for every $f:(\mathbb{R}^d)^\ell\longrightarrow\mathbb{R}$ bounded Borel (resp. continuous) function

$$\forall\, x\in\mathbb{R}^d, \qquad \frac{1}{n}\sum_{k=0}^{n-1} f(X_k,\cdots,X_{k+\ell-1}) \stackrel{n\to+\infty}{\longrightarrow} \mathbb{E}_\nu(f(X_0,\cdots,X_{\ell-1})) \qquad \mathbb{P}_x\text{-}a.s. \tag{2}$$

The rate of convergence in (2) is ruled by several classical theorems like the Central Limit Theorem or the Law of the Iterated Logarithm under some standard assumptions (see $e.g.$ [9]).

When f is no longer a function of finitely many X_n's but is a functional F defined on the whole canonical space $((\mathbb{R}^d)^{\otimes\mathbb{N}}, \mathcal{B}(\mathbb{R}^d)^{\otimes\mathbb{N}})$ of the chain, the computation of $\mathbb{E}_\nu(F)$ either by simulation or from statistical data cannot be carried out as easily. The first natural idea is to implement the usual Monte Carlo Method (MCM). However this approach turns out to be costly in terms of C.P.U. time. Starting from experimental facts on the Shift on Independent

Innovation Method (SIIM), one can try using equation (1) under $\mathbb{P}_\nu$ and shift on the chain itself. We will call this method the Shift Process Method (SPM).

The Markov assumption on $(X_n)_{n\in\mathbb{N}}$ will be dropped in the theoretical part of the paper. The theoretical results will then be applied to the α-mixing case. Then the three methods (MCM, SIIM, SPM) will be compared numerically. They will be applied to two α-mixing Markovian models $-X_{n+1} = h(X_n, Y_n)-$ where the underlying innovations Y_n are independent.

The paper is organized as follows. Section 2 is devoted to some background on the main tools used in the rest of the paper : the definition of an α-mixing process is recalled along with the Ibragimov Central Limit Theorem for α-mixing sequences satisfying the Ibragimov assumption (subsection 2.1). The Gál-Koksma Theorem in the L^2-stationary setting is recalled at subsection 2.2. This will be our basic result when dealing with the $a.s.$ rate of convergence (except for the Law of the Iterated Logarithm investigated in section 5).

Section 3 deals with the $a.s.$ rate of convergence of the shift method for stopping functionals. This result essentially relies on the finiteness of a pseudo-variance, denoted $\sigma^2(F)$. In section 4, a Central Limit Theorem is established under the same hypothesis. In section 5, after recalling Philipp and Stout's Theorem, a Law of the Iterated Logarithm is established, only for a subclass of stopping functionals having finite polynomial moments. Section 6 is dedicated to the Markov setting. Some standard α-mixing criteria for (stationary) Markov chains are recalled (subsection 6.1) and the simulation framework is presented (subsection 6.2). Some numerical simulations are processed in section 7 on three $\alpha-$mixing Markov processes satisfying the Ibragimov assumption. A simple Metropolis like algorithm (subsection 7.1) is considered in two different settings so that the invariant distribution ν is alternately explicitly known (subsubsection 7.1.1) and not explicitly known (subsubsection 7.1.2). The third example, a Vector Quantization algorithm, will illustrate some possible false convergence phenomenon (subsection 7.2) when ν is not explicitly known.

Throughout the text $L^p(\Omega, \mathcal{A}, \mathbb{P})$ will denote the set of $\mathcal{A}$-measurable real-valued functionals F whose L^p-norm $||F||_p := \left(\int |F|^p d\mathbb{P}\right)^{\frac{1}{p}}$ is finite. From now on the shift operator θ will be the canonical shift on $(\mathbb{R}^d)^{\mathbb{N}}$.

2 Some background

2.1 α-mixing sequences and Central Limit Theorem

We are going to recall some results on α-mixing processes (see $e.g.$ Doukhan [7]). Let α be a sequence $(\alpha(n),\ n \in \mathbb{N})$ of real numbers, satisfying $\lim_{n\to\infty} \alpha(n) = 0$, and let $(X_n)_{n\in\mathbb{N}}$ be a $\mathbb{R}^d$-valued process defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. $(X_n)_{n\in\mathbb{N}}$ is α-mixing if for every $k,\ n \in \mathbb{N},\ n \geq 1$

$$\forall A \in \mathcal{F}_0^k,\ \forall B \in \mathcal{F}_{k+n}^\infty, \qquad |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \leq \alpha(n).$$

Intuitively if $\alpha(n)$ is small then B and A are essentially independent, hence for an α-mixing process the future is asymptotically independent from the present and the past. One finds in the literature various notions of mixing that quantify the dependence between the past

and the future. Just for comparison, in the φ-mixing for example we measure the quantity $|\mathbb{P}(B/A) - \mathbb{P}(B)|$. The notion of α-mixing is therefore weaker. It is in fact the weakest when compared to all usual notions of strong mixing (see [7]). However, this assumption on the sequence $(X_n)_{n\in\mathbb{N}}$ turns out to be quite adequate. Furthermore, we will say that an $\alpha-$mixing process $(X_n)_{n\in\mathbb{N}}$ satisfies the *Ibragimov assumption* if

$$\sum_{n\geq 0} \alpha^{\frac{\delta}{2+\delta}}(n) < +\infty \text{ for some } \delta > 0. \tag{3}$$

Historically, the Central Limit Theorem for α-mixing processes is due to Ibragimov (1962) (see [7] or Hall & Heyde [12]). It essentially holds under the above assumption (3).

Theorem 1 *Suppose that $(X_n)_{n\in\mathbb{N}}$ is a centered real valued strictly stationary[3] α-mixing process with $\sum_{n=0}^{\infty} \alpha(n)^{\frac{\delta}{2+\delta}} < +\infty$ and $\mathbb{E}|X_0|^{2+\delta} < +\infty$ for some $\delta > 0$. Then the sequence*

$$\sigma^2 := Var(X_0) + 2\sum_{k=1}^{\infty} Cov(X_0, X_k)$$

is absolutely convergent. Furthermore, if $\sigma > 0$, then

$$\frac{X_0 + X_1 + \cdots + X_{n-1}}{\sigma\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0;1) \text{ as } n \to +\infty,$$

where $\mathcal{N}(0;1)$ denotes the standard normal distribution and $\xrightarrow{\mathcal{L}}$ is for the convergence in law.

■

Except for the fundamental underlying Central Limit Theorem for the martingale increments, this result mainly relies on the covariance inequality below (see [7] p.9).

Proposition 1 *Let $(X_n)_{n\in\mathbb{N}}$ be a strictly stationary α-mixing process. Then:*
$\forall\ r, p, q \geq 1$ with $\frac{1}{r} + \frac{1}{p} + \frac{1}{q} = 1$, $\forall\ F \in L^p(\mathcal{F}_0^k)$ and $\forall\ G \in L^q(\mathcal{F}_{k+n}^{\infty})$

$$|Cov(F, G)| \leq 8\alpha^{\frac{1}{r}}(n)\, ||F||_p\, ||G||_q .$$

Application to cylindrical functions: Let us go back to the framework described in the introduction *i.e.* the canonical projections $(X_n)_{n\in\mathbb{N}}$ are α-mixing on the canonical dynamical space $((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^d)^{\otimes\mathbb{N}}, \mathbb{P})$ with a rate $\alpha := (\alpha(n))_{n\in\mathbb{N}}$. Then, the real valued measurable functions F on $((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^d)^{\otimes\mathbb{N}}, \mathbb{P})$ that actually depend on finitely many components behave like the sequence $(X_n)_{n\in\mathbb{N}}$ itself in the following sense: if F only depends on the first N components then $X_n^{(F,N)} := F \circ \theta^n,\ n \in \mathbb{N}$, is an α_N-mixing process with rate

$$\alpha_N(n) = \begin{cases} 1 & \text{if} \quad n \leq N \\ \alpha(n-N) & \text{if} \quad n > N \end{cases} .$$

[3] A $\mathbb{R}^d$-valued process $(X_n)_{n\in\mathbb{N}}$ is strictly stationary if for every $k \in \mathbb{N}$, $(X_{n+k})_{n\in\mathbb{N}}$ and $(X_n)_{n\in\mathbb{N}}$ have the same distribution that is, if $\mathbb{P}$ denotes the distribution of $(X_n)_{n\in\mathbb{N}}$ on the canonical space $((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^d)^{\otimes\mathbb{N}})$, $\mathbb{P} \circ \theta = \mathbb{P}$ with the notations of section 1.

Note that

$$\sum_{n=0}^{\infty} \alpha^{\frac{\delta}{2+\delta}}(n) < +\infty \iff \sum_{n=0}^{\infty} \alpha_N^{\frac{\delta}{2+\delta}}(n) < +\infty.$$

So, one straightforwardly derives the following

Proposition 2 *Let $(X_n)_{n\in\mathbb{N}}$ be a strictly stationary $\mathbb{R}^d$-valued α-mixing process. If there is $\delta > 0$ satisfying $\sum_{n=0}^{\infty} \alpha^{\frac{\delta}{2+\delta}}(n) < +\infty$, then for every cylindrical function $F \in L^{2+\delta}(\mathbb{P})$ with $\mathbb{E}(F)=0$,*

(a) the sequence $\sigma^2(F) := Var(F) + 2\sum_{k=1}^{\infty} Cov(F\circ\theta^k, F)$ is absolutely convergent,

(b) Furthermore, if $\sigma(F) > 0$, then: $\frac{1}{\sigma(F)\sqrt{n}} \sum_{k=0}^{n-1} F\circ\theta^k \xrightarrow{\mathcal{L}} \mathcal{N}(0;1)$ as $n \to +\infty$.

2.2 Rate of almost sure convergence

As a first step we recall the Gál and Koksma Theorem, established in their article "*Sur l'ordre de grandeur des fonctions sommables*"([11]). We will restrict to the L^2-stationary process setting (see [1] for a probabilistic proof in a full general setting).

Theorem 2 *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let $(X_n)_{n\in\mathbb{N}}$ be a L^2-stationary sequence of random variables such that $\mathbb{E}|X_1 + X_2 + \cdots + X_n|^2 = O(n)$. Then*

$$\forall\, \varepsilon > 0 \qquad X_1 + X_2 + \cdots + X_n = o\left(n^{\frac{1}{2}}(\log(n))^{\frac{3}{2}+\varepsilon}\right) \quad \mathbb{P}\text{-}a.s.$$

■

Coming back to the canonical dynamical system $((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^d)^{\otimes\mathbb{N}}, \mathbb{P}, \theta)$, we derive from the previous theorem a strong ergodic result. *i.e.* a speed of *a.s.* convergence in Birkhoff's pointwise ergodic Theorem.

