An empirical central limit theorem for dependent sequences

Jérôme DEDECKER*and Clémentine PRIEUR†

Abstract

We prove a central limit theorem for the d-dimensional distribution function of a class of stationary sequences. The conditions are expressed in terms of some coefficients which measure the dependence between a given σ-algebra and indicators of quadrants. These coefficients are weaker than the corresponding mixing coefficients, and can be computed in many situations. In particular, we show that they are well adapted to functions of mixing sequences, iterated random functions, and a class of dynamical systems.

Mathematics Subject Classifications (2000): 60F17, 60G10, 62G30.
Key words: empirical distribution function, central limit theorem, dependence coefficients, mixing, dynamical systems.
Short title: Empirical CLT for dependent sequences

1 Introduction

In 1952, Donsker proved the weak convergence of the empirical distribution function of iid random variables to a Brownian bridge, which provides as a straightforward consequence the asymptotic behavior of the Kolmogorov-Smirnov statistics. Dudley (1966) clarified the notion of weak convergence on nonseparable metric spaces, and obtained a central limit theorem for the d-dimensional empirical distribution function.

An early result of Billingsley (1968) extended Donsker's theorem to ϕ-mixing sequences in the sense of Ibragimov (1962), provided that $\sum_{k>0} k^2\sqrt{\phi(k)} < \infty$. In 1979, Yoshihara obtained the same result for α-mixing sequences in the sense of Rosenblatt (1956) under the condition $\alpha(n) = O(n^{-a})$ for some $a > 3$, and

*Laboratoire de Statistique Théorique et Appliquée, Université Paris 6, Site Chevaleret, 13 rue Clisson, 75013 Paris, France. E-mail: dedecker@ccr.jussieu.fr

†GMM, LSP, INSA, 135, avenue de Rangueil, F-31077 Toulouse Cedex 04, France. E-mail: prieur@cict.fr

Dhompongsa (1984) proved the weak convergence of the d-dimensional empirical distribution function provided $\alpha(n) = O(n^{-a})$ for some $a > 2 + d$. The rate given by Yoshihara has been first improved by Shao and Yu (1996) to $a > 1 + \sqrt{2}$ and next by Rio (2000, Theorem 7.2) to $a > 1$. In fact, in Theorem 7.3 of his book, Rio (2000) has shown that the rate $\alpha(n) = O(n^{-a})$ for some $a > 1$ is sufficient for the weak convergence of the d-dimensional empirical distribution function. This last result is remarkable, for the rate of mixing does not depend on the dimension d. In the case of β-mixing sequences in the sense of Rozanov and Volkonskii (1959), Rio (1998 Theorem 1, 2000 Corollary 8.2) obtained the slightly better condition $\sum_{k>0} \beta(k) < \infty$, as a consequence of more general results for classes of functions.

Unfortunately, mixing is a rather restrictive condition, and many simple Markov chains are not mixing. For instance, if $(\epsilon_i)_{i\geq 1}$ is iid with marginal $\mathcal{B}(1/2)$, then the stationary solution $(X_i)_{i\geq 0}$ of the equation

$$X_n = \frac{1}{2}(X_{n-1} + \epsilon_n)\,, \quad X_0 \text{ independent of } (\epsilon_i)_{i\geq 1} \tag{1.1}$$

is not α-mixing (more precisely $\alpha(\sigma(X_0), \sigma(X_n)) = 1/4$ for any n). This example is not an exception: the chain satisfying (1.1) is the Markov chain associated to the dynamical system generated by the map $T(x) = 2x - [2x]$ on the space $[0, 1]$ equipped with the Lebesgue measure, and it is well known that such dynamical systems are not α-mixing in the sense that $\alpha(\sigma(T), \sigma(T^n))$ does not tend to zero as n tends to infinity. Once again, the first work to mention in this framework is that of Billingsley (1968, Theorem 2.2), who proved an empirical central limit theorem for functions of ϕ-mixing processes. Functions of mixing processes cover a large class of examples, such as linear processes or more general time series, as well as certain dynamical systems. More precisely, assume that T is a map from $[0, 1]$ to $[0, 1]$, with finite partition $\{I_1, \ldots, I_N\}$ of $[0, 1]$ into intervals of continuity and monoticity of T such that $|T'| \geq \lambda > 1$ on any interval of the partition. If moreover the absolutely continuous T-invariant probability measure μ is unique and (T, μ) is weakly mixing, Hofbauer and Keller (1982) have proved that the label process defined by $\xi_n(x) = i$ if $T^n(x) \in I_i$, is β-mixing with exponential mixing rate, and $T^n = f((\xi_i)_{i\geq n})$ for some measurable f.

In their Theorem 5, Borovkova, Burton and Dehling (2001) obtained the weak convergence of the empirical distribution function for functions $X_k = f((\xi_{i+k})_{i\in\mathbb{Z}})$ of β-mixing processes. Their assumption is two-part: a rate of mixing on the underlying sequence $(\xi_i)_{i\in\mathbb{Z}}$ and a condition involving the rate of convergence to zero of the quantities $a_l = \|f((\xi_i)_{i\in\mathbb{Z}}) - \mathbb{E}(f((\xi_i)_{i\in\mathbb{Z}})|\sigma(\xi_j, |j| \leq l))\|_1$. Many examples are given in this paper. In particular, the result applies to the empirical distribution function of the sequence $X_i = T^i$ on the probability space $([0, 1], \mu)$, when T is an expanding map as described in the paper by Hofbauer and Keller (1982) (in that case a_l decreases to zero with an exponential rate).

The example of uniformly expanding maps T of $[0,1]$ has been also studied by Collet, Martinez and Schmitt (2004). As in Hofbauer and Keller (1982), they assume that there is a finite partition of $[0,1]$ into intervals of continuity and monoticity of T, but the condition on T is weakened to: there exist $A > 0$ and $\lambda > 1$ such that $|(T^n)'| \geq A\lambda^n$ for any positive integer n, on any interval of the partition associated to T^n. If moreover the absolutely continuous T-invariant probability measure μ is unique and (T, μ) is weakly mixing, they have proved the weak convergence of the empirical distribution function of the sequence $X_i = T^i$ on the probability space $([0,1], \mu)$. Their proof is classical: they use the spectral properties of the adjoint operator of T to derive some covariance inequalities as well as some appropriate moments inequalities, from which both finite dimensional convergence and tightness follow. Note that in this context, it would have been much easier to apply a result given in Doukhan and Louhichi (1999), which we shall recall hereafter.

The idea of Doukhan and Louhichi (1999) is simple: since the only functions we want to control are indicators of half lines, a dependence condition involving only the functions $x \to \mathbb{1}_{x \leq t}$, or at most the differences $f_{s,t}(x) = \mathbb{1}_{x\leq t} - \mathbb{1}_{x \leq s}$, should be enough to obtain an empirical central limit theorem. Starting from a tightness criterion given in Shao and Yu (1996), they proved the weak convergence of the empirical distribution function of a stationary sequence $(X_i)_{i>0}$ of real-valued random variables provided

$$\sup_{s,t\in\mathbb{R}} \sup_{i\leq j, j+n\leq k\leq l} |\mathrm{Cov}(f_{s,t}(X_i)f_{s,t}(X_j), f_{s,t}(X_k)f_{s,t}(X_l))| = \mathcal{O}(n^{-a}) \text{ for } a > 5/2. \quad (1.2)$$

Note that (1.2) is satisfied by the class of expanding maps T considered in Collet, Martinez and Schmitt (2004): more precisely, using the spectral properties of the adjoint operator of T in the space of bounded-variation functions, one can easily see that the decay of the correlations is exponential (see Section 6.3 for more details). In fact (1.2) is satisfied for many other non mixing processes, as one can see from the examples of Section 6.

However, when applied to strongly mixing sequences, the condition (1.2) leads to the rate $\alpha(n) = n^{-a}$ for some $a > 5/2$: it is better than Yoshihara's (1979), but clearly worse than Rio's (2000). Two natural questions are: can we obtain a better rate than Doukhan and Louhichi, for some measure of dependence based on indicator of half lines? If yes, can we obtain similar results for the d-dimensional empirical distribution function? Let us now describe one of the main results of this paper. Let $(X_i)_{i\in\mathbb{Z}}$ be a stationary sequence of $\mathbb{R}^d$-valued random variables with common distribution function $F(t) = \mathbb{P}(X_0 \leq t)$ (as usual $x \leq t$ if and only if $x^{(i)} \leq t^{(i)}$ for any $1 \leq i \leq d$), and let $\mathcal{M}_0 = \sigma(X_i, i \leq 0)$. Define the function

$g_t(x) = \mathbb{1}_{x \le t} - F(t)$. If there exists $a > 1$ such that

$$\sup_{k \ge n} \Big\| \sup_{t \in \mathbb{R}} |\mathbb{E}(g_t(X_k)|\mathcal{M}_0)| \Big\|_\infty = \mathcal{O}(n^{-a}) \quad \text{and}$$

$$\sup_{k > l \ge n} \Big\| \sup_{s,t \in \mathbb{R}} |\mathbb{E}(g_t(X_k)g_s(X_l)|\mathcal{M}_0) - \mathbb{E}(g_t(X_k)g_s(X_l))| \Big\|_\infty = \mathcal{O}(n^{-a}) \,, \tag{1.3}$$

then we can prove a central limit theorem for the d-dimensional empirical distribution function in the space of bounded functions from $\mathbb{R}^d$ to $\mathbb{R}$ equipped with the uniform norm. Here, as in Rio's (2000) result for strongly mixing sequences, the rate $\mathcal{O}(n^{-a})$ for $a > 1$ does not depend on the dimension d. Condition (1.3) is a consequence of a more general result given in Theorem 1: one can take any other $\mathbb{L}^p$-norm instead of the $\mathbb{L}^\infty$-norm in (1.3), but the rate will depend on p and on the dimension d. For instance, for $p = 1$, we obtain the rate $\mathcal{O}(n^{-a})$ for $a > 2d$, which is again better than the rate given by (1.2) in the one-dimensional case.

As one can see from (1.3), the dependence coefficients which appear are generalization of the coefficients introduced in Dedecker and Prieur (2005): the difference is that we need to control the dependence between two points (X_k, X_l) in the future and the σ-algebra $\mathcal{M}_0$. Nevertheless, we shall see in Section 6 that all the examples given in our 2005 paper may be handled similarly for these new coefficients. The main tools for the proof of the empirical central limit theorem are a Rosenthal-type inequality given in Dedecker (2001) and a new tightness criterion inspired from that given in Andrews and Pollard (1994). As in Theorem 8.4 in Rio (2000), another important point is to control the size of the class $\mathcal{F} = \{\mathbb{1}_{x \le t}, t \in \mathbb{R}^d\}$ with respect to an appropriate measure Q related to the dependence structure of the sequence $(X_i)_{i \in \mathbb{Z}}$ (see equation (4.14) and inequality (4.15)).

2 Definitions

We first introduce the following dependence coefficients:

Definition 1 *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, let $\mathcal{M}$ be a sub σ-algebra of $\mathcal{A}$, and let d be a given positive integer. Let $X = (X_1, \ldots, X_k)$ be a random variable with values in $\mathbb{R}^{kd}$. Let $\mathbb{P}_X$ be the distribution of X and let $\mathbb{P}_{X|\mathcal{M}}$ be a conditional distribution of X given $\mathcal{M}$. For $1 \le i \le k$ and t in $\mathbb{R}^d$, let $g_{t,i}(x) = \mathbb{1}_{x \le t} - \mathbb{P}(X_i \le t)$, where $x \le t$ means that $x^{(j)} \le t^{(j)}$ for any $1 \le j \le d$. Define the random variable*

$$b(\mathcal{M}, X_1, \ldots, X_k) = \sup_{(t_1, \ldots, t_k) \in \mathbb{R}^{kd}} \Big| \int \prod_{i=1}^k g_{t_i,i}(x_i) \mathbb{P}_{X|\mathcal{M}}(dx) - \int \prod_{i=1}^k g_{t_i,i}(x_i) \mathbb{P}_X(dx) \Big| \,.$$

with $\mathbb{P}_{X|\mathcal{M}}(dx) = \mathbb{P}_{X|\mathcal{M}}(dx_1, \ldots, dx_k)$ and $\mathbb{P}_X(dx) = \mathbb{P}_X(dx_1, \ldots, dx_k)$. For any p in $[1, \infty]$, define the coefficient

$$\beta_p(\mathcal{M}, X_1, \ldots, X_k) = \| b(\mathcal{M}, X_1, \ldots, X_k) \|_p$$

For $p=1$ or ∞, we shall use the notations $\beta_1(\mathcal{M}, X_1, \ldots, X_k) = \beta(\mathcal{M}, X_1, \ldots, X_k)$ and $\beta_\infty(\mathcal{M}, X_1, \ldots, X_k) = \phi(\mathcal{M}, X_1, \ldots, X_k)$.

Let $\Lambda_1(\mathbb{R}^{kd})$ be the space of functions f satisfying

$$|f(x_1, \ldots, x_{kd}) - f(y_1, \ldots, y_{kd})| \le \sum_{i=1}^{kd} |x_i - y_i| \, .$$

Let $p \ge 1$ and assume that $X_i^{(j)}$ belongs to $\mathbb{L}^p(\mathbb{P})$ for any $1 \le j \le d$ and any $1 \le i \le k$. Define the coefficient

$$\tau_p(\mathcal{M}, X_1, \ldots, X_k) = \Big\| \sup\Big\{ \Big| \int f(x) \mathbb{P}_{X|\mathcal{M}}(dx) - \int f(x) \mathbb{P}_X(dx) \Big|, f \in \Lambda_1(\mathbb{R}^{kd}) \Big\} \Big\|_p \, .$$

The coefficients τ_1 and τ_∞ have been introduced and studied in Dedecker and Prieur (2005), Section 7. Note that the coupling properties of these coefficients, given in Section 7.1 of our 2005 paper, follow immediately from Proposition 6 in Rüschendorf (1985) (the reference to this article is clearly missing in our 2005 paper).

In the particular case where $d=1$, the coefficients β_p can be defined *via* some appropriate function spaces.

Proposition 1 *Let BV_1 be the space of left continuous functions f whose bounded variation norm is smaller than 1, that is df is a signed measure such that $\|df\| = \sup\{|df(g)|, \|g\|_\infty \le 1\} \le 1$. Let $X = (X_1, \ldots, X_k)$ be a random variable with values in $\mathbb{R}^k$. If f is a function in BV_1, let $f^{(i)}(x) = f(x) - \mathbb{E}(f(X_i))$. Keeping the same notations as in Definition 1, we have the equality*

$$b(\mathcal{M}, X_1, \ldots, X_k) = \sup_{f_1, \ldots, f_k \in BV_1} \Big| \int \prod_{i=1}^k f_i^{(i)}(x_i) \mathbb{P}_{X|\mathcal{M}}(dx) - \int \prod_{i=1}^k f_i^{(i)}(x_i) \mathbb{P}_X(dx) \Big| \, .$$

Proof of Proposition 1. Assume without loss that $f_i(-\infty) = 0$. Then

$$f_i^{(i)}(x) = -\int \left(\mathbb{1}_{x \le t} - \mathbb{P}(X_i \le t) \right) df_i(t)$$

Hence, with the notations of Definition 1,

$$\int \prod_{i=1}^k f_i^{(i)}(x_i) \mathbb{P}_{X|\mathcal{M}}(dx) = (-1)^k \int \Big(\int \prod_{i=1}^k g_{t_i,i}(x_i) \mathbb{P}_{X|\mathcal{M}}(dx) \Big) \prod_{i=1}^k df_i(t_i) \, ,$$

and the same is true for $\mathbb{P}_X$ instead of $\mathbb{P}_{X|\mathcal{M}}$. From these inequalities and the fact that $|df_i|(\mathbb{R}) \le 1$, we infer that

$$\sup_{f_1, \ldots, f_k \in BV_1} \Big| \int \prod_{i=1}^k f_i^{(i)}(x_i) \mathbb{P}_{X|\mathcal{M}}(dx) - \int \prod_{i=1}^k f_i^{(i)}(x_i) \mathbb{P}_X(dx) \Big| \le b(\mathcal{M}, X_1, \ldots, X_k)$$

The converse inequality follows by noting that $x \to \mathbb{1}_{x \le t}$ belongs to BV_1. $\square$

We now define the coefficients $\beta_{k,p}$ and $\tau_{k,p}$ for a sequence of σ-algebras and a sequence of $\mathbb{R}^d$-valued random variables.

Definition 2 *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let $(X_i)_{i\geq 0}$ be a sequence of random variables with values in $\mathbb{R}^d$, and let $(\mathcal{M}_i)_{i\geq 0}$ be a sequence of σ-algebras of $\mathcal{A}$. For any $p \geq 1$, $k \in \mathbb{N}^* \cup \{\infty\}$ and $n \geq 1$, define*

$$\begin{aligned}\beta_{k,p}(n) &= \max_{1\leq l\leq k} \sup_{i+n\leq j_1<\cdots<j_l} \beta_p(\mathcal{M}_i, X_{j_1}, \ldots, X_{j_l}) \\ \tau_{k,p}(n) &= \max_{1\leq l\leq k} \frac{1}{l} \sup_{i+n\leq j_1<\cdots<j_l} \tau_p(\mathcal{M}_i, X_{j_1}, \ldots, X_{j_l}) \,.\end{aligned}$$

For $p = 1$ or ∞, we shall use the notation $\beta_k(n) = \beta_{k,1}(n)$ and $\phi_k(n) = \beta_{k,\infty}(n)$.

3 Results

Let $(X_i)_{i\in\mathbb{Z}}$ be a stationary sequence of random variables with values in $\mathbb{R}^d$ and common distribution function F, and let $\mathcal{M}_i = \sigma(X_j, j \leq i)$. Let F_n be the empirical distribution function: for any t in $\mathbb{R}^d$, $F_n(t) = n^{-1}\sum_{i=1}^n \mathbb{1}_{X_i\leq t}$. In Theorem 1 below, we give sufficient conditions for the process $\{\sqrt{n}(F_n(t) - F(t)), t \in \mathbb{R}^d\}$ to converge in distribution to a tight Gaussian process on the space $\ell^\infty(\mathbb{R}^d)$ of bounded functions from $\mathbb{R}^d$ to $\mathbb{R}$ equipped with the uniform norm $|\cdot|_\infty$ (for more details on weak convergence on the non separable space $\ell^\infty(\mathbb{R}^d)$, we refer to van der Vaart and Wellner (1996); in particular, we shall not discuss any measurability problems, which can be handled by using the outer probability).

Recall that a random variable X with values in $\ell^\infty(\mathbb{R}^d)$ is tight if for any positive ϵ, there exists a compact set K_ϵ of $(\ell^\infty(\mathbb{R}^d), |\cdot|_\infty)$ such that $\mathbb{P}(X \in K_\epsilon) \geq 1 - \epsilon$. A random variable G with values in $\ell^\infty(\mathbb{R}^d)$ is a gaussian process if every one of its finite dimensional marginals $(G(t_1), \ldots, G(t_k))$ is normally distributed. If G is a tight gaussian process then it is also Gaussian as an $\ell^\infty(\mathbb{R}^d)$-valued random variable: for every element d of the dual of $\ell^\infty(\mathbb{R}^d)$, the real-valued random variable $d(G)$ is normally distributed.

Consider the two conditions :

$(\mathcal{C}_1)$ There exist ε in $]0, 1]$ and $p' > d(2 + \varepsilon)/2\varepsilon$ such that $\beta_{2,p'}(k) = \mathcal{O}(k^{-1-\varepsilon})$.

$(\mathcal{C}_2)$ There exists $\varepsilon > 0$ such that $\sum_{k=1}^{+\infty} k\, \beta_{2,d+\varepsilon}(k) < +\infty$.

Theorem 1 *If either $(\mathcal{C}_1)$ or $(\mathcal{C}_2)$ holds, then $\{\sqrt{n}(F_n(t)-F(t))\,,\, t \in \mathbb{R}^d\}$ converges weakly in $\ell^\infty(\mathbb{R}^d)$ to a tight Gaussian process with covariance function*

$$\Gamma(t, s) = \sum_{k\in\mathbb{Z}} \mathrm{Cov}(\mathbb{1}_{X_0\leq t}, \mathbb{1}_{X_k\leq s})\,. \tag{3.1}$$

In the next proposition, we give sufficient conditions for $(\mathcal{C}_1)$ and $(\mathcal{C}_2)$ to hold, in terms of the coefficients ϕ_2, β_2, $\tau_{2,\infty}$ and $\tau_{2,1}$. Consider the conditions $(\mathcal{C}_3)$, $(\mathcal{C}_4)$, $(\mathcal{C}_5)$ and $(\mathcal{C}_6)$:

($\mathcal{C}_3$) There exists $\varepsilon > 0$ such that $\phi_2(k) = \mathcal{O}\left(k^{-1-\varepsilon}\right)$.

($\mathcal{C}_4$) There exists $\varepsilon > 0$ such that $\beta_2(k) = \mathcal{O}\left(k^{-2d-\varepsilon}\right)$.

($\mathcal{C}_5$) Each component of X_1 has a bounded density and there exists $\varepsilon > 0$ such that $\tau_{2,\infty}(k) = \mathcal{O}\left(k^{-2-\varepsilon}\right)$.

($\mathcal{C}_6$) Each component of X_1 has a bounded density and there exists $\varepsilon > 0$ such that $\tau_{2,1}(k) = \mathcal{O}\left(k^{-4d-\varepsilon}\right)$.

Proposition 2 *The following implications hold:*

$$(\mathcal{C}_5) \Rightarrow (\mathcal{C}_3) \Rightarrow (\mathcal{C}_1) \quad \textit{and} \quad (\mathcal{C}_6) \Rightarrow (\mathcal{C}_4) \Rightarrow (\mathcal{C}_2).$$

Remark 1. In this remark, we shall discuss on the optimality of the conditions ($\mathcal{C}_3$) and ($\mathcal{C}_4$). This is a delicate matter, because very few is known on this subject. A reasonable conjecture is **C**: the minimal conditions in terms of the coefficients β_2 (resp. ϕ_2) to obtain the weak convergence of the empirical distribution function are the minimal conditions in terms of the coefficients β_2 (resp. ϕ_2) to obtain the central limit theorem for bounded random variables.

For the coefficient $\beta_2(i)$, the minimal condition to obtain the central limit theorem for bounded random variables is $\sum_{k>0} \beta_2(k) < \infty$ (for the optimality, use Theorem 4 in Bradley (1997) and the fact that $\beta(\mathcal{M}, X_1, X_2) \leq \beta(\mathcal{M}, \sigma(X_1, X_2))$, according to Proposition 9 of Section 5). If the conjecture **C** is true, the condition ($\mathcal{C}_4$) is not optimal, even if $d = 1$. By contrast Rio's result (1998) for β-mixing sequences in the sense of Rozanov and Volkonskii (1959) is optimal for any d.

For the coefficient $\phi_2(i)$, we do not know what is the minimal condition to obtain the central limit theorem for bounded random variables. What we can prove is that it holds provided that $\sum_{k>0} k^{-1/2}\phi_2(k) < \infty$. It follows that, if the conjecture **C** is true, the condition ($\mathcal{C}_3$) is not optimal. However, in that case, the loss does not increase with the dimension.

Remark 2. If $d = 1$, Prieur (2002) proved an empirical central limit theorem under a condition on the s-weak dependence coefficient $\theta(i)$, which is similar (but weaker) to our coefficient $\tau_{\infty,1}(i)$. From Prieur's result, we infer that if X_0 has a bounded density and $\tau_{\infty,1}(n) = \mathcal{O}(n^{-2-2\sqrt{2}-\epsilon})$ then the conclusion of Theorem 1 holds. Since $\tau_{2,1}(n) \leq \tau_{\infty,1}(n)$, our condition ($\mathcal{C}_6$) gives a better rate. Moreover, for all the examples studied in Prieur (2002), we can obtain the same bounds for $\tau_{\infty,1}(i)$ as those obtained for $\theta(i)$.