Proposition 3 *Let $F \in L^2((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^d)^{\otimes\mathbb{N}}), \mathbb{P})$ such that $\mathbb{E}(F)=0$. If $\sigma^2(F) := Var(F) + 2\sum_{k=1}^{\infty} Cov(F\circ\theta^k, F)$ converges, then*

$$\forall\, \varepsilon > 0, \qquad \frac{1}{n}\sum_{k=0}^{n-1} F\circ\theta^k = o\left(n^{-\frac{1}{2}}(\log(n))^{\frac{3}{2}+\varepsilon}\right) \quad \mathbb{P}\text{-}a.s..$$

Proof: Using the convergence of the series $\sigma^2(F)$ and the fact that θ preserves the measure $\mathbb{P}$, we first prove that (see *e.g.* [1])

$$\mathbb{E}\left(\sum_{k=0}^{n-1} F\circ\theta^k\right)^2 = n\sigma^2(F) - 2\sum_{k=1}^{n} kCov(F\circ\theta^k, F) - 2n\sum_{k=n+1}^{\infty} Cov(F\circ\theta^k, F).$$

The finiteness of $\sigma^2(F)$ along with the Kronecker lemma yield

$$\lim_{n\to\infty}\frac{1}{n}\int_{(\mathbb{R}^d)^{\mathbb{N}}}|\sum_{k=0}^{n-1}F\circ\theta^k|^2 d\mathbb{P}=\sigma^2(F).$$

The Gál and Koksma's Theorem completes the proof. ■

By its very construction, a functional F that can be simulated on a computer naturally appears as a stopping functional with respect to its (*a.s.* finite) stopping rule T. So from now on, we will focus on such $\mathcal{F}_T$-measurable functionals.

3 An *a.s.* rate of convergence for stopping functionals

3.1 A class of $\mathcal{F}_T$-measurable functionals with finite $\sigma^2(F)$

Set $\mathcal{F}_m^n:=\sigma(X_m,\cdots,X_n)$ and $\mathcal{F}_m^\infty:=\sigma(X_k,k\geq m)$. T will denote a $\mathcal{F}_0^n$-stopping time and F a $\mathcal{F}_T$-measurable functional. Finally $[x]$ will denote the integral part of x.

Theorem 3 below provides a bound for the covariance $Cov(F\circ\theta^k,F)$ from which the absolute convergence of the series $\sigma^2(F)=Var(F)+2\sum_{k=1}^{\infty}Cov(F\circ\theta^k,F)$ follows. It is the key result of this work.

Theorem 3 *Let $(X_n)_{n\in\mathbb{N}}$ be a $\mathbb{R}^d$-valued stationary α-mixing process. Assume there is some $\delta>0$ such that $\sum_{n=0}^{\infty}\alpha^{\frac{\delta}{2+\delta}}(n)<+\infty$. If T is a stopping time and $T\in L^p((\mathbb{R}^d)^{\mathbb{N}},\mathbb{P})$ for some $p>\frac{2+\delta}{1+\delta}$ then, for every $F\in L^{2+\delta}((\mathbb{R}^d)^{\mathbb{N}},\mathcal{F}_T,\mathbb{P})$ with $\mathbb{E}(F)=0$, we have*

$$\left|Cov(F\circ\theta^k,F)\right|\leq 16\,||F||_{2+\delta}^2\,\alpha^{\frac{\delta}{2+\delta}}(k-[k/2])+||F||_{\frac{2+\delta}{1+\delta}}\,||F||_{2+\delta}\,\frac{(\mathbb{E}(T^p))^{\frac{1+\delta}{2+\delta}}}{[k/2]^{p\frac{1+\delta}{2+\delta}}}. \tag{4}$$

Proof: To establish inequality (4), first notice that

$$\left|Cov(F\circ\theta^k,F)\right|\leq\left|Cov(F\circ\theta^k,F\cdot\mathbf{1}_{\{T\leq[k/2]\}})\right|+\left|Cov(F\circ\theta^k,F\cdot\mathbf{1}_{\{T>[k/2]\}})\right|. \tag{5}$$

Now $F\circ\theta^k$ is $\mathcal{F}_k^\infty$-measurable and $F\cdot\mathbf{1}_{\{T\leq[k/2]\}}$ is $\mathcal{F}_0^{[k/2]}$-measurable. By applying Proposition 1 with $r=1+\frac{2}{\delta}$ and $p=q=2+\delta$, we obtain

$$\begin{aligned}\left|Cov(F\circ\theta^k,F\cdot\mathbf{1}_{\{T\leq[k/2]\}})\right| &\leq 8\alpha^{\frac{\delta}{2+\delta}}(k-[k/2])\left|\left|F\cdot\mathbf{1}_{\{T\leq[k/2]\}}\right|\right|_{2+\delta}||F||_{2+\delta}\\ &\leq 8\alpha^{\frac{\delta}{2+\delta}}(k-[k/2])\,||F||_{2+\delta}^2\,. \end{aligned}\tag{6}$$

For the second term on the right of inequality (5), the standard *Hölder* inequality with $p=2+\delta$ and $q=\dfrac{2+\delta}{1+\delta}$ first provides

$$\left|Cov(F\circ\theta^k,F\cdot\mathbf{1}_{\{T>[k/2]\}})\right|\leq\left|\left|F\circ\theta^k\cdot\mathbf{1}_{\{T>[k/2]\}}\right|\right|_{\frac{2+\delta}{1+\delta}}||F||_{2+\delta}\,. \tag{7}$$

It is straightforward that

$$\mathbb{E}\left(|F|^{\frac{2+\delta}{1+\delta}}\circ\theta^k\cdot\mathbf{1}_{\{T>[k/2]\}}\right)\leq\left|Cov(|F|^{\frac{2+\delta}{1+\delta}}\circ\theta^k,\mathbf{1}_{\{T>[k/2]\}})\right|+\mathbb{E}\left(|F|^{\frac{2+\delta}{1+\delta}}\right)\mathbb{P}(T>[k/2])$$

At this stage, we observe that $F\circ\theta^k$ is $\mathcal{F}_k^\infty$-measurable and $\{T>[k/2]\}$ belongs to $\mathcal{F}_0^{[k/2]}$. Still applying Proposition 1 but this time with $r=1+\dfrac{1}{\delta}$, $p=1+\delta$ and $q=+\infty$ yields

$$\mathbb{E}\left(|F|^{\frac{2+\delta}{1+\delta}}\circ\theta^k\cdot\mathbf{1}_{\{T>[k/2]\}}\right)\leq 8\alpha^{\frac{\delta}{1+\delta}}(k-[k/2])\left(\mathbb{E}\left(|F|^{2+\delta}\right)\right)^{\frac{1}{1+\delta}}+\mathbb{E}\left(|F|^{\frac{2+\delta}{1+\delta}}\right)\mathbb{P}(T>[k/2]),$$

that is

$$\left|\left|F\circ\theta^k\cdot\mathbf{1}_{\{T>[k/2]\}}\right|\right|_{\frac{2+\delta}{1+\delta}}\leq\left(8\alpha^{\frac{\delta}{1+\delta}}(k-[k/2])\left(\mathbb{E}\left(|F|^{2+\delta}\right)\right)^{\frac{1}{1+\delta}}+\mathbb{E}\left(|F|^{\frac{2+\delta}{1+\delta}}\right)\mathbb{P}(T>[k/2])\right)^{\frac{1+\delta}{2+\delta}}.$$

Plugging this bound in inequality (7) and using inequality $(x+y)^\beta\leq x^\beta+y^\beta$, $0<\beta<1$, $x,y\geq 0$ leads to

$$\begin{aligned}\left|Cov(F\circ\theta^k,F\cdot\mathbf{1}_{\{T>[k/2]\}})\right| &\leq 8^{\frac{1+\delta}{2+\delta}}\alpha^{\frac{\delta}{2+\delta}}(k-[k/2])\,||F||_{2+\delta}^2+||F||_{\frac{2+\delta}{1+\delta}}\,||F||_{2+\delta}\left(\mathbb{P}(T\geq[k/2])\right)^{\frac{1+\delta}{2+\delta}}\\ &\leq 8\alpha^{\frac{\delta}{2+\delta}}(k-[k/2])\,||F||_{2+\delta}^2+||F||_{\frac{2+\delta}{1+\delta}}\,||F||_{2+\delta}\left(\mathbb{P}(T\geq[k/2])\right)^{\frac{1+\delta}{2+\delta}}. \quad (8)\end{aligned}$$

As $T\in L^p((\mathbb{R}^d)^{\mathbb{N}},\mathcal{B}(\mathbb{R}^d)^{\otimes\mathbb{N}},\mathbb{P})$, $\mathbb{P}(T>[k/2])\leq\dfrac{\mathbb{E}(T^p)}{[k/2]^p}$. Hence, collecting inequalities (5), (6) and (8) finally yield:

$$\left|Cov(F\circ\theta^k,F)\right|\leq 16\,||F||_{2+\delta}^2\,\alpha^{\frac{\delta}{2+\delta}}(k-[k/2])+||F||_{\frac{2+\delta}{1+\delta}}\,||F||_{2+\delta}\,\frac{\left(\mathbb{E}(T^p)\right)^{\frac{1+\delta}{2+\delta}}}{[k/2]^{p\frac{1+\delta}{2+\delta}}},$$

which completes the proof. ■

Remarks and improvements:

(a) A careful reading of the above proof (namely equation (8)) shows that the assumption $T\in L^p$ for some $p>\frac{2+\delta}{1+\delta}$ can be slightly improved into

$$\sum_{k=1}^{\infty}\mathbb{P}(T>k)^{\frac{1+\delta}{2+\delta}}<+\infty.$$

(b) As $\frac{2+\delta}{1+\delta}<2$, the moment assumption on T is always fulfilled as soon as that: $T\in L^2((\mathbb{R}^d)^{\mathbb{N}},\mathcal{B}(\mathbb{R}^d)^{\otimes\mathbb{N}},\mathbb{P})$.

(c) If the functional F is bounded, then we can simply assume that T is integrable. Indeed, if T is integrable then $\sum_{k=1}^{\infty}\mathbb{P}(T>[k/2])<+\infty$ and the proof can be modified – in fact simplified! – in this setting (which formally corresponds to $\delta=+\infty$).

(d) Our assumptions on the process $(X_n)_{n\in\mathbb{N}}$ and the functional F are satisfactory in the following sense:

– Both conditions $\sum_{n=0}^{\infty} \alpha^{\frac{\delta}{2+\delta}}(n) < \infty$ and $F \in L^{2+\delta}$ do not differ from those of the original Ibragimov Central Limit Theorem which studies functions only depending on one variable (*i.e.* $F(x_0, \cdots, x_n, \cdots) := f(x_0)$).

– When $\alpha(n)=0$, $n \geq 1$, we find again the results of [2] obtained in the independent setting.

(e) The $\mathcal{F}_T$-measurability of the functional F for some stopping time T is crucial. In fact, we cannot obtain this result as a consequence of some results on functionals that can be approximated by a sequence $(F_k)_{k\in\mathbb{N}}$ of $\mathcal{F}_0^k$-measurable cylindrical functions so as $\sum_{k=1}^{\infty} ||F - F_k||_2 < +\infty$. By such a simple approach (setting $F_k := F.\mathbf{1}_{\{T\leq k\}}$), we get the result under the more stringent assumptions:

$$F \in L^{2+\delta} \text{ and } T \text{ has a moment of order } p > \tfrac{4(2+\delta)}{\delta}.$$

3.2 An $a.s$ rate of convergence

As it has been emphasized in paragraph 2.2 on the $a.s.$ convergence rate, the condition $\sigma^2(F) < +\infty$ is the basic assumption to apply Gál and Koksma Theorem (Theorem 2). Therefore, we derive from the previous theorem the following $a.s.$ convergence result.

Theorem 4 *Under the assumptions of Theorem 3, one has:*

$$\forall\, \varepsilon > 0, \qquad \frac{1}{n}\sum_{k=0}^{n-1} F \circ \theta^k = o\left(n^{-\frac{1}{2}}(\log(n))^{\frac{3}{2}+\varepsilon}\right) \quad \mathbb{P}\text{-}a.s..$$

■

4 A Central Limit Theorem for stopping functionals

Theorem 5 *Let $(X_n)_{n\in\mathbb{N}}$ be a $\mathbb{R}^d$-valued stationary α-mixing process. Assume there is some $\delta > 0$ such that $\sum_{n=0}^{\infty} \alpha^{\frac{\delta}{2+\delta}}(n) < +\infty$. If T is a stopping time and $T \in L^p((\mathbb{R}^d)^{\mathbb{N}}, \mathbb{P})$ for some $p > \frac{2+\delta}{1+\delta}$ then, for every $F \in L^{2+\delta}((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{F}_T, \mathbb{P})$ with $\mathbb{E}(F) = 0$, we have:*

$$\sigma^2(F) > 0 \quad \Longrightarrow \quad \frac{1}{\sigma(F)\sqrt{n}}\sum_{n=0}^{n-1} F \circ \theta^k \xrightarrow{\mathcal{L}} \mathcal{N}(0;1) \text{ as } n \to +\infty, \tag{9}$$

where $\mathcal{N}(0;1)$ denotes the standard normal distribution.

To establish the Central Limit Theorem we compute the limits of $\sigma^2(F.\mathbf{1}_{\{T\leq \ell\}})$ and $\sigma^2(F.\mathbf{1}_{\{T>\ell\}})$ when ℓ tends to $+\infty$. Indeed, if for every $\ell \in \mathbb{N}$ we set:

$$\sigma_\ell^2 := \sigma^2(F.\mathbf{1}_{\{T\leq \ell\}}) = Var(F.\mathbf{1}_{\{T\leq \ell\}}) + 2\sum_{k=1}^{\infty} Cov(F.\mathbf{1}_{\{T\leq \ell\}} \circ \theta^k, F.\mathbf{1}_{\{T\leq \ell\}}) \tag{10}$$

and

$$\tau_\ell^2 := \sigma^2(F.\mathbf{1}_{\{T>\ell\}}) = Var(F.\mathbf{1}_{\{T>\ell\}}) + 2\sum_{k=1}^{\infty} Cov(F.\mathbf{1}_{\{T>\ell\}} \circ \theta^k, F.\mathbf{1}_{\{T>\ell\}}),$$

then we have the following results.

Lemma 1 *Under the assumptions of Theorem 5,*

$$\lim_{\ell\to\infty} \sigma_\ell^2 = \sigma^2 \text{ and } \lim_{\ell\to\infty} \tau_\ell^2 = 0. \tag{11}$$

Proof : Following Theorem 3, if we replace the function F by $F.\mathbf{1}_{\{T\le\ell\}} - \mathbb{E}\left(F.\mathbf{1}_{\{T\le\ell\}}\right)$ (still $\mathcal{F}_T$-measurable and centered), inequality (4) yields an upper bound for $\left|Cov(F.\mathbf{1}_{\{T\le\ell\}} \circ \theta^k, F.\mathbf{1}_{\{T\le\ell\}})\right|$, namely

$$\left|Cov(F.\mathbf{1}_{\{T\le\ell\}} \circ \theta^k, F.\mathbf{1}_{\{T\le\ell\}})\right| \le 16 \left|\left|F.\mathbf{1}_{\{T\le\ell\}} - \mathbb{E}_\nu(F.\mathbf{1}_{\{T\le\ell\}})\right|\right|_{2+\delta}^2 \alpha^{\frac{\delta}{2+\delta}}(k-[k/2])$$

$$+ \left|\left|F.\mathbf{1}_{\{T\le\ell\}} - \mathbb{E}(F.\mathbf{1}_{\{T\le\ell\}})\right|\right|_{\frac{2+\delta}{1+\delta}} \left|\left|F.\mathbf{1}_{\{T\le\ell\}} - \mathbb{E}(F.\mathbf{1}_{\{T\le\ell\}})\right|\right|_{2+\delta} \frac{(\mathbb{E}(T^p))^{\frac{1+\delta}{2+\delta}}}{[k/2]^{p\frac{1+\delta}{2+\delta}}}.$$

Now for every functional $G \in L^p,\ p>1,\ ||G-\mathbb{E}(G)||_p \le 2||G||_p$, therefore,

$$\begin{aligned}\forall\, \ell \in \mathbb{N}, \quad \left|Cov(F.\mathbf{1}_{\{T\le\ell\}} \circ \theta^k, F.\mathbf{1}_{\{T\le\ell\}})\right| &\le 72 \left|\left|F.\mathbf{1}_{\{T\le\ell\}}\right|\right|_{2+\delta}^2 \alpha^{\frac{\delta}{2+\delta}}(k-[k/2]) \\ &\quad +4 \left|\left|F.\mathbf{1}_{\{T\le\ell\}}\right|\right|_{\frac{2+\delta}{1+\delta}} \left|\left|F.\mathbf{1}_{\{T\le\ell\}}\right|\right|_{2+\delta} \frac{(\mathbb{E}(T^p))^{\frac{1+\delta}{2+\delta}}}{[k/2]^{p\frac{1+\delta}{2+\delta}}},\end{aligned}$$

which in turn implies that, for every $\ell \in \mathbb{N}$

$$\begin{aligned}\left|Cov(F.\mathbf{1}_{\{T\le\ell\}} \circ \theta^k, F.\mathbf{1}_{\{T\le\ell\}})\right| &\le 72 ||F||_{2+\delta}^2 \alpha^{\frac{\delta}{2+\delta}}(k-[k/2]) \\ &\quad +4 ||F||_{\frac{2+\delta}{1+\delta}} ||F||_{2+\delta} \frac{(\mathbb{E}(T^p))^{\frac{1+\delta}{2+\delta}}}{[k/2]^{p\frac{1+\delta}{2+\delta}}}.\end{aligned}$$

Hence equation (10) implies that σ_ℓ^2 is defined by an absolutely convergent sequence, uniformly, with respect to ℓ. As each term of the series converges towards $Cov(F \circ \theta^k, F)$, one finally has $\lim_{\ell\to\infty} \sigma_\ell^2 = \sigma^2$.

If we note that $F.\mathbf{1}_{\{T>\ell\}} - \mathbb{E}\left(F.\mathbf{1}_{\{T>\ell\}}\right)$ is $\mathcal{F}_T$-measurable and centered we obtain in the same way:

$$\begin{aligned}\left|Cov(F.\mathbf{1}_{\{T>\ell\}} \circ \theta^k, F.\mathbf{1}_{\{T>\ell\}})\right| &\le 72 ||F||_{2+\delta}^2 \alpha^{\frac{\delta}{2+\delta}}(k-[k/2]) \\ &\quad +4 ||F||_{\frac{2+\delta}{1+\delta}} ||F||_{2+\delta} \frac{(\mathbb{E}(T^p))^{\frac{1+\delta}{2+\delta}}}{[k/2]^{p\frac{1+\delta}{2+\delta}}}.\end{aligned}$$

Hence τ_ℓ^2 is also defined by an absolutely convergent sequence uniformly with respect to ℓ. Since each term of the series converges towards 0, $\lim_{\ell\to\infty} \tau_\ell^2 = 0$.

■

Let us prove now the Central Limit Theorem.