Proof of Proposition 2. The facts that ($\mathcal{C}_5$) $\Rightarrow$ ($\mathcal{C}_3$) and ($\mathcal{C}_6$) $\Rightarrow$ ($\mathcal{C}_4$) follow from Proposition 6 of Section 4. The fact that ($\mathcal{C}_3$) $\Rightarrow$ ($\mathcal{C}_1$) is straightforward. Since the random variable $b(\mathcal{M}, X, Y)$ is almost surely bounded by 1 (see Proposition 9, Section 5), we infer that $\beta_{2,d+\varepsilon}(k) \leq (\beta_2(k))^{1/(d+\varepsilon)}$, so that ($\mathcal{C}_4$) $\Rightarrow$ ($\mathcal{C}_2$). □

4 Proof of Theorem 1

We first recall two moment inequalities given in Dedecker [9].

Proposition 3 *Let $(X_i)_{i\in\mathbb{Z}}$ be a stationary sequence of centered and square integrable random variables and let $S_n = X_1 + \cdots + X_n$. Let $\mathcal{M}_i = \sigma(X_j, j \le i)$. The following upper bound holds*

$$\|S_n\|_p \le (pnV_\infty)^{1/2} + \Big(3p^2 n\Big(\|X_0^3\|_{p/3} + M_1(p) + M_2(p) + M_3(p)\Big)\Big)^{1/3},$$

where

$$\begin{aligned}
V_N &= \mathbb{E}(X_0^2) + 2\sum_{k=1}^{N} |\mathbb{E}(X_0X_k)| \\
M_1(p) &= \sum_{l=1}^{+\infty}\sum_{m=0}^{l-1} \|X_0X_m\mathbb{E}(X_{l+m}|\mathcal{M}_m)\|_{p/3} \\
M_2(p) &= \sum_{l=1}^{+\infty}\sum_{m=l}^{+\infty} \|X_0\mathbb{E}(X_mX_{l+m} - \mathbb{E}(X_mX_{l+m})|\mathcal{M}_0)\|_{p/3} \\
M_3(p) &= \frac{1}{2}\sum_{k=1}^{+\infty} \|X_0\mathbb{E}(X_k^2 - \mathbb{E}(X_k^2)|\mathcal{M}_0)\|_{p/3}\,.
\end{aligned}$$

Proposition 4 *We keep the same notations as in Proposition 3. For any positive integer N, the following upper bound holds*

$$\|S_n\|_p \le (pn\,(V_{N-1} + 2M_0(p)))^{1/2} + \Big(3p^2 n\Big(\|X_0^3\|_{p/3} + \tilde{M}_1(p) + \tilde{M}_2(p) + M_3(p)\Big)\Big)^{1/3},$$

where

$$\begin{aligned}
M_0(p) &= \sum_{l=N}^{+\infty} \|X_0\mathbb{E}(X_l|\mathcal{M}_0)\|_{p/2} \\
\tilde{M}_1(p) &= \sum_{l=1}^{N-1}\sum_{m=0}^{l-1} \|X_0X_m\mathbb{E}(X_{l+m}|\mathcal{M}_m)\|_{p/3} \\
\tilde{M}_2(p) &= \sum_{l=1}^{N-1}\sum_{m=l}^{+\infty} \|X_0\mathbb{E}(X_mX_{l+m} - \mathbb{E}(X_mX_{l+m})|\mathcal{M}_0)\|_{p/3}\,.
\end{aligned}$$

Next we recall the notion of number of brackets.

Definition 3 *Let Q be a finite measure on a measurable space $\mathcal{X}$. For any measurable function f from $\mathcal{X}$ to $\mathbb{R}$, let $\|f\|_{Q,r} = Q(|f|^r)^{1/r}$. If $\|f\|_{Q,r}$ is finite, one says that f belongs to L^r_Q. Let $\mathcal{F}$ be some subset of L^r_Q. The number of brackets $\mathcal{N}_{Q,r}(\varepsilon, \mathcal{F})$ is the smallest integer N for which there exist some functions $f_1^- \le f_1, \ldots, f_N^- \le f_N$ in $\mathcal{F}$ such that: for any integer $1 \le i \le N$ we have $\|f_i - f_i^-\|_{Q,r} \le \varepsilon$, and for any function f in $\mathcal{F}$ there exists an integer $1 \le i \le N$ such that $f_i^- \le f \le f_i$.*

Before proving Theorem 1 we state an uniform law of large numbers under a bracketing condition.

Proposition 5 *Let $(X_i)_{i>0}$ be a sequence of identically distributed random variables with values in some measurable space $\mathcal{X}$, with common marginal distribution P. Assume that for any f in $\mathcal{F}$, $n^{-1}\sum_{i=1}^n f(X_i)$ converges almost surely to $P(f)$. If $\mathcal{N}_{P,1}(x,\mathcal{F})$ is finite for any $x>0$, then*

$$\lim_{n\to\infty}\sup_{f\in\mathcal{F}}|P_n(f)-P(f)|=0 \quad \mathbb{P}\, p.s.$$

Proof of Proposition 5. The same as for i.i.d. sequences (see for instance van der Vaart and Wellner (1996), proof of Theorem 2.4.1 page 122). □

The main step in the proof of Theorem 1 is the following proposition, whose proof is based on a decomposition given in Andrews and Pollard (1994) (see also Louhichi (2000)).

Proposition 6 *Let $(X_i)_{i\geq 1}$ be a sequence of identically distributed random variables with values in a measurable space $\mathcal{X}$, with common distribution P. Let P_n be the empirical measure $P_n=n^{-1}\sum_{i=1}^n \delta_{X_i}$, and let Z_n be the normalized empirical process $Z_n=\sqrt{n}(P_n-P)$. Let Q be any finite measure on $\mathcal{X}$ such that $Q-P$ is a positive measure. Let $\mathcal{F}$ be a class of functions from $\mathcal{X}$ to $\mathbb{R}$ and $\mathcal{G}=\{f-l,(f,l)\in\mathcal{F}\times\mathcal{F}\}$. Assume that there exist $r\geq 2$, $p\geq 1$ and $q>2$ such that for any function g of $\mathcal{G}$, we have*

$$\|Z_n(g)\|_p\leq C(\|g\|_{Q,1}^{1/r}+n^{1/q-1/2})\,,$$

where the constant C does not depend on g nor n. If moreover

$$\int_0^1 x^{(1-r)/r}(\mathcal{N}_{Q,1}(x,\mathcal{F}))^{1/p}dx<\infty \quad \text{and} \quad \lim_{x\to 0} x^{p(q-2)/q}\mathcal{N}_{Q,1}(x,\mathcal{F})=0\,,$$

then

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{E}\Big(\sup_{g\in\mathcal{G},\|g\|_{Q,1}\leq\delta}|Z_n(g)|^p\Big)=0\,. \tag{4.1}$$

Proof of Proposition 6. It follows the line of Andrews and Pollard (1994) and Louhichi (2000). It is based on the following inequality: given N real-valued random variables, we have

$$\|\max_{1\leq i\leq N}|Z_i|\|_p\leq N^{1/p}\max_{1\leq i\leq N}\|Z_i\|_p\,. \tag{4.2}$$

For any positive integer k, denote by $\mathcal{N}_k=\mathcal{N}_{Q,1}(2^{-k},\mathcal{F})$ and by $\mathcal{F}_k$ a family of functions $f_1^{k,-}\leq f_1^k,\ldots,f_{\mathcal{N}_k}^{k,-}\leq f_{\mathcal{N}_k}^k$ in $\mathcal{F}$ such that $\|f_i^k-f_i^{k,-}\|_{Q,1}\leq 2^{-k}$, and for any f in $\mathcal{F}$, there exists an integer $1\leq i\leq\mathcal{N}_k$ such that $f_i^{k,-}\leq f\leq f_i^k$.

First step. We shall construct a sequence $h_{k(n)}(f)$ belonging to $\mathcal{F}_{k(n)}$ such that

$$\lim_{n\to\infty}\Big\|\sup_{f\in\mathcal{F}}|Z_n(f)-Z_n(h_{k(n)}(f))|\Big\|_p=0\,. \tag{4.3}$$

For any f in $\mathcal{F}$, there exist two functions g_k^- and g_k^+ in $\mathcal{F}_k$ such that $g_k^- \leq f \leq g_k^+$ and $\|g_k^+ - g_k^-\|_{Q,1} \leq 2^{-k}$. Since $Q - P$ is a positive measure, we have the inequalities

$$\begin{aligned} Z_n(f) - Z_n(g_k^-) &\leq Z_n(g_k^+) - Z_n(g_k^-) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{E}((g_k^+ - f)(X_i)) \\ &\leq |Z_n(g_k^+) - Z_n(g_k^-)| + \sqrt{n} 2^{-k} \,. \end{aligned}$$

Since $g_k^- \leq f$, we also have that $Z_n(g_k^-) - Z_n(f) \leq \sqrt{n}2^{-k}$, which enables us to conclude that $|Z_n(f) - Z_n(g_k^-)| \leq |Z_n(g_k^+) - Z_n(g_k^-)| + \sqrt{n}2^{-k}$. Consequently

$$\sup_{f \in \mathcal{F}} |Z_n(f) - Z_n(g_k^-)| \leq \max_{1 \leq i \leq \mathcal{N}_k} |Z_n(f_i^k) - Z_n(f_i^{k,-})| + \sqrt{n} 2^{-k} \,. \tag{4.4}$$

Combining (4.2) and (4.4), we obtain that

$$\Big\| \sup_{f \in \mathcal{F}} |Z_n(f) - Z_n(g_k^-)| \Big\|_p \leq \mathcal{N}_k^{1/p} \max_{1 \leq i \leq \mathcal{N}_k} \|Z_n(f_i^k) - Z_n(f_i^{k,-})\|_p + \sqrt{n} 2^{-k} \,. \tag{4.5}$$

Starting from (4.5) and applying the inequality of Proposition 6, we obtain

$$\Big\| \sup_{f \in \mathcal{F}} |Z_n(f) - Z_n(g_k^-)| \Big\|_p \leq C(\mathcal{N}_k^{1/p} 2^{-k/r} + \mathcal{N}_k^{1/p} n^{1/q-1/2}) + \sqrt{n} 2^{-k} \,. \tag{4.6}$$

From the integrability condition on $\mathcal{N}_{Q,1}(x, \mathcal{F})$, and since $x \to x^{(1-r)/r} \mathcal{N}(x, \mathcal{F})^{1/p}$ is non increasing, we infer that $\mathcal{N}_k^{1/p} 2^{-k/r}$ tends to 0 as k tends to infinity. Take $k(n)$ such that $2^{k(n)} = \sqrt{n}/a_n$ for some sequence a_n decreasing to zero. Then $\sqrt{n}2^{-k(n)}$ tends to 0 as n tends to infinity. Il remains to control the second term on right hand in (4.6). By definition of $\mathcal{N}_{k(n)}$, we have that

$$\mathcal{N}_{k(n)} n^{p(1/q-1/2)} = \mathcal{N}_{Q,1}\Big(\frac{a_n}{\sqrt{n}}, \mathcal{F}\Big) \Big(\frac{1}{\sqrt{n}}\Big)^{p(q-2)/q} \,. \tag{4.7}$$

Since $x^{p(q-2)/p} \mathcal{N}_{Q,1}(x, \mathcal{F})$ tends to 0 as x tends to zero, we can find a sequence a_n such that the right hand term in (4.7) converges to 0. The function $h_{k(n)}(f) = g_{k(n)}^-$ satisfies (4.3).