Proof: Let $F \in L^2((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{F}_T, \mathbb{P})$. For every $\ell \in \mathbb{N}$ we write

$$F = \left(F.\mathbf{1}_{\{T\leq \ell\}} - \mathbb{E}(F.\mathbf{1}_{\{T\leq \ell\}})\right) + \left(F.\mathbf{1}_{\{T>\ell\}} - \mathbb{E}(F.\mathbf{1}_{\{T>\ell\}})\right).$$

Then

$$\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1} F\circ\theta^k = \frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}\left(F.\mathbf{1}_{\{T\leq \ell\}} - \mathbb{E}(F.\mathbf{1}_{\{T\leq \ell\}})\right)\circ\theta^k + \frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}\left(F.\mathbf{1}_{\{T>\ell\}} - \mathbb{E}(F.\mathbf{1}_{\{T>\ell\}})\right)\circ\theta^k.$$

$F.\mathbf{1}_{\{T\leq \ell\}} - \mathbb{E}(F.\mathbf{1}_{\{T\leq \ell\}})$ is a cylindrical function only depending on the first ℓ variables. According to proposition 2, the first term on the right of the equality converges in distribution towards $\mathcal{N}(0,\sigma_\ell^2)$, for every $\ell \in \mathbb{N}$, with

$$\sigma_\ell^2 = \sigma^2(F.\mathbf{1}_{\{T\leq \ell\}}) = Var(F.\mathbf{1}_{\{T\leq \ell\}}) + 2\sum_{k=1}^{\infty} Cov(F.\mathbf{1}_{\{T\leq \ell\}}\circ\theta^k, F.\mathbf{1}_{\{T\leq \ell\}}).$$

From Lemma 1, one derives the convergence in distribution of $\mathcal{N}(0,\sigma_\ell^2)$ towards $\mathcal{N}(0,\sigma^2)$. Consequently, it amounts to prove that, for every $\varepsilon > 0$,

$$\lim_{\ell\to\infty}\limsup_{n\to\infty}\mathbb{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}\left(F.\mathbf{1}_{\{T>\ell\}} - \mathbb{E}(F.\mathbf{1}_{\{T>\ell\}})\right)\circ\theta^k\right| \geq \varepsilon\right) = 0.$$

Then Bienaymé-Tchebichev's inequality yields

$$\mathbb{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}\left(F.\mathbf{1}_{\{T>\ell\}} - \mathbb{E}(F.\mathbf{1}_{\{T>\ell\}})\right)\circ\theta^k\right| \geq \varepsilon\right)$$
$$\leq \frac{1}{n\varepsilon^2}\int_{(\mathbb{R}^d)^{\mathbb{N}}}\left|\sum_{k=0}^{n-1}\left(F.\mathbf{1}_{\{T>\ell\}} - \mathbb{E}(F.\mathbf{1}_{\{T>\ell\}})\right)\circ\theta^k\right|^2 d\mathbb{P}.$$

The convergence of $\frac{1}{n}\int_{(\mathbb{R}^d)^{\mathbb{N}}}\left|\sum_{k=0}^{n-1}\left(F.\mathbf{1}_{\{T>\ell\}} - \mathbb{E}(F.\mathbf{1}_{\{T>\ell\}})\right)\circ\theta^k\right|^2 d\mathbb{P}$ towards τ_ℓ^2 yields:

$$\limsup_{n\to\infty}\mathbb{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}\left(F.\mathbf{1}_{\{T>\ell\}} - \mathbb{E}(F.\mathbf{1}_{\{T>\ell\}})\right)\circ\theta^k\right| \geq \varepsilon\right) \leq \frac{\tau_\ell^2}{\varepsilon^2}.$$

Lemma 1 completes the proof.

■

Remark : This CLT is satisfactory since it holds under the same Ibragimov assumption that rules the standard CLT for $\alpha-$mixing processes. However some recent work by Doukhan, Massart and Rio [8] shows that the (functional) CLT holds for a stationary $\alpha-$mixing processes $(X_n)_{n\in\mathbb{N}}$ whenever

$$\int_0^1 \alpha^{-1}(t)Q^2(t)dt < +\infty \tag{12}$$

where $t \longmapsto \alpha^{-1}(t)$ denotes the canonical inverse of the monotonic function $t \longmapsto \alpha([t])$ and Q denotes the quantile function of X_0.

5 The law of the Iterated Logarithm

The *a.s.* estimates for the convergence rate obtained in paragraph 3.2 using the Gál and Koksma Theorem (see Theorem 4) are obviously weaker than those of the standard Law of the Iterated Logarithm (LIL) property. The usefulness of these results is to provide an estimate close to the iterated logarithm but under weak and natural assumptions in simulation. However it is possible to prove the true LIL under more stringent assumptions on the functional F and the stopping time T.

Several results are available in the literature on the asymptotic behavior of the partial sums $\sum_{k=0}^{n-1} X_k$ of a "weakly dependent" $(X_k)_{k\in\mathbb{N}}$ process or on the partial sums $\sum_{k=0}^{n-1} F(X_k, X_{k+1}, \cdots)$ of a functional F depending on a "weakly dependent" process (see [3],[17]). Thus, W. Philipp and W. Stout in [17] provide several invariance principles for the partial sums of "weakly dependent" random variable sequences. Among them some are related to the sum of the functional of a delayed α-mixing process.

Philipp and Stout's Theorem: For the sake of simplicity we state Philipp and Stout's Theorem in the α-mixing stationary case still using the same notations. We go back to the canonical dynamic system $((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^d)^{\otimes\mathbb{N}}, \mathbb{P}, \theta)$.

Theorem 6 *(W. Philipp and W. Stout [17]) Let F be a centered function $\in L^{2+\delta}((\mathbb{R}^d)^{\mathbb{N}}, \mathbb{P})$ for some $0 < \delta \leq 2$, and $(F_k)_{k\in\mathbb{N}}$ an approximating sequence of $\mathcal{F}_0^k$-measurable functions. We assume that:*

(i) There is some constant C satisfying:

$$\forall\ n \in \mathbb{N} \qquad ||F - F_n||_{2+\delta} \leq \frac{C}{n^{2+\frac{7}{\delta}}}. \tag{13}$$

(ii)

$$\mathbb{E}\left(\sum_{k=0}^{n-1} F \circ \theta^k\right)^2 = n + O(n^{1-\frac{\delta}{30}}) \ \ as\ \ n \to \infty. \tag{14}$$

(iii) $(X_n)_{n\in\mathbb{N}}$ is a $\mathbb{R}^d$-valued stationary α-mixing sequence with:

$$\alpha(n) = o\left(n^{-168(1+\frac{2}{\delta})}\right). \tag{15}$$

Then the Law of the Iterated Logarithm holds, that is: $\mathbb{P}$*-a.s.*

$$\limsup_{n\to\infty} \frac{\sum_{k=0}^{n-1} F \circ \theta^k}{\sqrt{2n\log\log n}} = 1 \qquad \text{and} \qquad \liminf_{n\to\infty} \frac{\sum_{k=0}^{n-1} F \circ \theta^k}{\sqrt{2n\log\log n}} = -1.$$

■

The proof of this theorem is available in chapter 8 of [17].

We will apply now this theorem to $\mathcal{F}_T$-measurable functionals.

Application to stopping functionals of an α-mixing process: We study now some classes of functions depending on a stopping time. Hence we consider a $(\mathcal{F}_0^n)_{n\in\mathbb{N}}$-stopping time T, and a $\mathcal{F}_T$-measurable functional F.

Theorem 7 *Let $(X_n)_{n\in\mathbb{N}}$ be a $\mathbb{R}^d$-valued stationary α-mixing sequence and $\delta \in (0,2]$. Assume that $\alpha(n) = o\left(n^{-168(1+\frac{2}{\delta})}\right)$. If T is a stopping time and $T \in L^{p(\delta)}((\mathbb{R}^d)^{\mathbb{N}}, \mathbb{P})$ for some $p(\delta) > \dfrac{2(2+\delta)(1+\delta)(2\delta+7)}{\delta^2}$ ([4]) then, for every $F \in L^{2+2\delta}((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{F}_T, \mathbb{P})$ with $\mathbb{E}(F) = 0$ and $\sigma^2(F) > 0$, the Law of the Iterated Logarithm is satisfied, that is:*

$$\mathbb{P}\text{-}a.s. \quad \limsup_{n\to\infty} \frac{\sum_{k=0}^{n-1} F\circ\theta^k}{\sqrt{2n\log\log n}} = \sigma(F) \qquad \text{and} \qquad \liminf_{n\to\infty} \frac{\sum_{k=0}^{n-1} F\circ\theta^k}{\sqrt{2n\log\log n}} = -\sigma(F).$$

Remark: Note that $\varphi(\delta) := \dfrac{2(2+\delta)(1+\delta)(2\delta+7)}{\delta^2}$ is a decreasing function on $(0,2]$ so $\varphi(\delta) \geq \varphi(2) = 66$. For any practical implementation, such a requirement amounts to assuming that the stopping time T has moments of every order.

Proof: W.l.g., one may assume that $\sigma^2(F) = 1$ and $\alpha(n)$ is a non increasing sequence. We will now show that the assumptions of Theorem 6 are fulfilled. According to the proof of Proposition 3 one has:

$$\mathbb{E}\left(\sum_{k=0}^{n-1} F\circ\theta^k\right)^2 = n - 2\sum_{k=1}^{n} kCov(F\circ\theta^k, F) - 2n\sum_{k=n+1}^{\infty} Cov(F\circ\theta^k, F). \tag{16}$$

Let $p \in \left(\frac{2+\delta}{1+\delta}, p(\delta)\right)$. So, one has, following Theorem 3:

$$\begin{aligned} \forall F \in L^{2+\delta}((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{F}_T, \mathbb{P}) \qquad \left|Cov(F\circ\theta^k, F)\right| &<< \alpha^{-168}(k - [k/2]) + \frac{1}{[k/2]^{p\frac{1+\delta}{2+\delta}}} \\ &<< \alpha^{-168}(k) + k^{-p\frac{1+\delta}{2+\delta}} \end{aligned} \tag{17}$$

where $<<$ means that the left term is upper bounded by the right one up to a multiplicative constant. Let $A_n := \sum_{k=n+1}^{\infty} Cov(F\circ\theta^k, F)$. It follows from (17) that

$$\begin{aligned} |A_n| \quad &\leq \quad \sum_{k=n+1}^{\infty} |Cov(F\circ\theta^k, F)| << \sum_{k=n+1}^{\infty} \alpha^{-168}(k) + \sum_{k=n+1}^{\infty} k^{-p\frac{1+\delta}{2+\delta}} \\ &<< \quad n^{-167} + n^{-p\frac{1+\delta}{2+\delta}} \\ &<< \quad n^{-167} + n^{-\frac{\delta}{30}} \qquad \text{by setting } p := \frac{2+\delta}{1+\delta}\frac{30+\delta}{30} \in \left(\frac{2+\delta}{1+\delta}, p(\delta)\right) \\ &<< \quad n^{-\frac{\delta}{30}} \qquad \text{since} \quad \frac{\delta}{30} \leq \frac{1}{15} \leq 167. \end{aligned}$$

[4] Note that $\frac{2(2+\delta)(1+\delta)(2\delta+7)}{\delta^2} > \frac{2+\delta}{1+\delta}$ at least on $(0,2]$.

Now we need to estimate $\sum_{k=1}^{n} kCov(F\circ\theta^k, F)$. An Abel transform yields

$$\begin{aligned}\left|\sum_{k=1}^{n} kCov(F\circ\theta^k, F)\right| &= \left|\sum_{k=1}^{n} k(A_{k-1}-A_k)\right| \\ &\leq \sum_{k=0}^{n}|A_k| + n|A_n| << n^{1-\frac{\delta}{30}} \qquad \text{since } 0<\frac{\delta}{30}\leq\frac{1}{15}<1\end{aligned}$$

Plugging both estimates in equation (16) shows that assumption (ii) is fulfilled *i.e.*:

$$\mathbb{E}\left(\sum_{k=0}^{n-1} F\circ\theta^k\right)^2 = n + O(n^{1-\frac{\delta}{30}}).$$

In order to fulfill (i), if we set $F_k := F\cdot\mathbf{1}_{\{T\leq k\}}$, $r := 2(1+\delta)$ and $s := \dfrac{2(2+\delta)(1+\delta)}{\delta}$. The *Hölder* inequality (applied with the conjugate exponents $\frac{2(1+\delta)}{2+\delta}$ and $\frac{2(1+\delta)}{\delta}$) along with the Bienaymé-Tchebichev inequality finally lead to

$$||F-F_k||_{2+\delta} \leq ||F||_r \mathbb{P}(T>k)^{\frac{1}{s}} \leq ||F||_r \frac{\mathbb{E}\left(T^{(2+\frac{7}{\delta})s}\right)^{\frac{1}{s}}}{k^{2+\frac{7}{\delta}}}.$$

So (i) holds whenever T admits a moment of order $\dfrac{2(2+\delta)(1+\delta)(2\delta+7)}{\delta^2}$. This completes the proof. ■

Application: The restriction on the α-mixing coefficient is here very drastic. In practice it is essentially satisfied in the geometric framework. Thus, one recovers the result of the i.i.d. setting (*i.e.* $\alpha(n)=0$, $n\geq 1$).

Remark : The very same remark as that made in section 4 holds here as a recent paper by Rio (see [18]) shows that the LIL holds under assumption (12).

6 Markov setting

6.1 α-mixing Markov chains

We consider here a $\mathbb{R}^d$-valued Markov chain defined by its transition probabilities $(\pi(x,\cdot))_{x\in\mathbb{R}^d}$. We denote by $\mathbb{P}_x$ the probability distribution on the canonical space $((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^d)^{\otimes\mathbb{N}}, \mathbb{P})$ for which the sequence of canonical projections $(X_n)_{n\in\mathbb{N}}$ is a Markov chain with transition π and initial distribution δ_x. To deal efficiently with our expectation computation problem, it is necessary to suppose the existence of a stationary distribution ν, (*i.e.* $\nu\pi=\nu$ in other words: $\forall A\in\mathcal{B}(\mathbb{R}^d)^{\otimes\mathbb{N}}$ $, \nu(A) = \displaystyle\int_{\mathbb{R}^d}\pi(x,A)\nu(dx)$. Let $\mathbb{P}_\nu := \displaystyle\int_{\mathbb{R}^d}\mathbb{P}_x\nu(dx)$. One has $\theta(\mathbb{P}_\nu)=\mathbb{P}_\nu$ *i.e.* under $\mathbb{P}_\nu$ the process $(X_n)_{n\in\mathbb{N}}$ is strictly stationary. Let us recall now a characterization of the α-mixing in the framework of Markov chains.

Proposition 4 *(cf. [7]). Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with stationary distribution ν and transition probabilities $(\pi(x,.))_{x\in\mathbb{R}^d}$. If*

$$\lim_{n\to+\infty}\int_{\mathbb{R}^d}\sup_{A\in\mathcal{B}(\mathbb{R}^d)}|\pi^n(x,A)-\nu(A)|\,\nu(dx)=0$$

then the process $(X_n)_{n\in\mathbb{N}}$ is α-mixing under the stationary distribution $\mathbb{P}_\nu$ with

$$\alpha(n)\le 2\int_{\mathbb{R}^d}\sup_{A\in\mathcal{B}(\mathbb{R}^d)}|\pi^n(x,A)-\nu(A)|\,\nu(dx).$$

We provide the proof for the reader's convenience.

Proof: Let $A\in\mathcal{F}_0^k$ and $B\in\mathcal{F}_{k+n}^\infty$. Set $h(x):=\mathbb{E}_x(\mathbf{1}_B)$. We will denote by $\|\mu\|$ the total variation of a signed measure μ (*i.e.* $\|\mu\|:=|\mu|(\mathbb{R}^d)$). First note that there is some $\tilde{B}\in\mathcal{F}_0^\infty$ s.t. $B=(\theta^{n+k})^{-1}(\tilde{B})$. Hence $h(x):=\mathbb{E}_x(\mathbf{1}_{\tilde{B}}\circ\theta^{n+k})$. The stationarity of ν implies that $\nu(h)=\int\nu(dx)\mathbb{E}_x(\mathbf{1}_{\tilde{B}}\circ\theta^{n+k})=\int\nu(dx)\mathbb{E}_x(\mathbf{1}_B)=\mathbb{P}_\nu(B)$. Then

$$\begin{aligned}
|\mathbb{P}_\nu(A\cap B)-\mathbb{P}_\nu(A)\mathbb{P}_\nu(B)| &= \left|\mathbb{E}_\nu\left(\mathbf{1}_A(\mathbb{E}_\nu(\mathbf{1}_{\tilde{B}}\circ\theta^{n+k}/\mathcal{F}_0^{n+k})-\mathbb{P}_\nu(B))\right)\right|\\
&= \left|\mathbb{E}_\nu\left(\mathbf{1}_A(\mathbb{E}_{X_{n+k}}(\mathbf{1}_{\tilde{B}})-\mathbb{P}_\nu(B))\right)\right|\\
&= |\mathbb{E}_\nu\left(\mathbf{1}_A(h(X_{n+k})-\nu(h))\right)|\\
&= \left|\mathbb{E}_\nu\left(\mathbf{1}_A\int_{\mathbb{R}^d}h(x)(\pi^n(X_k,dx)-\nu(dx))\right)\right|\\
&\le \mathbb{E}_\nu\left(\mathbf{1}_A\,\|\pi^n(X_k,dx)-\nu(dx)\|\,\|h\|_\infty\right)
\end{aligned}$$

As the functions h and $\mathbf{1}_A$ are $[0,1]$-valued, it follows that

$$|\mathbb{P}_\nu(A\cap B)-\mathbb{P}_\nu(A)\mathbb{P}_\nu(B)|\le\mathbb{E}_\nu\left(\|\pi^n(X_k,.)-\nu\|\right)=\int_{\mathbb{R}^d}\|\pi^n(x,\cdot)-\nu\|\nu(dx).$$

The inequality $\|\mu\|\le 2\sup_{A\in\mathcal{B}(\mathbb{R}^d)}|\mu(A)|$ completes the proof.

■

Connection with the ergodicity properties of a Markov chain: In the Markov chain literature, two classes have especially been studied. Let A be a non negative function such that $\int_{\mathbb{R}^d}A(x)\nu(dx)<+\infty$. We assume that

$$\sup_{B\in\mathcal{B}(\mathbb{R}^d)}|\pi^n(x,B)-\nu(B)|\le A(x)u_n\quad\text{ with }\lim_{n\to\infty}u_n=0. \tag{18}$$

If $u_n:=\rho^n$, $0<\rho<1$, (18) is called the *geometric ergodicity* property and if $u_n:=\frac{1}{n^\gamma}$, $\gamma>0$, (18) is called the *Riemann recurrence* property. In all cases, under $\mathbb{P}_\nu$, we have an α-mixing Markov chain with $\alpha(n)=O(u_n)$.

General remark: Concerning the convergence in distribution, one has to note that results which are stated below exclusively hold under $\mathbb{P}_\nu$. On the other hand all results dealing with almost sure convergence, established under the distribution $\mathbb{P}_\nu$, remain true under $\mathbb{P}_x$ for ν *almost every* x.

Comment: Theorem 7 stresses the interest of the Strong Ergodic Theorem obtained in section 3.2 as it yields a rather similar result but under much looser assumptions on the α-mixing coefficient. Thus Theorem 5 holds in the case of Riemann recurrence as soon as $\gamma > 1 + \frac{2}{\delta}$ while $\gamma > 168(1 + \frac{2}{\delta})$ is necessary to get the LIL.

6.2 Mathematical and simulation framework

Let $(X_n)_{n\in\mathbb{N}}$ be an homogeneous Markov chain admitting a representation of the form

$$X_{n+1} = h(X_n, Y_{n+1}),\ n \geq 0,\ X_0 \overset{\mathcal{L}}{\sim} \mu_0 \tag{19}$$

where $h : \mathbb{R}^d \times \mathbb{R}^p \longrightarrow \mathbb{R}^d$ is a Borel function ([5]), $(Y_n)_{n\in\mathbb{N}}$ is an i.i.d. sequence of μ distributed innovations and μ_0 is a (starting) distribution on $\mathbb{R}^d$. One can notice that, if F is a functional of the Markov chain $(X_n)_{n\in\mathbb{N}}$ starting at $x \in \mathbb{R}^d$, there is some functional G on $\mathbb{R}^d \times \mathbb{R}^{\mathbb{N}}$ such that

$$F(X_0, X_1, \cdots, X_n, \cdots) = G(x, Y_1, Y_2, \cdots, Y_n, \cdots).$$

So, the shift on independent innovation method (SIIM) naturally yields an estimate and an error bound for $\mathbb{E}_x(F) = \int_{\mathbb{R}^{\mathbb{N}}} G(x, y_1, \cdots, y_n, \cdots) d\mu^{\otimes\mathbb{N}}(y_0, y_1, \ldots, y_n, \ldots, \ldots)$.

It is possible the same way round to approximate $\mathbb{E}_{\mu_0}(F)$ with some similar weak and strong error bounds whenever the starting distribution μ_0 of X_0 can be simulated from the distribution μ of the innovation Y_0 (*i.e.* there exists some φ s.t. $\varphi(Y_0) \sim \mu_0$).

Among all the possible starting distributions μ_0 for the chain, the invariant one plays an essential role for obvious reasons related to Statistics or Simulation. However the computation of functional expectations under $\mathbb{P}_\nu$ makes problem when this invariant distribution is not explicitly known from the distribution of the innovation μ. This is usually due to the fact that no information is available on ν except for its existence and uniqueness. One way to apply the SIIM or the MCM is to prove that $\mathcal{L}(X_n)$ converges fast enough to ν for any starting value $x \in \mathbb{R}^d$ (that is $\mu\pi^n(dy) \to \nu(dy)$ in distribution or, if possible, in variation). This will be the case when the chain $(X_n)_{n\in\mathbb{N}}$ has further properties as Doeblin or geometrical ergodicity, Riemann recurrence or even stability (see below and [9] or [7]).