Second step. We shall prove that for any $\epsilon > 0$ and n large enough, there exists a function $h_m(f)$ in $\mathcal{F}_m$ such that

$$\Big\| \sup_{f \in \mathcal{F}} |Z_n(h_m(f)) - Z_n(h_{k(n)}(f))| \Big\|_p \leq \epsilon \,. \tag{4.8}$$

Given h in $\mathcal{F}_k$, choose a function $T_{k-1}(h)$ in $\mathcal{F}_{k-1}$ such that $\|h - T_{k-1}(h)\|_{Q,1} \leq 2^{-k+1}$. Denote by $\pi_{k,k} = Id$ and for $l < k$, $\pi_{l,k}(h) = T_l \circ \cdots \circ T_{k-1}(h)$. We consider the function $h_m(f) = \pi_{m,k(n)}(h_{k(n)}(f))$. We have that

$$\Big\| \sup_{f \in \mathcal{F}} |Z_n(h_m) - Z_n(h_{k(n)})| \Big\|_p \leq \sum_{l=m+1}^{k(n)} \Big\| \sup_{f \in \mathcal{F}} |Z_n(\pi_{l,k(n)}(h_{k(n)}) - Z_n(\pi_{l-1,k(n)}(h_{k(n)})| \Big\|_p \,. \tag{4.9}$$

Clearly

$$\Big\|\sup_{f\in\mathcal{F}}|Z_n(\pi_{l,k(n)}(h_{k(n)}) - Z_n(\pi_{l-1,k(n)}(h_{k(n)})|\Big\|_p \leq \Big\|\max_{f\in\mathcal{F}_l}|Z_n(f) - Z_n(T_{l-1}(f))|\Big\|_p .$$

Applying the inequality of Proposition 6 to (4.9) we obtain

$$\Big\|\sup_{f\in\mathcal{F}}|Z_n(h_m) - Z_n(h_{k(n)})|\Big\|_p \leq C\sum_{l=m+1}^{k(n)}(2^{1/r}\mathcal{N}_l^{1/p}2^{-l/r} + \mathcal{N}_l^{1/p}n^{1/q-1/2})$$

Clearly

$$\sum_{l=m+1}^{\infty}\mathcal{N}_l^{1/p}2^{-l/r} \leq 2\int_0^{2^{-m-1}} x^{(1-r)/r}(\mathcal{N}_{Q,1}(x,\mathcal{F}))^{1/p}dx\,,$$

which by assumption is as small as we wish. To control the second term, write

$$n^{1/q-1/2}\sum_{l=m+1}^{k(n)}\mathcal{N}_l^{1/p} \leq n^{1/q-1/2}\sum_{l=0}^{k(n)}2^l\mathcal{N}_l^{1/p}2^{-l} \leq 2n^{1/q-1/2}\int_{2^{-k(n)}}^1\frac{1}{x}(\mathcal{N}_{Q,1}(x,\mathcal{F}))^{1/p}dx\,.$$

It is easy to see that if $x^{p(q-2)/q}\mathcal{N}_{Q,1}(x,\mathcal{F})$ tends to 0 as x tends to 0, then

$$\lim_{x\to 0}x^{(q-2)/q}\int_x^1\frac{1}{y}(\mathcal{N}_{Q,1}(y,\mathcal{F}))^{1/p}dy = 0\,.$$

Consequently, we can choose the decreasing sequence a_n such that

$$\lim_{n\to\infty}\Big(\frac{1}{\sqrt{n}}\Big)^{(q-2)/q}\int_{a_n n^{-1/2}}^1\frac{1}{x}(\mathcal{N}_{Q,1}(x,\mathcal{F}))^{1/p}dx = 0\,.$$

The function $h_m(f) = \pi_{m,k(n)}(h_{k(n)}(f))$ satisfies (4.8).

Third step. From steps 1 and 2, we infer that for any $\epsilon > 0$ and n large enough, there exists $h_m(f)$ in $\mathcal{F}_m$ such that

$$\Big\|\sup_{f\in\mathcal{F}}|Z_n(f) - Z_n(h_m(f))|\Big\|_p \leq 2\epsilon\,.$$

Using the same argument as in Andrews and Pollard (1994) (see the paragraph "Comparison of pairs" page 124), we obtain that, for any f and g in $\mathcal{F}$,

$$\Big\|\sup_{\|f-g\|_{Q,1}\leq\delta}|Z_n(f) - Z_n(g)|\Big\|_p \leq 8\epsilon + \mathcal{N}_m^{2/r}\sup_{\|f-g\|_{Q,1}\leq\delta}\|Z_n(f) - Z_n(g)\|_p\,.$$

We conclude the proof by noting that

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\Big\|\sup_{g\in\mathcal{G},\|g\|_{Q,1}\leq\delta}|Z_n(g)|\Big\|_p \leq 8\epsilon\,. \quad \square$$

Proof of Theorem 1. Let $\mathcal{F} = \{x \to \mathbb{1}_{x\leq t}\,,\ t\in\mathbb{R}^d\}$, and let $\mathcal{G} = \{f - h\,,\ f,h\in\mathcal{F}\}$. Using Theorem 1 of Dedecker and Rio (2000), we get the convergence of the finite dimensional distributions of $\{\sqrt{n}(F_n(t) - F(t))\,,\, t \in \mathbb{R}^d\}$ to that of the Gaussian process with covariance function Γ as soon as the sequence $(X_i)_{i\in\mathbb{Z}}$ is ergodic and $\sum_{i>0} \beta_1(i)$ is finite. To be precise, let $\mathbf{f} = (f_1, \ldots, f_k)$ be an element of $\mathcal{F}^k$ and for any x in $\mathbb{R}^k$ let $< x, \mathbf{f} - P(\mathbf{f}) >= x_1(f_1 - P(f_1)) + \cdots + x_k(f_k - P(f_k))$. Define the matrix C by $C_{i,j} = \Gamma(t_i, t_j)$, where t_i is such that $f_i = \mathbb{1}_{x\le t_i}$. Since $(X_i)_{i\in\mathbb{Z}}$ is ergodic, we infer from Dedecker and Rio (2000) that the random variable $Z_n(< x, \mathbf{f} - P(\mathbf{f}) >)$ converges in distribution to a mean-zero normal distribution with variance $x^t C x$ as soon as

$$\sum_{i\ge 0} \| < x, \mathbf{f} - P(\mathbf{f}) > (X_0)\mathbb{E}(< x, \mathbf{f} - P(\mathbf{f}) > (X_i)|\mathcal{M}_0)\|_1 < \infty\,. \tag{4.10}$$

Consequently, if (4.10) holds, the random vector $(Z_n(f_1), \ldots, Z_n(f_k))$ converges in distribution to a Gaussian vector with covariance matrix C. By Definition 1, we obtain that

$$\begin{aligned}\| < x, \mathbf{f} - P(\mathbf{f}) > (X_0)&\mathbb{E}(< x, \mathbf{f} - P(\mathbf{f}) > (X_i)|\mathcal{M}_0)\|_1 \\ &\le \| < x, \mathbf{f} - P(\mathbf{f}) > (X_0)\|_\infty \Big(\sum_{j=1}^k x_j\Big)\beta_1(\mathcal{M}_0, X_i)\,, \end{aligned}\tag{4.11}$$

so that (4.10) holds as soon as $\sum_{i\ge 0} \beta_1(i)$ is finite.

If we do not assume that the sequence $(X_i)_{i\in\mathbb{Z}}$ is ergodic, the limit may be non Gaussian. However the ergodicity assumption may be dropped by assuming instead that $\sum_{i>0} \beta_2(i) < +\infty$. Let T be the shift operator from $\mathbb{R}^{\mathbb{Z}}$ to $\mathbb{R}^{\mathbb{Z}}$: $(T(x))_i = x_{i+1}$. Let $\mathcal{I}$ be the σ-algebra of T-invariant elements of $\mathcal{B}(\mathbb{R}^{\mathbb{Z}})$. Let $\mathbf{X} := (X_i)_{i\in\mathbb{Z}}$. Since $\beta_2(n)$ tends to 0 as n tends to infinity, we can prove that $\mathbb{E}(f(X_0)g(X_k)|\mathbf{X}^{-1}(\mathcal{I})) = \mathbb{E}(f(X_0)g(X_k))$ for any measurable functions f, g. Once again, we conclude by using Theorem 1 in Dedecker and Rio (2000).

To obtain the weak convergence of the empirical distribution function in the space $\ell^\infty(\mathbb{R}^d)$, it remains to prove that the process $\{Z_n(f), f \in \mathcal{F}\}$ is asymptotically tight, that is there exists a semi metric ρ on $\mathcal{F}$ such that $(\mathcal{F}, \rho)$ is totally bounded, and, for every $\epsilon > 0$,

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{P}\Big(\sup_{\rho(f,g)\le\delta,\, f,g\in\mathcal{F}} |Z_n(f) - Z_n(g)| > \epsilon\Big) = 0\,. \tag{4.12}$$

Since $\mathcal{N}_{Q,1}(x, \mathcal{F}) = \mathcal{O}(x^{-d})$ for any finite measure Q on $\mathbb{R}^d$, the set $(\mathcal{F}, \|\cdot\|_{Q,1})$ is totally bounded. Consequently, the property (4.12) follows from (4.1) by applying Markov's inequality.

Let us prove that condition $(\mathcal{C}_2)$ implies (4.1) for some appropriate measure Q. For any s, t in $\mathbb{R}^d$, let $f_{s,t}(x) = \mathbb{1}_{x\le t} - \mathbb{1}_{x\le s}$ and $\tilde{f}_{s,t}(x) = f_{s,t}(x) - \int f_{s,t}(x)P(dx)$. Applying Proposition 3 to $(\tilde{f}_{s,t}(X_i))_{i\in\mathbb{Z}}$, we obtain that, for any $p \ge 1$,

$$\|Z_n(\tilde{f}_{s,t})\|_p \le \sqrt{pV_\infty} + n^{1/3-1/2}\Big(3p^2(\|\tilde{f}_{s,t}(X_0)^3\|_{p/3} + M_1(p) + M_2(p) + M_3(p))\Big)^{1/3}, \tag{4.13}$$

where V_∞, $M_1(p)$, $M_2(p)$ and $M_3(p)$ are defined in Proposition 3. Define now the measure Q on $\mathcal{X}$ by

$$Q(dx) = B(x)\, P(dx) = (1 + 4 \sum_{k=1}^{+\infty} b_k(x))\, P(dx)\,, \qquad (4.14)$$

where $b_k(x)$ is the function from $\mathbb{R}^d$ to $[0,1]$ such that $b(\sigma(X_0), X_k) = b_k(X_0)$ and P is the law of X_0. Note that Q is finite as soon as $\sum_{k=1}^{+\infty} \beta_1(k)$ is finite. For any k in $\mathbb{N}^*$, we have that

$$\begin{aligned} |\mathrm{Cov}(f_{s,t}(X_0), f_{s,t}(X_k))| &= |\mathbb{E}(f_{s,t}(X_0)\mathbb{E}(\tilde{f}_{s,t}(X_k)|X_0))| \\ &\leq 2\mathbb{E}(|f_{s,t}(X_0) b_k(X_0)|) = 2\int |f_{s,t}(x)| b_k(x) P(dx) \end{aligned}$$

Consequently,

$$V_\infty = \mathrm{Var}(f_{s,t}(X_0)) + 2\sum_{k=1}^{+\infty} |\mathrm{Cov}(f_{s,t}(X_0), f_{s,t}(X_k))| \leq \int |f_{s,t}(x)| Q(dx)\,. \qquad (4.15)$$

In the same way, since $\|\tilde{f}_{s,t}(\cdot)\|_\infty \leq 1$,

$$M_1(p) \leq 2 \sum_{l=1}^{+\infty} \sum_{m=0}^{l-1} \|b(\mathcal{M}_m, X_{l+m})\|_{p/3} = 2 \sum_{k=1}^{+\infty} k\, \beta_{1\,,\,p/3}(k)\,. \qquad (4.16)$$

Since

$$\begin{aligned} \tilde{f}_{s,t}(X_m)\tilde{f}_{s,t}(X_l) = \tilde{f}_{-\infty,t}(X_m)\tilde{f}_{-\infty,t}(X_l) + \tilde{f}_{-\infty,s}(X_m)\tilde{f}_{-\infty,s}(X_l) \\ - \tilde{f}_{-\infty,t}(X_m)\tilde{f}_{-\infty,s}(X_l) - \tilde{f}_{-\infty,s}(X_m)\tilde{f}_{-\infty,t}(X_l)\,, \quad (4.17) \end{aligned}$$

we have that

$$|\mathbb{E}(\tilde{f}_{s,t}(X_m)\tilde{f}_{s,t}(X_l)|\mathcal{M}_0) - \mathbb{E}(\tilde{f}_{s,t}(X_m)\tilde{f}_{s,t}(X_l))| \leq 4b(\mathcal{M}_0, X_m, X_l)\,.$$