Anyway, whatever method is used – MCM or SIIM – getting an approximate for $\mathbb{E}_\nu(F)$ needs, *prior to* the simulation of *every independent trial* of F (or every $F\circ\theta^k$), to *re-simulate* an approximately ν-distributed starting value. Practically, it amounts to assigning $X_0 := X_{n_0}$ for a large enough n_0. Such preliminary simulations are C.P.U time and random number consuming. The major drawback eventually remains that the method actually converges to $\mathbb{E}_{\mathcal{L}(X_{n_0})}(F)$ with no available control on the bias.

On the other hand, if the chain $(X_n)_{n\in\mathbb{N}}$ is ergodic, the Birkhoff's Theorem applied on $((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^d)^{\otimes\mathbb{N}}, \mathbb{P}_\nu)$ to the canonical shift θ yields

$$\mathbb{P}_\nu\text{-}a.s. \qquad \frac{1}{n}\sum_{k=0}^{n-1} F \circ \theta^k \longrightarrow \mathbb{E}_\nu(F) \text{ as } n \to +\infty. \tag{20}$$

[5] For notational convenience we will assume that $p=1$.

It follows that the convergence also holds $\mathbb{P}_x$-$a.s.$ for ν-almost every $x \in \mathbb{R}^d$. So, whenever the $(X_n)_{n\in\mathbb{N}}$ is $\alpha-$mixing the above theoretical results would provide some valuable information about expectation computation based on a direct shifting of the $(X_n)_{n\in\mathbb{N}}$'s (SPM). The SPM has the same possible advantages as the SIIM (shift on the innovations): preservation of the random number generator, (more) time saving by a drastic reduction of the numerical computations. Besides it converges to the true value $\mathbb{E}_\nu(F)$ and there is no longer theoretical bias. On a more practical point of view, no preliminary simulation (or computation) of an approximately ν-distributed starting value X_{n_0} is really necessary.

7 Applications and simulations

Most stopping times used in simulations actually are some hitting time T_A of a given Borel set A that is

$$T_A := \min\{n \in \mathbb{N} / \, X_n \in A\}.$$

So we will concentrate in the examples below to functionals related to such stopping times (these functionals can be the stopping time itself).

7.1 A simple Metropolis like algorithm

The simplified version of the Metropolis algorithm used below for the testing procedure is mentioned in [19]. Let ρ be a $[0,1]$-valued function defined on the whole real line and set

$$X_{n+1} := \begin{cases} X_n & \text{if } U_{n+1} > \rho(X_n) \\ Z_{n+1} & \text{if } U_{n+1} \le \rho(X_n) \end{cases} \tag{21}$$

where $Y_n := (U_n, Z_n)_{n\ge 1}$ is a sequence of i.i.d. random vectors with distribution $U([0,1]) \otimes \mu$.

Usually, such a procedure is implemented to provide some approximately ν-distributed numbers. It looks a bit like the rejection method except that the number of trials before getting one (almost)-ν distributed number can be *a priori* bounded.

As soon as $\int \rho^{-1} d\mu < +\infty$, $\nu := \dfrac{\rho^{-1} d\mu}{\int \rho^{-1} d\mu}$ is the unique invariant distribution for the chain.

7.1.1 A first example (ν is explicitly known)

When μ has a compact support, say the unit interval $[0,1]$, it is obvious that for every $[0,1]$-valued starting value x, the whole process $(X_n)_{n\in\mathbb{N}}$ lives in $[0,1]$. Then, if ρ is continuously defined on $[0,1]$, the transition $(\pi(x,dy))_{x\in[0,1]}$ is Feller on $[0,1]$. Under the above uniqueness assumption of the invariant probability distribution, the chain is then ν-stable.

For example, if $\mu := \beta(a+1,1)$ and $\rho(x) := \begin{cases} 0 & \text{if } x \ge 0 \\ x & \text{if } x \in [0,1] \\ 1 & \text{if } x \ge 1 \end{cases}$ then ν-stability holds for $\nu = \beta(a,1)$. In that special setting, one can show following [8], that the chain is in fact α-mixing (even β-mixing).

For our purpose it is more significant to focus on a rather general stopping time, say

$$F := T := \inf\{n \geq 1, \sum_{k=1}^{n} X_k \notin [-10, 10]\}$$

as the simulated functional.

Numerical comparison of the methods: This first set up makes possible a sketch of comparison between the three methods: MCM, SIIM (shift on the innovations) and SPM (shift on the chain itself), in terms of C.P.U. time, random number generator, etc. As a matter of fact an exact simulation procedure for the invariant distribution ν is available (set $X_0 := U_0^{\frac{1}{a}}$ for some uniformly distributed r.v. U_0).

The results are following for $a := 2$

	C.P.U. time (s)			Used random numbers			$\mathbb{E}_\nu(F)$		
Iterations	MCM	SIIM	SPM	MCM	SIIM	SPM	MCM	SIIM	SPM
n=1 000	0.370	0.180	0.060	27 064	12 508	1 714	15.740	15.796	15.796
n=5 000	1.770	0.790	0.240	135 627	61 695	8 412	15.792	16.091	15.691
n=10 000	2.650	1.990	0.360	272 721	123 514	16 811	15.930	15.994	15.657
n=50 000	13.190	11.780	2.420	1361 032	616 114	83 384	15.916	15.910	15.973
n=100 000	32.440	18.480	4.610	2 722 984	1 229 784	167 019	15.932	15.925	15.872

The table of the C.P.U. time shows that, *for a given number of iterations*, the SPM method is 7 times faster than the MCM while the SIIM method is roughly speaking twice faster. Of course such factors are strongly depending of the setting and can in no way be adopted as general rules.

INSERT FIGURE 1 AROUND HERE

Taking into account these factors, while n iterations are processed for the MCM algorithm, $7n$ (resp. $2n$) iterations are processed for the SPM (resp. SIIM) algorithm. The graphic in Figure 1 was plotted so that the abscissa axis represents the C.P.U time (expressed in equivalent MCM iteration number). Thus, above $n := 2.10^4$ are plotted the approximates of $\mathbb{E}_\nu(T)$ obtained by the three methods while 2.10^4 iterations were processed by the MCM.

It turns out that the speed improvement of the method satisfactorily compensates the increase of $\sigma^2(T)$. Finally the gain true essentially lies in the *saving of the random number generator*.

7.1.2 A second example (ν is not explicitly known)

When $\rho > \varepsilon_0$, one readily checks that the transition

$$\pi(x, dy) := (1 - \rho(x))\delta_x(dy) + \rho(x)\mu(dy) \geq \varepsilon_0 \mu(dy)$$

is Doeblin recurrent, hence α-mixing. We set for this example $\rho(x) = 0.25 + 0.5 \times e^{-|x|}$ and $\mu := \mathcal{N}(0; 1)$.

Numerical comparison of the methods: This second set up is in some sense more realistic as it requires a preliminary simulation of the invariant distribution. This preprocessing, supposed to geometrically converge due to Doeblin recurrence, was made using $n_0 = 1\,000$ trials of the chain (see section 6.2). The simulations are displayed in Figure 2.

INSERT FIG. 2 AROUND HERE

The three methods clearly converge but, seemingly, toward two separate limits: the MCM and SIIM methods going on one side, the SPM on another. Two interpretations can reasonably be proposed: either the SPM method is too slow and a false convergence phenomenon occurs or the SPM value is right and both MCM and SIIM (which estimate $\mathbb{E}_{\mathcal{L}(X_{n_0})}(T)$) are wrong *i.e.* $\mathbb{E}_{\mathcal{L}(X_{n_0})}(T) \not\approx \mathbb{E}_\nu(T)$.

Taking into account the theoretical properties of the methods developed in subsection 6.2, along with some further simulations processed below with another algorithm: the Competitive Learning Vector Quantization, we guess that the best estimate is the one provided by the SPM method. However this remains debatable and would need some large scale simulations to draw some general rule.

7.2 A Vector Quantization Algorithm (1-dim setting)

The 1-dim Competitive Learning Vector Quantization algorithm: Let μ be a probability measure on $[0,1]$. One defines on $F_n^+ := \{u \in [0,1]^n \,/\, 0 < u_1 < \cdots < u_n < 1\}$ the so-called *n-distortion* $E_n^\mu(x)$ of a n-tuple $x := (x_1, \cdots, x_n) \in F_n^+$ by

$$E_n^\mu(x_1, \cdots, x_n) := \int \min_{1 \le i \le n} (x_i - u)^2 \mu(du).$$

This function measures how the n-tuple $(x_1, \cdots, x_n)$ can be considered as a good "skeleton" or "quantification" of the distribution μ: the lower $E_n^\mu(x)$ is, the better x quantifies μ. The multi-dimensional version of the distortion is widely used in Automatic Classification to optimally reduce the size of a data set. Some applications to Numerical Integration are also developed (see [15] or [16]). So it is important to reach an element of $\mathrm{argmin}_{F_n^+} E_n^\mu$.

It is well-known (see *e.g.* [6]) that

$$\nabla E_n^\mu(x_1, \cdots, x_n) = 2 \left(\int_{\tilde{x}_i}^{\tilde{x}_{i+1}} (x_i - u) \mu(du) \right)_{1 \le i \le n} \quad \text{where } \tilde{x}_i := \frac{x_i + x_{i+1}}{2},\ 2 \le i \le n-1,$$

$\tilde{x}_1 := 0$, $\tilde{x}_{n+1} := 1$. Note that ∇E_n^μ admits an obvious continuous extension on the closure $\overline{F}_n^+$ of F_n^+. A classical deterministic minimizing procedure $x^{t+1} := x^t - \varepsilon \nabla E_n^\mu(x^t)$ ($\varepsilon \in (0, 1/2)$ seems intractable because of the integral form of the gradient when the μ is not uniform. On the other hand, whenever μ has a simple simulating procedure, the so-called "Competitive Learning Vector Quantization" related stochastic gradient descent

$$X^0 \in F_n^+,\ X_i^{t+1} := X_i^t - \varepsilon \mathbf{1}_{]\tilde{X}_i^t, \tilde{X}_{i+1}^t]}(\omega^{t+1})(X_i^t - \omega^{t+1}),\ ; (\omega^t) \text{ i.i.d. \& } \mu\text{-distributed} \qquad (22)$$

can straightforwardly be implemented on a computer. Note that the CLVQ algorithm is also mentioned in the Neural Network literature as the Kohonen algorithm with 0 neighbor. Although such an algorithm with constant step cannot converge in the *a.s.* sense, most

practitioners implement it that way to avoid the metastability problems encountered with the algorithm with decreasing step.

Equation (22) defines for every $\varepsilon \in (0,1/2)$ an homogeneous $\overline{F}_n^+$-valued Markov chain. When μ is diffuse (*i.e.* weights no single point) this chain admits a Feller extension on $\overline{F}_n^+$ by properly defining the algorithm on n-tuples having stuck components [6]. Then the existence of an invariant probability measure ν^ε is straightforward and it can be shown that, in fact, $\nu^\varepsilon(F_n^+)=1$. On the other hand, it has been established that whenever

$$\exists O \subset [0,1], \text{ open set, s.t. } \mu_{|O} \geq \alpha \lambda_{|O} \tag{23}$$

the CLVQ algorithm is *Doeblin* recurrent on F_n^+ (and subsequently has a unique invariant probability measure ν^ε even if μ is not diffuse). Many simulations processed with this algorithm shows that the geometric convergence ratio ρ is very close to 1.

This family of compactly supported distributions ν^ε, $\varepsilon \in (0,1/2)$ is tight, and still assuming that μ is diffuse, one may show (see [6] or [10]) that any weak limiting value ν^0 of the tight family $(\nu^\varepsilon)_{\varepsilon\in(0,1/2)}$ satisfies $\mathrm{supp}(\nu^0) \subset \{\nabla E_n^\mu = 0\}$. Thus whenever $\{\nabla E_n^\mu = 0\}$ is reduced to a single point x^*, one has $\nu^\varepsilon \Rightarrow \delta_{x^*}$.

This is the case *e.g.* if $\mu(du)=f(u)du$ where f is either strictly ln-concave or ln-concave with $f(0_+)+f(1_-)>0$ (see [14]). Furthermore, when $\mu:=U([0,1])$, $x^*:=(\frac{2k-1}{2n})_{1\leq i\leq n}$.

The simulations They were processed with the uniform distribution $\mu:=U([0,1])$, $n:=10$, $\varepsilon:=0.1$. All the preliminary simulations were always processed starting from the equilibrium point x^*. We considered the hitting time

$$F := T := \min\left\{ t\in\mathbb{N} \,/\, \left| X^t - (\frac{2k-1}{2n})_{1\leq k\leq n} \right|_2 < \frac{1.1}{n} \right\}.$$

• The MCM: Our aim was to study the convergence of the MCM method as a function of $\mathcal{L}(X_{n_0})$ where n_0 denotes the number of preliminary simulation (keep in mind that $\mathcal{L}(X_{n_0})$ geometrically converges in variation to ν^ε). On the other hand, it has been highlighted that the MCM actually approximates $\mathbb{E}_{\mathcal{L}(X_{n_0})}(T)$ instead of $\mathbb{E}_\nu(T)$ and that no control of $|\mathbb{E}_{\mathcal{L}(X_{n_0})}(T) - \mathbb{E}_\nu(T)|$ as a function of $\|\mathcal{L}(X_{n_0}) - \nu\| = O(\rho^n)$ is known.

INSERT FIG 3 AROUND HERE

Figure 3 shows that, actually, if convergence holds, it is surprisingly slow. Any estimate of $\mathbb{E}_\nu(T)$ obtained by the MCM method with less than 5 000 preliminary simulations of the starting value for every path of the simulation of T provides a meaningless result.

The SIIM method fails exactly for the same reasons.

INSERT FIG. 4 AROUND HERE

• The SPM: Figure 4 displays a simulation by the SPM method processed on 10^8 trials. It shows how slow the convergence is toward an approximate value of $\mathbb{E}_\nu(T) \approx 5.1$.

INSERT FIG 5 AROUND HERE

• The comparison was lead (see Figure 5) like for the Metropolis like algorithm : the abscissa axis represents the number n of iterations of the MCM. Above every n are plotted the values obtained for $\mathbb{E}_\nu(T)$ by the three methods "while n iterations of the MCM were run". Roughly speaking this amounts to comparing the methods *vs* C.P.U. time.

Figure 5 shows that the SPM method converges much faster than MCM and SIIM, or to be more specific, is far less costly in term of random numbers consumption.

8 Conclusion

We have extend to $\alpha-$mixing stationary process satisfying the usual Ibragimov assumption ($\sum_{n\geq 0} \alpha^{\frac{\delta}{2+\delta}}(n) < +\infty$ for some $\delta > 0$) some weak (CLT) and strong (Gàl-Koksma and LIL) rates of convergence for the pointwise Birkhoff's Theorem on the canonical dynamical system $((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^d)^{\otimes\mathbb{N}}, \mathbb{P}, \theta)$, formerly obtained in the i.i.d case (*i.e.* $\mathbb{P} = \mu^{\otimes\mathbb{N}}$).

One promising application is the computation of expectations of stopping functionals of an $\alpha-$mixing Markov chain under its stationary distribution $\mathbb{P}_\nu$. The main interest of the shift method is that no preliminary simulation of the invariant distribution ν is required while such simulations are necessary in the MCM before every simulated path or before the first path when shifting on the i.i.d. innovations (SIIM). Furthermore, for a given number of iteration, it saves the pseudo-random generator by storing intermediary results.

Both the CLT and the LIL are ruled by a pseudo-variance $\sigma^2(F) := Var(F) + 2\sum_{k=1}^{\infty} Cov(F \circ \theta^k, F)$ which is typically greater than $Var(F)$(and difficult to estimate!). This may cause trouble when specifying the number of iteration of the simulation and even partially annihilate the advantages of the SPM.

Some first tests processed on two strongly mixing Markov chains seem promising if not completely conclusive. Some large scale tests should be carried out to check the validity of the method on a practical point of view.

On a theoretical point of view, next question is now to investigate the recent Doukhan-Massart-Rio assumption in relation with stopping functionals.

References

[1] M. B. Alaya: *Les théorèmes ergodiques en simulation*, PhDthesis, Thèse de l'école nationale des ponts et chaussées, 1992.

[2] M. B. Alaya: On the simulation of expectations of random variables depending on a stopping time, *Stochastic Analysis and Applications*, 11(2), pp.133-153, 1993.

[3] P. Billingsley: *Convergence of Probability Measures*, Wiley, New-York, 1968.

[4] N. Bouleau: On effective computation of expectations in large or infinite dimension, *Journal of Computational and Applied Mathematics*, 31, pp.23-34, 1990. North-Holland.

[5] N. Bouleau: On numerical integration by the shift and application to Wiener space, *Acta Applicandae Mathematicae*, 25(3), pp.201-220, 1991.

[6] C. Bouton and G. Pagès, Convergence in distribution of the one-dimensional Kohonen algorithms when the stimuli are not uniform, *Advances in Applied Probability*, 26(1), pp.80-103, 1994.

[7] P. Doukhan: *Mixing, Properties and examples*, *Lecture notes in Statistics volume 85*, Springer Verlag, Berlin, 1994.

[8] P. Doukhan, P. Massart and E. Rio: The functional central limit theorem for stationary mixing processes. *Ann. Inst. H. Poincaré Probab. statist*, 30, pp.63-82, 1994.

[9] M. Duflo: *Méthodes récursives aléatoires*, Masson, Paris, 1990.

[10] J. C. Fort and G. Pagès: On the asymptotic behavior of the invariant distributions of a stochastic algorithm with constant step, *technical report, 1996, submitted for publication.*

[11] I. S. Gál and J. F. Koksma: Sur l'ordre de grandeur des fonctions sommables, *Indigationes Math.*, 12, pp.192-207, 1950.

[12] P. Hall and C. C. Heyde: *Martingale limit theory and its applications*, Academic Press, 1980.

[13] U. Krengel: *Ergodic Theorems*, de Gruyter Studies in Mathematics, Berlin, New-York, 1985.

[14] D. Lamberton and G. Pagès: On the critical points of the 1-dimensional Competitive Learning Vector Quantization Algorithm, *Proceedings of the ESANN'96*, Bruges (Belgium), 1996.

[15] G. Pagès: Voronoï Tessellation, space quantization algorithm and numerical integration, *Proceedings of the ESANN'93*, M. Verleysen éd., Editions D Facto (ISBN 2-9600049-0-6), Bruxelles, pp.221-228, 1993.

[16] G. Pagès: *Numerical Integration by Space Quantization,* submitted for publication.

[17] W. Philipp and W. Stout: *Almost sure invariance principles for partial sums of weakly dependent variables*, Technical report, Memoirs of the American Mathematical Society, n^o161, 1975.

[18] E. Rio: The functional Law of the Iterated Logarithm for stationary strongly mixing sequences, *The Annals of Proba.*, 23(3), pp.1188-1203, 1995.

[19] C. P. Robert: Convergence assessments for Markov Chains Monte Carlo Methods, *Technical report* INSEE n^o9347.