Hence,

$$M_2(p) \leq 4 \sum_{l=1}^{+\infty} \sum_{m=l}^{+\infty} \|b(\mathcal{M}_0, X_m, X_{l+m})\|_{p/3} \leq 4 \sum_{k=1}^{+\infty} k\, \beta_{2\,,\,p/3}(k)\,. \qquad (4.18)$$

Applying (4.17) with $m = l = k$, since

$$(\tilde{f}_{-\infty,t}(X_k))^2 - \mathbb{E}((\tilde{f}_{-\infty,t}(X_k))^2) = (1 - 2F(t))\tilde{f}_{-\infty,t}(X_k)\,,$$

and since

$$\begin{aligned} \tilde{f}_{-\infty,t}(X_k)\tilde{f}_{-\infty,s}(X_k) - \mathbb{E}(\tilde{f}_{-\infty,t}(X_k)\tilde{f}_{-\infty,s}(X_k)) \\ = \tilde{f}_{-\infty,s\wedge t}(X_k) - F(t)\tilde{f}_{-\infty,s}(X_k) - F(s)\tilde{f}_{-\infty,t}(X_k)\,, \end{aligned}$$

we have that

$$M_3(p) \;\leq\; 4\sum_{k=1}^{+\infty} \|b(\mathcal{M}_0, X_k)\|_{p/3} = 4\sum_{k=1}^{+\infty} \beta_{1,p/3}(k)\,. \tag{4.19}$$

Since $\|\tilde{f}_{s,t}(X_0)^3\|_{p/3} \leq 1$, we obtain from (4.15), (4.16), (4.18), (4.19) and Proposition 3, that for any g in $\mathcal{G}$,

$$\|Z_n(g)\|_p \leq (p\,\|g\|_{Q,1})^{1/2} + n^{1/3-1/2}\Big(3p^2\Big(1 + 10\sum_{k=1}^{+\infty} k\beta_{2,p/3}(k)\Big)\Big)^{1/3}.$$

We then apply Proposition 6 with $r = 2$ and $q = 3$. Since $\mathcal{N}_{Q,1}(x, \mathcal{F}) = \mathcal{O}(x^{-d})$, we obtain that $\{Z_n(f), f \in \mathcal{F}\}$ is asymptotically tight as soon as $p > 3d$. The result follows.

Let us prove that condition $(\mathcal{C}_1)$ implies (4.1). Applying Proposition 4 to the sequence $(\tilde{f}_{s,t}(X_i))_{i\in\mathbb{Z}}$, we obtain, for any $p \geq 1$,

$$\|Z_n(\tilde{f}_{s,t})\|_p \leq (p(V_\infty + 2M_0(p)))^{1/2} + n^{1/3-1/2}\Big(3p^2\Big(\|\tilde{f}(X_0)^3\|_{p/3} + \tilde{M}_1(p) + \tilde{M}_2(p) + M_3(p)\Big)\Big)^{1/3},$$

where V_∞, $M_0(p)$, $\tilde{M}_1(p)$, $\tilde{M}_2(p)$ and $M_3(p)$ are defined in Proposition 4. It remains to bound $M_0(p)$, $\tilde{M}_1(p)$ and $\tilde{M}_2(p)$. Since $\|\tilde{f}_{s,t}(X_0)\|_\infty \leq 1$,

$$M_0(p) \leq \sum_{l=N}^{+\infty} \beta_{1,p/2}(l) \tag{4.20}$$

In the same way,

$$\tilde{M}_1(p) \leq 2\sum_{k=1}^{N-1} k\,\beta_{1\,,p/3}(k) \tag{4.21}$$

Using (4.17), we obtain that

$$\tilde{M}_2(p) \leq 4\sum_{l=1}^{N-1}\sum_{m=l}^{\infty} \beta_{2\,,p/3}(m) = 4\sum_{k=1}^{\infty} \beta_{2\,,p/3}(k)(k \wedge (N-1))\,. \tag{4.22}$$

Hence, using (4.15), (4.20), (4.21), (4.22), (4.19) and applying Proposition 4, we get

$$\|Z_n(g)\|_p \leq (p\|g\|_{Q,1})^{1/2} + \Big(2p\sum_{k=N}^{+\infty} \beta_{2,p/2}(k)\Big)^{1/2}$$
$$+ n^{1/3-1/2}\Big(3p^2\Big(1 + 2\sum_{k=1}^{N} k\beta_{2,p/2}(k) + 4\sum_{k=1}^{\infty} \beta_{2\,,p/3}(k)(k \wedge N) + 4\sum_{k=1}^{+\infty} \beta_{2,p/2}(k)\Big)\Big)^{1/3}. \tag{4.23}$$

We take now $p = 2p'$ and $N = n^\alpha$ with $\alpha = 1/(2+\varepsilon)$. If $(\mathcal{C}_1)$ holds, we infer from (4.23) that there exists some positive constant C such that, for any g in $\mathcal{G}$, $\|Z_n(g)\|_p \leq C\|g\|_{Q,1}^{1/2} + Cn^{-\varepsilon/(4+2\varepsilon)}$. To conclude we apply Proposition 6 with $r = 2$ and $q = 2+\varepsilon$. Since $\mathcal{N}_{Q,1}(x, \mathcal{F}) = \mathcal{O}(x^{-d})$, the process $\{Z_n(f), f \in \mathcal{F}\}$ is asymptotically tight as soon as $p' > d(2+\varepsilon)/2\varepsilon$. The result follows. $\square$

5 Comparison of coefficients

The following proposition is a useful tool to compute upper bounds for the coefficients $\beta_{k,p}$.

Proposition 7 *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $X = (X_1, \ldots, X_d)$ and $Y = (Y_1, \ldots, Y_d)$ two random variables with values in $\mathbb{R}^d$, and $\mathcal{M}$ a σ-algebra of $\mathcal{A}$. If (X^*, Y^*) is distributed as (X, Y) and independent of $\mathcal{M}$ then, assuming that each component X_k and Y_k has a continuous distribution function F_{X_k} and F_{Y_k}, we get for any $x_1, \ldots, x_d, y_1, \ldots, y_d$ in $[0, 1]$,*

$$\beta_p(\mathcal{M}, X) \leq \Big\| \sum_{k=1}^{d} x_k + \mathbb{P}(|F_{X_k}(X_k^*) - F_{X_k}(X_k)| > x_k|\mathcal{M}) \Big\|_p . \tag{5.1}$$

$$\begin{aligned}\beta_p(\mathcal{M}, X, Y) \leq & \Big\| \sum_{k=1}^{d} x_k + \mathbb{P}(|F_{X_k}(X_k^*) - F_{X_k}(X_k)| > x_k|\mathcal{M}) \Big\|_p \\ & + \Big\| \sum_{k=1}^{d} y_k + \mathbb{P}(|F_{Y_k}(Y_k^*) - F_{Y_k}(Y_k)| > y_k|\mathcal{M}) \Big\|_p .\end{aligned} \tag{5.2}$$

Proof of Proposition 7. Let Z be a random variable with values in $\mathbb{R}^m$ and let f be a function from $\mathbb{R}^m$ to $\mathbb{R}$ such that $|f(z_1, \ldots, z_i, \ldots, z_m) - f(z_1, \ldots, z_i', \ldots, z_m)| \leq |\mathbb{1}_{z_i \leq a_i} - \mathbb{1}_{z_i' \leq a_i}|$ for some real numbers $a_1, \ldots, a_m$. Let $\mathcal{U}$ be a σ-algebra and let Z^* be a random variable distributed as Z and independent of $\mathcal{U}$. Then

$$\begin{aligned}|f(Z) - f(Z^*)| &= \Big| \sum_{k=1}^{m} f(Z_1, \ldots Z_k, Z_{k+1}^*, \ldots Z_m^*) - f(Z_1, \ldots Z_{k-1}, Z_k^*, \ldots Z_m^*) \Big| \\ &\leq \sum_{k=1}^{m} |\mathbb{1}_{Z_k \leq a_k} - \mathbb{1}_{Z_k^* \leq a_k}| .\end{aligned}$$

Hence

$$|\mathbb{E}(f(Z)|\mathcal{U}) - \mathbb{E}(f(Z))| \leq \mathbb{E}(|f(Z) - f(Z^*)|\,|\mathcal{U}) \leq \sum_{k=1}^{m} \mathbb{E}\left(|\mathbb{1}_{Z_k \leq a_k} - \mathbb{1}_{Z_k^* \leq a_k}|\,|\mathcal{U}\right) . \tag{5.3}$$

We first apply (5.3) to $Z = X$, $Z^* = X^*$, $\mathcal{U} = \mathcal{M}$, and $f(z) = \mathbb{1}_{z \leq t}$ with $a_1 = t_1, \ldots, a_k = t_k$. Since $F_{X_k}^{-1}(F_{X_k}(X_k)) = X_k$ almost surely, we obtain that

$$\begin{aligned}|\mathbb{E}(\mathbb{1}_{X \leq t}|\mathcal{M}) - \mathbb{P}(X \leq t)| &\leq \sum_{k=1}^{d} \mathbb{E}(|\mathbb{1}_{X_k \leq t_k} - \mathbb{1}_{X_k^* \leq t_k}|\,|\mathcal{M}) \\ &\leq \sum_{k=1}^{d} \mathbb{E}(|\mathbb{1}_{F_{X_k}(X_k) \leq F_{X_k}(t_k)} - \mathbb{1}_{F_{X_k}(X_k^*) \leq F_{X_k}(t_k)}|\,|\mathcal{M}) .\end{aligned}$$

Using the same arguments as in Lemma 2 of Dedecker and Prieur (2005), the inequality (5.1) follows.

In the same way, applying (5.3) to $Z = (Z^{(1)}, Z^{(2)}) = (X, Y)$, $Z^* = (X^*, Y^*)$, $\mathcal{U} = \mathcal{M}$ and

$$f(z^{(1)}, z^{(2)}) = (\mathbb{1}_{z^{(1)} \le s} - F_X(s))(\mathbb{1}_{z^{(2)} \le t} - F_Y(t)),$$

we obtain that

$$\begin{aligned}|\mathbb{E}((\mathbb{1}_{X \le s} - F_X(s))(\mathbb{1}_{Y \le t} - F_Y(t))|\mathcal{M}) - \mathbb{E}((\mathbb{1}_{X \le s} - F_X(s))(\mathbb{1}_{Y \le t} - F_Y(t)))| \\ \le \sum_{k=1}^{d} \mathbb{E}(|\mathbb{1}_{X_k \le s_k} - \mathbb{1}_{X_k^* \le s_k}|\,|\mathcal{M}) + \sum_{k=1}^{d} \mathbb{E}(|\mathbb{1}_{Y_k \le t_k} - \mathbb{1}_{Y_k^* \le t_k}|\,|\mathcal{M}),\end{aligned}$$

and we conclude the proof of (5.2) by using the same arguments as for (5.1). □

Next, we apply Proposition 7 to compare $\beta_p(M, X)$ and $\tau_p(\mathcal{M}, X)$.

Proposition 8 *If each component of X and Y has a density bounded by K, then we have the following upper bounds*

$$\beta_p(\mathcal{M}, X) \le 2\sqrt{Kd\,\tau_p(\mathcal{M}, X)} \quad \textit{and} \quad \beta_p(\mathcal{M}, X, Y) \le 2\sqrt{2Kd\,\tau_p(\mathcal{M}, X, Y)}\,.$$

Proof of Proposition 8. Starting from (5.1) with $x_1 = \cdots = x_k = x$ and applying Markov's inequality, we infer that

$$\beta_p(\mathcal{M}, X) \le dx + \frac{K}{x}\Big\|\mathbb{E}\Big(\sum_{k=1}^{d} |X_k - X_k^*|\Big|\mathcal{M}\Big)\Big\|_p.$$

Now, from Proposition 6 in Rüschendorf (1985) (see also the equality (7.5) in Dedecker and Prieur (2005)), one can choose X^* such that

$$\Big\|\mathbb{E}\Big(\sum_{k=1}^{d} |X_k - X_k^*|\Big|\mathcal{M}\Big)\Big\|_p = \tau_p(\mathcal{M}, X)$$

Hence

$$\beta_p(\mathcal{M}, X) \le dx + \frac{K\tau_p(\mathcal{M}, X)}{x},$$

and the first inequality follows by minimizing in x. The second inequality may be proved in the same way. □

In the last part of this section, we show that the coefficient $\beta(\mathcal{M}, X_1, \ldots, X_k)$ (resp. $\phi(\mathcal{M}, X_1, \ldots, X_k)$), defined in Definition 1, is smaller than the usual β-mixing coefficient $\beta(\mathcal{M}, \sigma(X_1, \ldots, X_k))$ (resp. ϕ-mixing coefficient $\phi(\mathcal{M}, \sigma(X_1, \ldots, X_k))$) of Rozanov and Volkonskii (1959) (resp. Ibragimov (1962)). Let $\mathcal{X}$ be some Polish space, and let $\Lambda_1(\mathcal{X}, d_0)$ be the set of measurable functions from $\mathcal{X}$ to $\mathbb{R}$ which are 1-lipschitz with respect to the discrete metric $d_0(x, y) = \mathbb{1}_{x \ne y}$. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. For any random variable with values in $\mathcal{X}$, and any σ-algebra $\mathcal{M}$ of $\mathcal{A}$, define

$$b(\mathcal{M}, \sigma(X)) = \sup_{f \in \Lambda_1(\mathcal{X}, d_0)} \Big| \int f(x) \mathbb{P}_{X|\mathcal{M}}(dx) - \int f(x) \mathbb{P}_X(dx) \Big| ,$$

and recall that the mixing coefficients $\beta(\mathcal{M}, X)$ and $\phi(\mathcal{M}, X)$ may be defined as $\beta(\mathcal{M}, \sigma(X)) = \|b(\mathcal{M}, \sigma(X)\|_1$ and $\phi(\mathcal{M}, \sigma(X)) = \|b(\mathcal{M}, \sigma(X))\|_\infty$. In the next section, we shall give many non-mixing sequences for which the coefficient $\beta_{2,p}(n)$ of Definition 2 tends to zero as n tends to infinity.

Proposition 9 *Let $(X_1, \ldots X_k)$ be a random variable with values in $\mathbb{R}^{kd}$. We have that $b(\mathcal{M}, (X_1, \ldots, X_k)) \leq b(\mathcal{M}, \sigma(X_1, \ldots, X_k)) \leq 1$ almost surely.*

Proof of Proposition 9. The second inequality follows easily from the fact that $|f(x) - f(y)| \leq 1$ for any f in $\Lambda_1(\mathbb{R}^{kd}, d_0)$. To prove the first one, it suffices to see that the function $g : (x_1, \ldots, x_k) \to \Pi_{i=1}^k g_{t_i,i}(x_i)$ defined in Definition 1 belongs to $\Lambda_1(\mathbb{R}^{kd}, d_0)$ (this can be done by induction on k). □

6 Examples

In this section, we present three classes of examples for which we can compute upper bounds for $\beta_{2,p}(n)$ for any $p \geq 1$ and any $n \geq 1$. For the coefficients $\tau_{k,p}(n)$, many examples are given in Dedecker and Prieur (2005), Section 7.2.

6.1 Example 1 : causal functions of stationary sequences

Let $(\xi_i)_{i \in \mathbb{Z}}$ be a stationary sequence of random variables with values in a measurable space $\mathcal{X}$. Assume that there exists a function H defined on a subset of $\mathcal{X}^{\mathbb{N}}$, with values in $\mathbb{R}^d$ and such that $H(\xi_0, \xi_{-1}, \xi_{-2}, \ldots,)$ is defined almost surely. The stationary sequence $(X_n)_{n \in \mathbb{Z}}$ defined by $X_n = H(\xi_n, \xi_{n-1}, \xi_{n-2}, \ldots)$ is called a causal function of $(\xi_i)_{i \in \mathbb{Z}}$.

Assume that there exists a stationary sequence $(\xi'_i)_{i \in \mathbb{Z}}$ distributed as $(\xi_i)_{i \in \mathbb{Z}}$ and independent of $(\xi_i)_{i \leq 0}$. Define $X_n^* = H(\xi'_n, \xi'_{n-1}, \xi'_{n-2}, \ldots)$. Clearly X_n^* is independent of $\sigma(X_i, i \leq 0)$ and distributed as X_n. For any $x = (x_1, \ldots, x_d)$ in $\mathbb{R}^d$ let $|x|_\infty = \max(|x_1|, \ldots, |x_d|)$. For any $\alpha > 0$ (α may be infinite) define the sequence $(\delta_{i,\alpha})_{i>0}$ by

$$(\mathbb{E}(\,|X_i - X_i^*|_\infty^\alpha))^{1/\alpha} = \delta_{i,\alpha} \, . \tag{6.1}$$

Let $\mathcal{M}_i = \sigma(X_j, j \leq i)$. Since $(X_i)_{i \in \mathbb{Z}}$ is a strictly stationary sequence, we can write :

$$\beta_{2,p}(n) = \max \left(\beta_p(\mathcal{M}_0, X_n), \sup_{j_2 > j_1 \geq n} \beta_p(\mathcal{M}_0, X_{j_1}, X_{j_2}) \right) . \tag{6.2}$$

Let F_i be the distribution function of $X_0^{(i)}$. Using Proposition 7, we obtain for any $x \in [0,1]$ and any $y \in [0,1]$,

$$\beta_p(\mathcal{M}_0, X_{j_1}, X_{j_2}) \leq dx + \Big\| \sum_{k=1}^{d} \mathbb{P}(|F_k((X_{j_1}^*)^{(k)}) - F_k(X_{j_1}^{(k)})| > x|\mathcal{M}_0) \Big\|_p \\ + dy + \Big\| \sum_{k=1}^{d} \mathbb{P}(|F_k((X_{j_2}^*)^{(k)}) - F_k(X_{j_2}^{(k)})| > y|\mathcal{M}_0) \Big\|_p . \quad (6.3)$$

Assume now that each component of X_0 has a continuous distribution function, and let w be the maximum of the modulus of continuity, that is

$$w(x) = \max_{1\leq k\leq d} \sup_{|y-z|\leq x} |F_k(y) - F_k(z)| . \quad (6.4)$$

Define the function g_r by $g_r(y) = y(w(y))^{1/r}$. Clearly

$$dw(x) + \Big\| \sum_{k=1}^{d} \mathbb{P}(|F_k((X_{j_1}^*)^{(k)}) - F_k(X_{j_1}^{(k)})| > w(x)|\mathcal{M}_0) \Big\|_p \\ \leq dw(x) + \Big\| \sum_{k=1}^{d} \mathbb{P}(|(X_{j_1}^*)^{(k)} - X_{j_1}^{(k)}| > x|\mathcal{M}_0) \Big\|_p . \quad (6.5)$$

Now, using Markov inequality at order $r > 0$,

$$dw(x) + \Big\| \sum_{k=1}^{d} \mathbb{P}(|(X_{j_1}^*)^{(k)} - X_{j_1}^{(k)}| > x|\mathcal{M}_0) \Big\|_p \leq dw(x) + d\left(\frac{\delta_{j_1,pr}}{x}\right)^r \quad (6.6)$$

Combining (6.2), (6.3), (6.5) and (6.6), and taking $x = g_r^{-1}(\delta_{j_1,pr}), y = g_r^{-1}(\delta_{j_2,pr})$, we conclude that

$$\beta_{2,p}(n) \leq 2d \sup_{j_2>j_1\geq n} \left(\left(\frac{\delta_{j_1,pr}}{g_r^{-1}(\delta_{j_1,pr})} \right)^r + \left(\frac{\delta_{j_2,pr}}{g_r^{-1}(\delta_{j_2,pr})} \right)^r \right) . \quad (6.7)$$

From (6.2) and (6.3), we also have that

$$\phi_2(n) \leq 2d \sup_{j_2>j_1\geq n} w(\delta_{j_1,\infty} \vee \delta_{j_2,\infty}). \quad (6.8)$$

If each component of X_0 has a density bounded by K, we obtain :

$$dw(x) + \Big\| \sum_{k=1}^{d} \mathbb{P}(|(X_{j_1}^*)^{(k)} - X_{j_1}^{(k)}| > x|\mathcal{M}_0) \Big\|_p \leq dKx + d\left(\frac{\delta_{j_1,pr}}{x}\right)^r .$$

Minimizing the right hand term in this inequality, we get :

$$\beta_{2,p}(n) \leq C(r) d K^{\frac{r}{r+1}} \sup_{j_2>j_1\geq n} \left(\delta_{j_1,pr}^{\frac{r}{r+1}} + \delta_{j_2,pr}^{\frac{r}{r+1}} \right) , \quad (6.9)$$

with $C(r) = r^{\frac{1}{r+1}} + r^{-\frac{r}{r+1}}$ (note that $C(r) \leq 2$ and $C(\infty) = 1$).

In particular, the bounds (6.7)-(6.8) and (6.9) apply to the case where the sequence $(\xi_i)_{i\in\mathbb{Z}}$ is β-mixing. According to Theorem 4.4.7 in Berbee (1979), if Ω is rich enough, there exists $(\xi'_i)_{i\in\mathbb{Z}}$ distributed as $(\xi_i)_{i\in\mathbb{Z}}$ and independent of $(\xi_i)_{i\leq 0}$ such that $\mathbb{P}(\xi_i \neq \xi'_i \text{ for some } i \geq k) = \beta(\sigma(\xi_i, i \leq 0), \sigma(\xi_i, i \geq k))$. If the sequence $(\xi_i)_{i\in\mathbb{Z}}$ is iid, it suffices to take $\xi'_i = \xi_i$ for $i > 0$ and $\xi'_i = \xi''_i$ for $i \leq 0$, where $(\xi''_i)_{i\in\mathbb{Z}}$ is an independent copy of $(\xi_i)_{i\in\mathbb{Z}}$.

Application: causal linear processes in $\mathbb{R}^d$. Let $(\mathbb{B}, |\cdot|_{\mathbb{B}})$ be a Banach space. For any linear application A from $\mathbb{B}$ to $\mathbb{R}^d$, let $\|A\| = \sup\{|Ab|_\infty, |b|_{\mathbb{B}} \leq 1\}$. Let $(A_i)_{i\geq 0}$ be a sequence of linear operators from $\mathbb{B}$ to $\mathbb{R}^d$ such that $\sum_{i\geq 0} \|A_i\| < \infty$, and let $(\xi)_{i\in\mathbb{Z}}$ be a stationary sequence of $\mathbb{B}$-valued random variables. Define the random variables with values in $\mathbb{R}^d$

$$X_n = \sum_{j\geq 0} A_j \xi_{n-j} \,. \tag{6.10}$$

For any $p \geq 1$, we have that

$$\delta_{i,p} \leq \sum_{j\geq 0} \|A_j\| \| \, |\xi_{i-j} - \xi'_{i-j}|_{\mathbb{B}} \|_p \leq \| \, |\xi_0 - \xi'_0|_{\mathbb{B}} \|_p \sum_{j\geq i} \|A_j\| + \sum_{j=0}^{i-1} \|A_j\| \| \, |\xi_{i-j} - \xi'_{i-j}|_{\mathbb{B}} \|_p \,.$$

From Proposition 2.3 in Merlevède and Peligrad (2002), we obtain that

$$\delta_{i,p} \leq \| \, |\xi_0 - \xi'_0|_{\mathbb{B}} \|_p \sum_{j\geq i} \|A_j\| + \sum_{j=0}^{i-1} \|A_j\| \Big(2^p \int_0^{\beta(\sigma(\xi_k, k\leq 0), \sigma(\xi_k, k\geq i-j))} Q^p_{|\xi_0|_{\mathbb{B}}}(u) du \Big)^{1/p} .$$

where Q_{ξ_0} is the generalized inverse of $t \to \mathbb{P}(\|\xi_0\| > t)$ (note that in Merlevède and Peligrad the constant in front of the integral is 2^{p+2}. In fact it works with the constant 2^p).

If the sequence $(\xi_i)_{i\in\mathbb{Z}}$ is iid, it follows that for any $p \geq 1$,

$$\delta_{i,p} \leq \| \, |\xi_0 - \xi'_0|_{\mathbb{B}} \|_p \sum_{j\geq i} \|A_j\| \,. \tag{6.11}$$

For instance, if $\mathbb{B} = \mathbb{R}$, $A_i = 2^{-i-1}$ and $\xi_0 \sim \mathcal{B}(1/2)$, then $\delta_{i,\infty} \leq 2^{-i}$. Since X_0 is uniformly distributed over $[0,1]$, we have $\phi(i) \leq 2^{-i}$. Recall that this sequence is not strongly mixing.

Applying Theorem 1, Corollary 1 below gives some sufficient conditions for the empirical central limit theorem to hold, when the sequence $(\xi_i)_{i\in\mathbb{Z}}$ is iid.

Corollary 1 *Let $(\xi_i)_{i\in\mathbb{Z}}$ be an iid sequence of $\mathbb{B}$-valued random variables. Let $(A_i)_{i\geq 0}$ be a sequence of linear operators from $\mathbb{B}$ to $\mathbb{R}^d$ such that $\sum_{i\geq 0} \|A_i\| < \infty$, and let $R_n = \sum_{i\geq n} \|A_i\|$. Let $(X_n)_{n\in\mathbb{Z}}$ be the stationary sequence defined by (6.10), and assume that each component of X_0 has a density bounded by K. If one of the following condition holds, then $\{\sqrt{n}(F_n(t) - F(t)) \,,\, t \in \mathbb{R}^d\}$ converges weakly in $\ell^\infty(\mathbb{R}^d)$ to a tight Gaussian process with covariance function given by (3.1).*

1. *For* $1 \le m \le 7d/2$, *the random variable* $|\xi_0|_{\mathbb{B}}$ *belongs to* $\mathbb{L}^m$ *and* $R_n = O(n^{-a})$ *for* $a > 2(m+d)/m$.
2. *For* $m > 7d/2$, *the random variable* $|\xi_0|_{\mathbb{B}}$ *belongs to* $\mathbb{L}^m$ *and* $R_n = O(n^{-a})$ *for* $a > (\sqrt{2m+d} + \sqrt{2d})^2/2m$.

The proof of this corollary is immediate by using the bounds (6.9) and (6.11) with $r = m/p$, and by optimizing in p (1. follow from the condition $(\mathcal{C}_2)$ and 2. follows from the condition $(\mathcal{C}_1)$). For instance if $\mathbb{E}(|\xi_0|_{\mathbb{B}}) < \infty$, the rate is $R_n = O(n^{-a})$, for $a > 2d + 2$. If $\mathbb{E}(|\xi_0|_{\mathbb{B}}^d) < \infty$, the rate is $R_n = O(n^{-a})$, for $a > 4$. If $\mathbb{E}(|\xi_0|_{\mathbb{B}}^{2d}) < \infty$, the rate is $R_n = O(n^{-a})$, for $a > 3$. If $\| |\xi_0|_{\mathbb{B}} \|_\infty < \infty$, the rate is $R_n = O(n^{-a})$, for $a > 1$.

For $\mathbb{B} = \mathbb{R}$ and $d = 1$, Doukhan and Surgailis (1998) obtained an empirical central limit theorem under the condition $\mathbb{E}(|\xi_0|^{4\gamma}) < \infty$, $\sum_{k>0} |A_k|^\gamma < \infty$ for some $0 < \gamma \le 1$, and an additional condition on the law of ξ_0 (which implies that the distribution function of ξ_0 is Δ-Hölder for some $\Delta > 1/2$). Next, Wu (2006) Corollary 2, obtained an empirical central limit theorem by assuming only that $\mathbb{E}(|\xi_0|^2) < \infty$, $\sum_{k>0} |A_k| < \infty$, and that ξ_0 has a density belonging to the Sobolev space of order 2 (in particular, it is bounded and two times differentiable). We note that the conditions on $(A_i)_{i\ge0}$ obtained in the above papers are weaker than ours. However, the main difference between our result and that of Doukhan and Surgailis or Wu is that we do not make any assumption on the distribution of ξ_0 (except moment assumptions). For instance, we obtain the empirical central limit theorem for $A_i = 2^{-i-1}$ and $\xi_0 \sim \mathcal{B}(1/2)$, which does not follow from the results cited above.

6.2 Example 2 : iterated random functions

Let $(X_n)_{n\ge0}$ be a $\mathbb{R}^d$-valued stationary Markov chain, such that

$$X_n = F(X_{n-1}, \xi_n) \tag{6.12}$$

for some measurable function F and some i.i.d. sequence $(\xi_i)_{i>0}$ independent of X_0. Let X_0^* be a random variable distributed as X_0 and independent of $(X_0, (\xi_i)_{i>0})$. As in Shao and Wu (2004), define $X_n^* = F(X_{n-1}^*, \xi_n)$. The sequence $(X_n^*)_{n\ge0}$ is distributed as $(X_n)_{n\ge0}$ and independent of X_0. Let $\mathcal{M}_i = \sigma(X_j, 0 \le j \le i)$. As in Example 1, define the sequence $(\delta_{i,p})_{i>0}$ by (6.1). The coefficient $\beta_{2,p}(n)$ of the sequence $(X_n)_{n\ge0}$ satisfy (6.7)-(6.8) of Example 1.

Let μ be the distribution of X_0 and $(X_n^x)_{n\ge0}$ the chain starting from $X_0^x = x$. With these notations, we have that, for any $\alpha > 0$,

$$\delta_{i,\alpha} = \Big(\iint \mathbb{E}(|X_i^x - X_i^y|_\infty^\alpha)\mu(dx)\mu(dy) \Big)^{1/\alpha} \tag{6.13}$$

For instance, if there exists a sequence $(d_{i,\alpha})_{i\geq 0}$ of positive numbers such that

$$(\mathbb{E}(|X_i^x - X_i^y|_\infty^\alpha))^{1/\alpha} \leq d_{i,\alpha}|x-y|_\infty ,$$

then $\delta_{i,\alpha} \leq d_{i,\alpha}(\mathbb{E}(|X_0 - X_0^*|_\infty^\alpha))^{1/\alpha}$. For instance, in the usual case where

$$(\mathbb{E}(|F(x,\xi_0) - F(y,\xi_0)|_\infty^\alpha))^{1/\alpha} \leq \kappa|x-y|_\infty \tag{6.14}$$

for some $\kappa < 1$, we can take $d_{i,\alpha} = \kappa^i$.

An important example is $X_n = f(X_{n-1}) + \xi_n$ for some function f which is κ-lipschitz with respect to the norm $|\cdot|_\infty$. If $|X_0|_\infty$ has a moment of order α, then $\delta_{i,\alpha} \leq \kappa^i(\mathbb{E}(|X_0 - X_0^*|_\infty^\alpha)^{1/\alpha}$. We refer to the papers by Diaconis and Freedman (1999) and Shao and Wu (2004) for various examples of iterative random maps.

As in Shao and Wu (2004), we can apply our results to the case where the function w defined in (6.4) is such that $w(x) \leq K|\ln(x)|^{-\gamma}$. This leads to the following Corollary:

Corollary 2 *Let $(X_n)_{n\geq 0}$ be a $\mathbb{R}^d$-valued stationary Markov chain satisfying (6.12), and let $\delta_{i,\alpha}$ be the coefficients defined in (6.13). Assume that the function w defined in (6.4) is such that $w(x) \leq C|\ln(x)|^{-\gamma}$ for some $\gamma > 1$. If $\delta_{i,\alpha} \leq C\kappa^i$ for some $\kappa < 1$ and $\alpha > 0$, then $\{\sqrt{n}(F_n(t) - F(t)), t \in \mathbb{R}^d\}$ converges weakly in $\ell^\infty(\mathbb{R}^d)$ to a tight Gaussian process with covariance function given by (3.1). In particular, $\delta_{i,\alpha} \leq C\kappa^i$ holds as soon as (6.14) holds.*

The condition $\delta_{i,\alpha} \leq C\kappa^i$ is exactly Condition (2) of Theorem 4 in Shao and Wu (2004), in the case where $d = 1$. Our result improves on the corresponding one in Theorem 4 of Shao and Wu, which gives $\gamma > 5/2$ instead of $\gamma > 1$. Note that the constant $5/2$ in their result is obtained by applying the criterion (1.2). Here, we obtain the condition $\gamma > 1$ by applying the criterion $(\mathcal{C}_1)$ instead of (1.2).

Proof of Corollary 2. Starting from (6.3) and applying (6.6) with $r = \alpha/p$, we have

$$\beta_{2,p}(n) \leq 2d\Big(K|\ln(x)|^{-\gamma} + \Big(\frac{C\kappa^n}{x}\Big)^{\alpha/p}\Big) .$$

Taking $x = C\kappa^n n^{2p/\alpha}$, it follows that, for any $p > 1$,

$$\beta_{2,p}(n) \leq 2d\Big(K|\ln(c\kappa^n n^{2p/\alpha})|^{-\gamma} + \frac{1}{n^2}\Big) ,$$

and the condition $(\mathcal{C}_1)$ is satisfied by taking $p > d(\gamma+1)/2(\gamma-1)$.

6.3 Example 3 : dynamical systems on $[0, 1]$.

Let $I = [0, 1]$, T be a map from I to I and define $X_i = T^i$. If μ is invariant by T, the sequence $(X_i)_{i\geq 0}$ of random variables from (I, μ) to I is strictly stationary.

Denote by $\|g\|_{1,\lambda}$ the $\mathbb{L}^1$-norm with respect to the Lebesgue measure λ on I and by $\|\nu\| = |\nu|(I)$ the total variation of ν.

Covariance inequalities. In many interesting cases, one can prove that, for any BV function h and any k in $\mathbb{L}^1(I, \mu)$,

$$|\mathrm{Cov}(h(X_0), k(X_n))| \le a_n \|k(X_n)\|_1 (\|h\|_{1,\lambda} + \|dh\|) \,, \tag{6.15}$$

for some non increasing sequence a_n tending to zero as n tends to infinity. Note that if (6.15) holds, then

$$\begin{aligned} |\mathrm{Cov}(h(X_0), k(X_n))| &= |\mathrm{Cov}(h(X_0) - h(0), k(X_n))| \\ &\le a_n \|k(X_n)\|_1 (\|h - h(0)\|_{1,\lambda} + \|dh\|) \,. \end{aligned}$$

Since $\|h - h(0)\|_{1,\lambda} \le \|dh\|$, we obtain that

$$|\mathrm{Cov}(h(X_0), k(X_n))| \le 2a_n \|k(X_n)\|_1 \|dh\| \,. \tag{6.16}$$

If (6.16) holds, the upper bound $\phi(\sigma(X_n), X_0) \le 2a_n$ follows from Lemma 4 in Dedecker and Prieur.

The associated Markov chain. Define the operator $\mathcal{L}$ from $\mathbb{L}^1(I, \lambda)$ to $\mathbb{L}^1(I, \lambda)$ *via* the equality

$$\int_0^1 \mathcal{L}(h)(x) k(x) \lambda(dx) = \int_0^1 h(x)(k \circ T)(x) \lambda(dx)$$

where $h \in \mathbb{L}^1(I, \lambda)$ and $k \in \mathbb{L}^\infty(I, \lambda)$. The operator $\mathcal{L}$ is called the Perron-Frobenius operator of T. Assume that μ is absolutely continuous with respect to the Lebesgue measure, with density f_μ. Let I^* be the support of μ (that is $(I^*)^c$ is the largest open set in I such that $\mu((I^*)^c) = 0$) and choose a version of f_μ such that $f_\mu > 0$ on I^* and $f_\mu = 0$ on $(I^*)^c$. Note that one can always choose $\mathcal{L}$ such that $\mathcal{L}(f_\mu h)(x) = \mathcal{L}(f_\mu h)(x) \mathbb{1}_{f_\mu(x)>0}$. Define a Markov kernel associated to T by

$$K(h)(x) = \frac{\mathcal{L}(f_\mu h)(x)}{f_\mu(x)} \mathbb{1}_{f_\mu(x)>0} + \mu(h) \mathbb{1}_{f_\mu(x)=0}. \tag{6.17}$$

It is easy to check (see for instance Barbour *et al.* (2000)) that $(X_0, X_1, \ldots, X_n)$ has the same distribution as $(Y_n, Y_{n-1}, \ldots, Y_0)$ where $(Y_i)_{i \ge 0}$ is a stationary Markov chain with invariant distribution μ and transition kernel K. Here, we need the following result:

Lemma 1 *Let $(Y_i)_{i\ge0}$ be a real-valued Markov chain with transition kernel K. Assume that there exists a constant C such that*

$$\text{for any } BV \text{ function } f \text{ and any } n > 0, \quad \|dK^n(f)\| \le C \|df\| \,. \tag{6.18}$$

Then, for any $j > i \ge 0$, $\phi(\sigma(Y_k), Y_{k+i}, Y_{k+j}) \le (1 + C)\phi(\sigma(Y_k), Y_{k+i})$.

Consequently, if both (6.15) and (6.18) hold it follows that, for any $n \geq j > i \geq 0$,

$$\phi(\sigma(X_k, k \geq n), X_{n-i}, X_{n-j}) \leq (1+C)\phi(\sigma(X_k, k \geq n), X_{n-i}) \leq 2(1+C)a_i\,.$$

Proof of Lemma 1. Let $f_k(x) = f(x) - \mathbb{E}(f(Y_k))$. We have, almost surely,

$$\begin{aligned}\mathbb{E}(f_{k+i}(Y_{k+i})g_{k+j}(Y_{k+j})|Y_k) - \mathbb{E}(f_{k+i}(Y_{k+i})g_{k+j}(Y_{k+j})) = \\ \mathbb{E}(f_{k+i}(Y_{k+i})(K^{j-i}(g))_{k+i}(Y_{k+i})|Y_k) - \mathbb{E}(f_{k+i}(Y_{k+i})(K^{j-i}(g))_{k+i}(Y_{k+i}))\,.\end{aligned}$$

Let f and g be two functions in BV_1. It is easy to see that

$$\begin{aligned}\|d((K^{j-i}(g))_{k+i}f_{k+i})\| &\leq \|df_{k+i}\|\ \|(K^{j-i}(g))_{k+i}\|_\infty + \|d(K^{j-i}(g))_{k+i}\|\ \|f_{k+i}\|_\infty \\ &\leq (1 + \|d(K^{j-i}(g))_{k+i}\|)\,.\end{aligned}$$

Hence, applying (6.18), the function $(K^{j-i}(g))_{k+i}f_{k+i}/(1+C)$ belongs to BV_1. The result follows from Proposition 1.

Spectral gap. In many interesting cases, the spectral analysis of $\mathcal{L}$ in the Banach space of BV-functions equipped with the norm $\|h\|_v = \|dh\| + \|h\|_{1,\lambda}$ can be done by using the Theorem of Ionescu-Tulcea and Marinescu (see Lasota and Yorke (1974)). Assume that 1 is a simple eigenvalue of $\mathcal{L}$ and that the rest of the spectrum is contained in a closed disk of radius strictly smaller than one. Then there exists an unique T-invariant absolutely continuous probability μ whose density f_μ is BV, and

$$\mathcal{L}^n(h) = \lambda(h)f_\mu + \Psi^n(h) \quad \text{with } \Psi(f_\mu) = 0 \text{ and } \quad \|\Psi^n(h)\|_v \leq D\rho^n\|h\|_v. \tag{6.19}$$

for some $0 \leq \rho < 1$ and $D > 0$. Assume moreover that

$$\left\| \frac{1}{f_\mu}\mathbb{I}_{f_\mu>0} \right\|_v = \gamma < \infty\,. \tag{6.20}$$

Starting from (6.17), we have that

$$K^n(h) = \mu(h) + \frac{\Psi^n(hf_\mu)}{f_\mu}\mathbb{I}_{f_\mu>0}\,.$$

Let $\|\cdot\|_{\infty,\lambda}$ be the essential sup with respect to λ. Taking $C_1 = 2D\gamma(\|df_\mu\|+1)$, we obtain $\|K^n(h)-\mu(h)\|_{\infty,\lambda} \leq C_1\rho^n\|h\|_v$. This estimate implies (6.15) with $a_n = C_1\rho^n$. Indeed,

$$\begin{aligned}|\mathrm{Cov}(h(X_0), k(X_n))| &= |\mathrm{Cov}(h(Y_n), k(Y_0))| \\ &\leq \|k(Y_0)(\mathbb{E}(h(Y_n)|\sigma(Y_0)) - \mathbb{E}(h(Y_n)))\|_1 \\ &\leq \|k(Y_0)\|_1\|K^n(h) - \mu(h)\|_{\infty,\lambda} \\ &\leq C_1\rho^n\|k(Y_0)\|_1(\|dh\| + \|h\|_{1,\lambda})\,.\end{aligned}$$

Moreover, we also have that

$$\begin{aligned}\|dK^n(h)\| = \|dK^n(h-h(0))\| &\leq 2\gamma\|\Psi^n(f_\mu(h-h(0)))\|_v \\ &\leq 8D\rho^n\gamma(1+\|df_\mu\|)\|dh\|\end{aligned}$$

so that (6.18) holds with $C_2 = 8D\gamma(1+\|df_\mu\|)$. Finally, if (6.19) holds, the coefficients $\phi_2(i)$ of the chain $(Y_i)_{i\geq 0}$ with respect to $(\mathcal{M}_i = \sigma(Y_j, j \leq i))_{i\geq 0}$ satisfy

$$\phi_2(i) \leq 2C_1(1+C_2)\rho^i\,.$$

Application: expanding maps. A large class of expanding maps T is given in Broise (1996), Section 2.1, page 11. If Broise's condition are satisfied and if T is mixing in the ergodic-theoretic sense, then the Perron-Frobenius operator $\mathcal{L}$ satisfies the assumption (6.19). Let us recall some well know examples (see Section 2.2 in Broise):

1. $T(x) = \beta x - [\beta x]$ for $\beta > 1$. These maps are called β-transformations.

2. I is the finite union of disjoints intervals $(I_k)_{1\leq k\leq n}$, and $T(x) = a_k x + b_k$ on I_k, with $|a_k| > 1$.

3. $T(x) = a(x^{-1}-1) - [a(x^{-1}-1)]$ for some $a > 0$. For $a = 1$, this transformation is known as the Gauss map.

Aknowledgments. We wish to thank the referee for having pointed out some important references, which helped us to improve the first version of this paper. We also thank Herold Dehling for helpful discussions about the empirical distribution function of uniformly expanding maps.

References

[1] D. W. K. Andrews and D. Pollard, (1994). An introduction to functional central limit theorems for dependent stochastic processes. *Int. Stat. Rev.* **62** 119-132.

[2] A. D. Barbour, R. M. Gerrard and G. Reinert, (2000). Iterates of expanding maps. *Probab. Theory Relat. Fields* **116** 151-180.

[3] H. C. P. Berbee, (1979). Random walks with stationary increments and renewal theory. *Math. Cent. Tracts.* Amsterdam.

[4] P. Billingsley, (1968). *Convergence of probability measures.* Wiley, New-York.

[5] S. Borovkova, R. Burton and H. Dehling, (2001). Limit theorems for functionals of mixing processes with application to U-statistics and dimension estimation. *Trans. Amer. Math. Soc.* **353** 4261-4318.

[6] R. C. Bradley, (1997). On quantiles and the central limit question for strongly mixing sequences. *J. Theoret. Probab.* **10** 507-555.

[7] A. Broise, (1996). Transformations dilatantes de l'intervalle et théorèmes limites. Études spectrales d'opérateurs de transfert et applications. *Astérisque* **238** 1-109.

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

[9] J. Dedecker, (2001). Exponential inequalities and functional central limit theorems for random fields. *ESAIM Probab. & Stat.* `http://www.emath.fr/ps/` **5**, 77–104.

[10] J. Dedecker and C. Prieur, (2005). New dependence coefficients. Examples and applications to statistics. *Probab. Theory and Relat. Fields.* **132** 203-236.

[11] J. Dedecker and E. Rio, (2000). On the functional central limit theorem for stationary processes. *Ann. Inst. H. Poincaré Probab. Statist.* **36** 1-34.

[12] S. Dhompongsa, (1984). A note on the almost sure approximation of empirical process of weakly dependent random vectors. *Yokohama math. J.* **32** 113-121.

[13] P. Diaconis and D. Freedman, (1999). Iterated random functions. *SIAM Rev.* **41** 45-76.

[14] M. D. Donsker, (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statistics* **23**, 277-281.

[15] P. Doukhan and S. Louhichi, (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Process. Appl.* **84** 313-342.

[16] P. Doukhan and D. Surgailis, (1998). Functional central limit theorem for the empirical process of short memory linear processes. *C.R. Acad. Sci. Paris.* **326** 87-92.

[17] R. M. Dudley, (1966). Weak convergence of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces. *Illinois J. Math* **10** 109-126.

[18] F. Hofbauer and G. Keller, (1982). Ergodic properties of invariant measures for piecewise monotonic transformations. *Math. Z.* **180** 119–140.

[19] I. A. Ibragimov, (1962). Some limit theorems for stationary processes. *Theory Probab. Appl.* **7** 349-382.

[20] A. Lasota and J. A. Yorke (1974). On the existence of invariant measures for piecewise monotonic transformations. *Trans. Amer. Math. Soc.* **186** 481-488.

[21] S. Louhichi, (2000). Weak convergence for empirical processes of associated sequences. *Ann. Inst. H. Poincaré Probab. Statist.* **36** 547-567.

[22] F. Merlevède and M. Peligrad (2002). On the coupling of dependent random variables and applications. *Empirical process techniques for dependent data.* 171-193. Birkhäuser.

[23] C. Prieur, (2002). An Empirical Functional Central Limit Theorem For Weakly Dependent Sequences. *Probab. and Math. Statistics* Vol. **22** Fasc. **2** 259-287.

[24] E. Rio, (1998). Processus empiriques absoluments réguliers et entropie universelle. *Probab. Theory and Relat. Fields* **111** 585-608.

[25] E. Rio, (2000). *Théorie asymptotique des processus aléatoires faiblement dépendants.* Collection Mathématiques & Applications **31**. Springer, Berlin.

[26] M. Rosenblatt, (1956). A central limit theorem and a strong mixing condition, *Proc. Nat. Acad. Sci. U. S. A.* **42** 43-47.

[27] Y. A. Rozanov and V. A. Volkonskii, (1959). Some limit theorems for random functions I. *Theory Probab. Appl.* **4** 178-197.

[28] L. Rüschendorf, (1985). The Wasserstein Distance and Approximation Theorems. *Z. Wahrsch. Verw. Gebiete* **70** 117-129.

[29] X. Shao and W-B. Wu, (2004). Limit Theorems for Iterated Random Functions, *J. Appl. Probab.* **41-2** 425-436.

[30] Q-M. Shao and H. Yu, (1996). Weak convergence for weighted empirical processes of dependent sequences, *Ann. Probab.* **24** 2098-2127.

[31] A. W. van der Vaart and J. A. Wellner, (1996). *Weak convergence and empirical processes.* Springer, Berlin.

[32] W. B. Wu, (2006). Empirical process of stationary sequences, *to appear.* Available at `http://galton.u-chicago.edu/~wbwu/papers/empirical-apr26-06.pdf`

[33] K. Yoshihara, (1979). Note on an almost sure invariance principle for some empirical processes. *Yokohama math. J* **27**, 105-110.