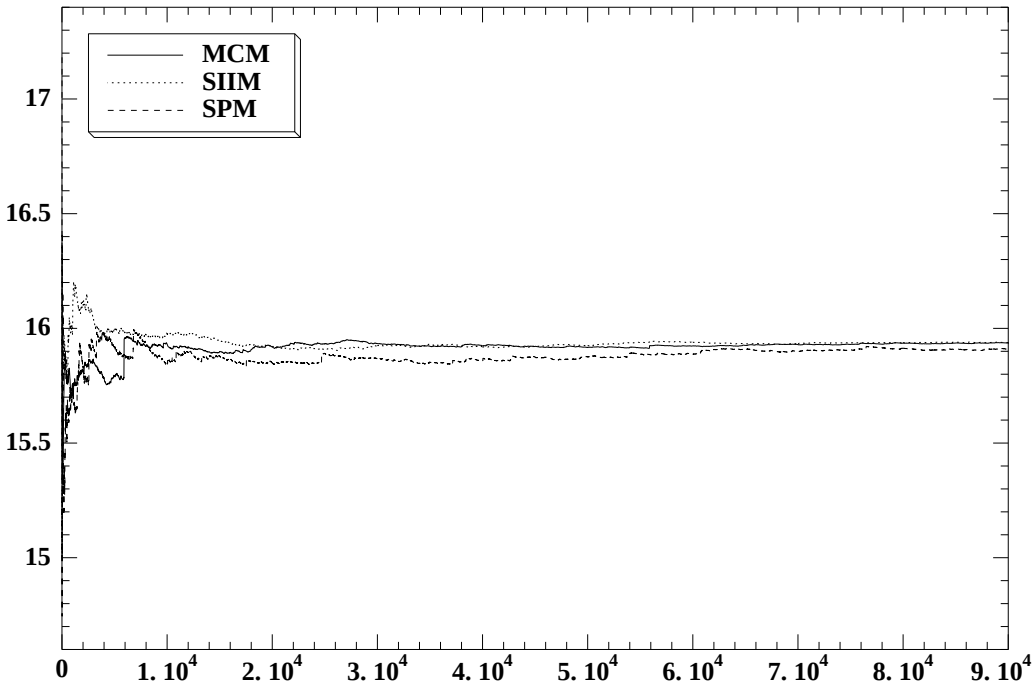

Figure 1: ν is explicitly known

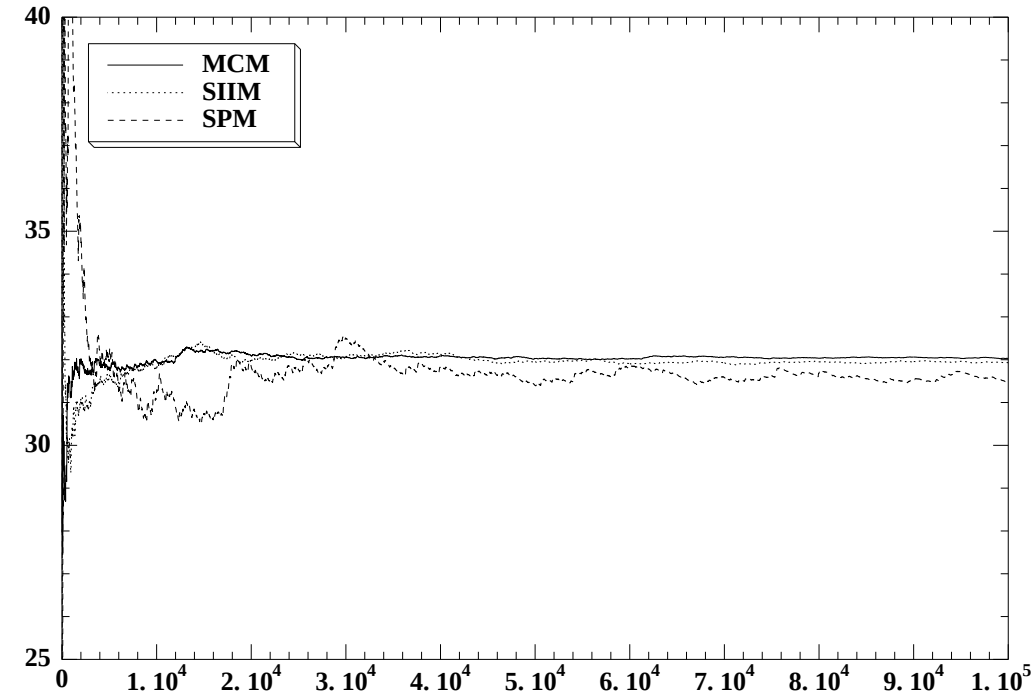

Figure 2: ν is not explicitly known

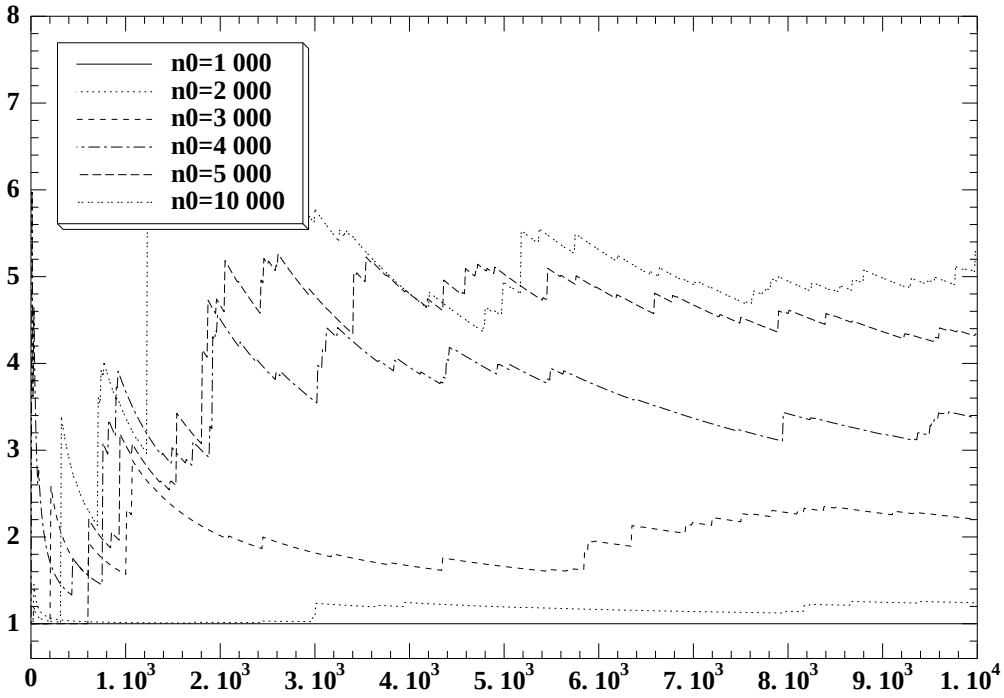

Figure 3: MCM for various n_0

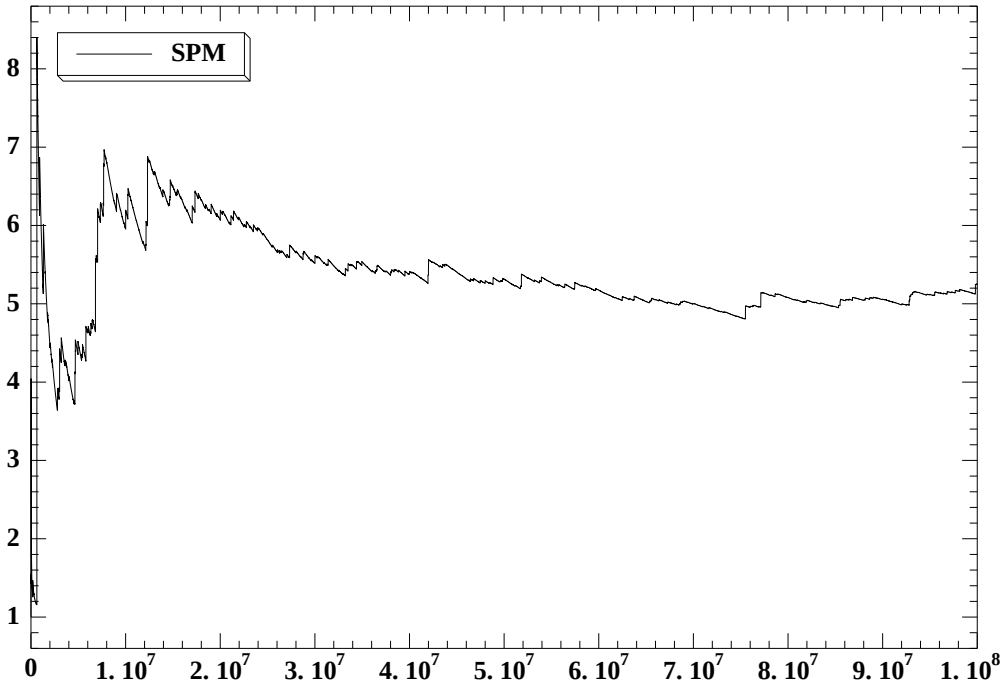

Figure 4: Convergence of SPM

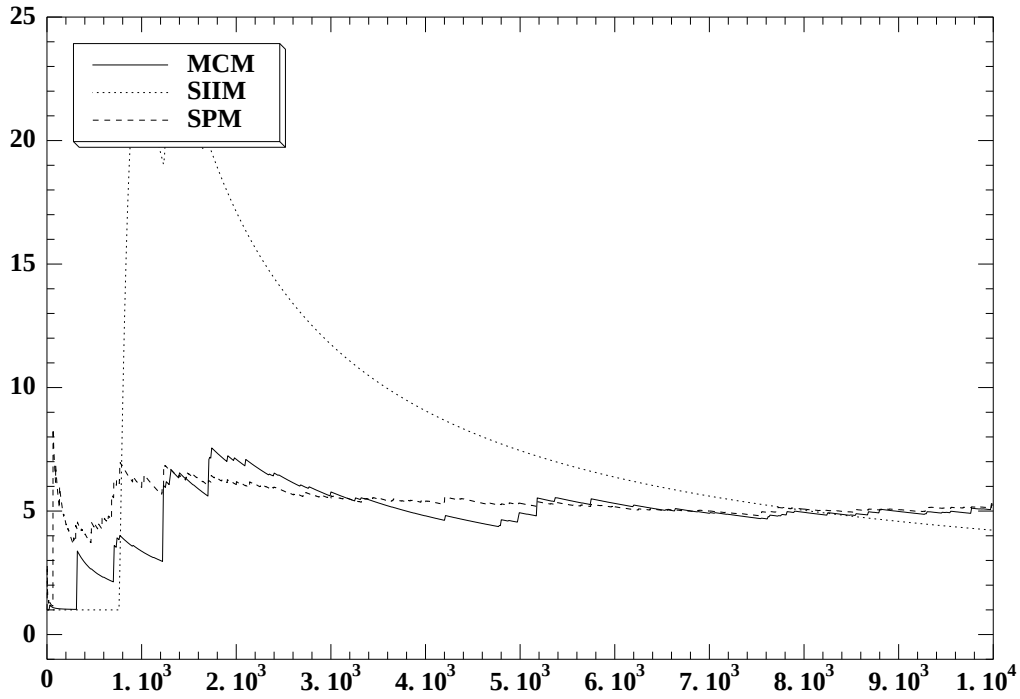
MCM
SIIM
SPM
25
20
15
10
5
0
0
1. 10^3
2. 10^3
3. 10^3
4. 10^3
5. 10^3
6. 10^3
7. 10^3
8. 10^3
9. 10^3
1. 10^4

Figure 5: