INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2001-35

Weak Dependence : Models and Applications

P. ANGO NZE[1]
P. DOUKHAN[2]

[1] Université de Lille 3, UFR AES, BP 149, 59653 Villeneuve d'Ascq Cedex, France. Email : angonze@univ-lille3.fr
[2] CREST et Université de Cergy-Pontoise, UPRESA 8088 CNRS-Mathématiques, bât. Les Chênes, 33 Bld du Port, 95011 Cergy-Pontoise Cedex, France. Email : paul.doukhan@math.u-cergy.fr

Weak dependence : models and applications.

Patrick Ango Nze and Paul Doukhan

Abstract We are aimed to develop a systematic introduction to a new weak dependence condition. We show that some popular models hold this property : stationary Markov models, bilinear models, and more generally, Bernoulli shifts. In some cases no mixing properties can be expected without additional regularity assumption of innovation's distribution where as weak dependence conditions can be easily derived. We also develop some of its standard applications. First, probabilistic results : weak invariance principle for Donsker line and empirical process. We also present a statistical application to kernel estimates for density and regression functions.

Résumé Nous développons ici une introduction systematique à une nouvelle condition de dépendance faible. Nous montrons en particulier que quelques modèles très populaires en séries chronologiques vérifient cette condition : des chaînes de Markov stationnaires, des modèles bilinéaires ou plus généralement des schémas de Bernoulli. On ne peut, en général, espérer que ces modèles satisfassent à des conditions de mélange.

Quelques applications classiques sont aussi envisagées. Nous prouvons d'abord des théorèmes limite probabilistes : le principe de Donsker et le théorème de limite centrale empirique. Nous obtenons aussi le comportement asymptotique d'estimations à noyau d'une densité et d'une fonction de régression; leur convergence uniforme presque sûre ainsi que leur comportement asymptotiquement gaussien sont prouvés comme dans le cas de suites indépendantes.

Patrick Ango Nze
Université Lille 3, UFR AES
BP 149,
F-59653 Villeneuve d'Ascq Cedex
France
angonze@univ-lille3.fr
and
Paul Doukhan
UPRESA 8088 CNRS-Mathématiques
Université de Cergy Pontoise
Bât. Les Chênes, 33, Bd. du Port
F-95011 Cergy-Pontoise Cedex and CREST-LS, timbre J340,
3, Av. Pierre Larousse, F-92245 Malakoff
France
Paul.Doukhan@math.u-cergy.fr

1 Introduction

In this paper we are aimed to provide a real alternative to the standard mixing conditions (see e.g. [24]) for modelling some of the stationary sequences commonly used in statistics.

We deal with sequences of random variables. We do not assume independence. Asymptotic results obtained do not need to be analogous to that of independent sequences: this is a first but complicated way to define weak dependence.

We are thus turning to much simpler conditions:

– The first preoccupation we bear in mind is to obtain conditions, simple enough to check on the really used models; our experience (see [11] and [2]) proves the real difficulty to handle strong mixing or absolute regularity conditions.

– The second condition expected from a tractable weak dependence condition is that the planned applications may ensue from such condition.

Our first concern was to establish two probabilistic limit theorems: the Donsker invariance principle and the empirical CLT. They are used in statistical applications: to determine change point estimation in the mean of a stationary sequence or to obtain Kolmogorov-Smirnov tests for distributions. Kernel functional estimation problems provide other statistical applications.

This paper is organized as follows.

A first section is devoted to the definitions our new version of a weak dependence conditions. The main source is the seminal works [13] and [4].

A second section provides the weak dependence properties (in this sense) of some of the classical stationary sequences used in statistics: Markovian models, associated sequences, and Bernoulli shifts. As a by-product, this new dependence condition gives access to some classes of models which could not be studied with former tools (mixing, for instance).

After this we turn to some moment inequalities to be used in the sequel. This is a first application of those weak dependence conditions. Indeed, a moment inequality for sums may be proved by combinatorial arguments. This condition also yields some sub-exponential inequalities described in the application to functional estimation.

Another technical trick is the Lindeberg-Rio (see [26]) method, easily handled in this weak dependence frame. The latter method is compared with Bernstein's blocking technique (see [17]). Each method provides a CLT. The dependence rates for CLT convergence are compared later on in the functional estimation frame. Surprisingly, there is not an uniformly best method.

We then turn to the Donsker invariance principle and the empirical CLT.

Finally we consider functional estimation. Besides the convergence results which are of a proper interest, this frame allows to compare

– the weak dependence conditions with the standard mixing.

– the methods for proving CLTs.

– the obstructions in view of optimal almost sure convergence of kernel estimators (that is the i.i.d. rates).

The number of applications is relatively limited due to the fact that the conditions are quite recent. We hope that this panorama of statistical applications may be enlarged in the years to come.

2 Independence and dependence

2.1 Independence and correlation

Recall that random variables (r.v.), taking values in $\mathbb{R}^d$ are independent in case

$$\forall f, g \ \text{ bounded measurable} : \operatorname{Cov}(f(X), g(Y)) = 0.$$

It is enough to consider classes of continuous or more regular functions.

Mixing conditions (recalled hereafter) were introduced by weakening such conditions: the null, right hand side term above is made "small".

Note that in some cases, orthogonality yields independence.

- a) Bernoulli r.v's (that is r.v.'s with two points support.)
- b) Gaussian vectors : $(X, Y) \in \mathbb{R}^{a+b}$.
- c) Associated random vectors.

The first case yields only pairwise independence but not independence, in the case of more than two r.v.'s; it follows from orthogonality of the four couples (X, Y), $(X, 1-Y)$, $(1-X, Y)$ and $(1-X, 1-Y)$ if the couple (X, Y) is orthogonal and the variables (X, Y) are both supported in $\{0, 1\}$.

Except for the first case, properties inherited from covariances are not easy to translate into properties of subjacent sigma-fields.

In the two latter cases b) and c), additional inequalities may be proved for Lipschitz functions. They take the form

$$|\operatorname{Cov}(f(X), g(Y))| \leq c(f, g) \sum_{i,j} |\operatorname{Cov}(X_i, Y_j))|.$$

2.2 The weak dependence conditions

As the covariances of the initial r.v.'s are much easier to to compute than mixing coefficients, we introduce dependence properties (rather than sigma-fields) for a process $(X_n)_{n \in \mathbb{Z}}$. Set $\mathbb{L}^\infty$ for the set of numerical bounded measurable functions on some space $\mathbb{R}^u$ and $\|\cdot\|_\infty$ the corresponding norm. We define the Lipschitz modulus of a function $h : \mathbb{R}^u \to \mathbb{R}$

$$\operatorname{Lip}(h) = \sup_{x \neq y} \frac{|h(x) - h(y)|}{\|x - y\|_1}, \qquad \text{where } \|(z_1, \ldots, z_u)\|_1 = |z_1| + \cdots + |z_u|.$$

Define

$$\mathcal{L} = \{h \in \mathbb{L}^\infty; \|h\|_\infty \leq 1, \operatorname{Lip}(h) < \infty\}. \tag{1}$$

Definition 2.1 (Doukhan and Louhichi [13]) *The sequence $(X_n)_{n \in \mathbb{Z}}$ is called $(\theta, \mathcal{L}, \psi)$-weakly dependent if there exists a sequence $\theta = (\theta_r)_{r \in \mathbb{N}}$ decreasing to zero at infinity, and a function ψ with arguments $(h, k, u, v) \in \mathcal{L}^2 \times \mathbb{N}^2$ such that for any u-tuple $(i_1, \ldots, i_u)$ and any v-tuple $(j_1, \ldots, j_v)$ with $i_1 \leq \cdots \leq i_u < i_u + r \leq j_1 \leq \cdots \leq j_v$, one has*

$$|\operatorname{Cov}(h(X_{i_1}, \ldots, X_{i_u}), k(X_{j_1}, \ldots, X_{j_v}))| \leq \psi(h, k, u, v)\theta_r \tag{2}$$

if the functions h and k are defined respectively on $\mathbb{R}^u$ and on $\mathbb{R}^v$.

The examples of interest involve the function $\psi_1'(h,k,u,v) = v\mathrm{Lip}(k)$ (*e.g.* in causal linear processes), $\psi_1(h,k,u,v) = u\mathrm{Lip}(h) + v\mathrm{Lip}(k)$, (*e.g.* in non causal linear processes), $\psi_2(h,k,u,v) = uv\mathrm{Lip}(h)\mathrm{Lip}(k)$ (*e.g.* in associated processes), and $\psi_2'(h,k,u,v) = v\mathrm{Lip}(h)\mathrm{Lip}(k)$.

Notice that the sequence θ depends on both the class $\mathcal{L}$ and the function ψ. The function ψ may really depend on all its arguments, in contrast with the case of bounded mixing sequences. This definition is hereditary through images by convenient functions, as it was noticed by [35] in view to check more easily a weak dependence property.

The following example, due to Rosenblatt ([29]) describes the case where mixing fails to hold. This is the (Markov) $AR(1)$-process with binomial innovations ($\mathbb{P}(\xi_0 = 0) = \mathbb{P}(\xi_0 = 1) = \frac{1}{2}$) :

$$X_n = \frac{1}{2}(X_{n-1} + \xi_n).$$

This is also the Bernoulli shift (see definition below) $X_n = H(\xi_n)$ with $H(x) = \sum_{k=0}^{\infty} 2^{-(k+1)} x_k$. This model has the stationary uniform distribution on the interval but it satisfies no mixing condition (the past may entirely be recovered from present). For a more rigorous proof see [17], page 375.

Define also the class,

$$\mathcal{I} = \left\{ \bigotimes_{i=1}^{u} g_{x_i}; x_i \in \mathbb{R}_+^*, u \in \mathbb{N}^* \right\}, \text{ where } g_x(y) = \mathbb{1}_{x \le y} - \mathbb{1}_{y \le -x}, \forall x \in \mathbb{R}_+^*. \qquad (3)$$

The following lemma links $\mathcal{I}$-weak dependence with $\mathcal{L}$-weak dependence. Indeed, examples are proved to satisfy a weak dependence condition w.r.t. the class $\mathcal{L}$. Consider the weaker $\mathcal{L}_0 \cap \mathcal{C}_b^1$-weak dependence condition defined with

$$\mathcal{L}_0 = \left\{ \bigotimes_{i=1}^{u} f_i;\ f_i \in \mathcal{L},\ f_i : \mathbb{R} \to \mathbb{R}, i = 1, \ldots, u,\ u \in \mathbb{N}^* \right\}.$$

Here $\mathcal{C}_b^1$ stands for the set of partially differentiable functions with bounded partial derivatives and the function ψ is either ψ_1 or ψ_2. This latter condition will thus imply $\mathcal{I}$-weak dependence under concentration assumptions.

Lemma 2.2 *Let (X_n) be a sequence of r.v's. Suppose that, for some real $\alpha > 0$*

$$C(\lambda) = \sup_{x \in \mathbb{R}} \sup_i \mathbb{P}\left(x \le X_i \le x + \lambda\right) \le C\lambda^\alpha. \qquad (4)$$

Suppose that the sequence (X_n) is

- *$(\theta, \mathcal{L}_0 \cap \mathcal{C}_b^1, \psi_1)$-weakly dependent, then it is $(\theta_{\mathcal{I}}, \mathcal{I}, \psi)$-weakly dependent with*

$$\theta_{\mathcal{I},r} = \theta_r^{\frac{\alpha}{1+\alpha}} \text{ and } \psi(h,k,u,v) = 2(8C)^{\frac{1}{1+\alpha}}(u+v).$$

- *$(\theta, \mathcal{L}_0 \cap \mathcal{C}_b^1, \psi_2)$-weakly dependent, then it is $(\theta_{\mathcal{I}}, \mathcal{I}, \psi)$-weakly dependent with*

$$\theta_{\mathcal{I},r} = \theta_r^{\frac{\alpha}{2+\alpha}} \text{ and } \psi(h,k,u,v) = (8C)^{\frac{2}{2+\alpha}}(u+v)^{\frac{2(1+\alpha)}{2+\alpha}}.$$

In order to understand better the mechanism of proofs we recall the proof of this lemma.

Proof of Lemma 2.2. First recall the following useful inequality, valid for real numbers $0 \le x_i, y_i \le 1$, $|x_1 \dots x_m - y_1 \dots y_m| \le \sum_{i=1}^m |x_i - y_i|$. Let $g, f \in \mathcal{I}$, then $g(y_1, \dots, y_u) = g_{x_1}(y_1) \dots g_{x_u}(y_u)$, and $f(y_1, \dots, y_v) = g_{x'_1}(y_1) \dots g_{x'_v}(y_v)$ for some $u, v \in \mathbb{N}^*$ and $x_i, x'_j \ge 0$. For fixed $x > 0$ and $a > 0$ let

$$f_x(y) = \mathbb{1}_{y>x} - \mathbb{1}_{y\le -x} + \left(\frac{y}{a} - \frac{x}{a} + 1\right) \mathbb{1}_{x-a<y<x} + \left(\frac{y}{a} + \frac{x}{a} - 1\right) \mathbb{1}_{-x<y<-x+a}.$$

Therefore $\mathrm{Lip}(f_x) = a^{-1}$ and $\|f_x\|_\infty = 1$ and $\mathrm{Lip}(h) \le a^{-1}$, $\mathrm{Lip}(k) \le a^{-1}$ if we set

$$h(y_1, \dots, y_u) = f_{x_1}(y_1) \dots f_{x_u}(y_u), \qquad k(y_1, \dots, y_v) = f_{x'_1}(y_1) \dots f_{x'_v}(y_v).$$

Consider $i_1 \le \dots \le i_u \le i_u + r \le j_1 \le \dots \le j_v$ and set

$$\mathrm{Cov}(h, k) := \mathrm{Cov}(h(X_{i_1}, \dots, X_{i_u}), k(X_{j_1}, \dots, X_{j_v})).$$

• $(\theta, \mathcal{L}_0 \cap \mathcal{C}_b^1, \psi_1)$ (resp. $(\theta, \mathcal{L}_0 \cap \mathcal{C}_b^1, \psi_2)$)-weak dependence implies, with either $c(u, v) = (u + v)$ or $c(u, v) = (u + v)^2$,

$$|\mathrm{Cov}(h, k)| \le \frac{1}{a} c(u, v)\theta_r \qquad \left(\text{resp. } |\mathrm{Cov}(h, k)| \le \frac{1}{a^2} c(u, v)\theta_r\right).$$

• Inequality (4) yields $|\mathrm{Cov}(g, f) - \mathrm{Cov}(h, k)| \le 8Ca^\alpha(u + v)$. Hence $|\mathrm{Cov}(g, f)| \le 8Ca^\alpha(u + v) + \frac{1}{a} c(u, v)\theta_r$ (resp. $|\mathrm{Cov}(g, f)| \le 8Ca^\alpha(u + v) + \frac{1}{a^2} c(u, v)\theta_r$).
The Lemma follows by setting

$$a = \left[\frac{c(u, v)\theta_r}{8C(u + v)}\right]^{1/(1+\alpha)} \qquad \left(\text{resp. } a = \left[\frac{c(u, v)\theta_r}{8C(u + v)}\right]^{1/(2+\alpha)}\right).$$

2.3 Mixing

For the completeness sake, we recall here the definitions of the main mixing coefficients. For more details, the reader is deferred to Doukhan [11].

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let $\mathcal{U}, \mathcal{V}$ be two sub σ-algebras of $\mathcal{A}$, various measures of dependence between $\mathcal{U}$ and $\mathcal{V}$ have been introduced; among them let us recall

$$\begin{aligned}
\alpha(\mathcal{U}, \mathcal{V}) &= \sup\{|\mathbb{P}(U \cap V) - \mathbb{P}(U)\mathbb{P}(V)|; U \in \mathcal{U}, V \in \mathcal{V}\}, \\
\beta(\mathcal{U}, \mathcal{V}) &= \mathbb{E} \sup\{|\mathbb{P}(V|\mathcal{U}) - \mathbb{P}(V)|; V \in \mathcal{V}\}, \\
\rho(\mathcal{U}, \mathcal{V}) &= \sup\{|\mathrm{Corr}(u, v)|; u \in \mathbb{L}^2(\mathcal{U}), v \in \mathbb{L}^2(\mathcal{V})\}, \\
\phi(\mathcal{U}, \mathcal{V}) &= \sup\{|\mathbb{P}(V|U) - \mathbb{P}(V)|; U \in \mathcal{U}, V \in \mathcal{V}\}.
\end{aligned}$$

Those coefficients are respectively Rosenblatt's [28] strong mixing coefficient, $\alpha(\mathcal{U}, \mathcal{V})$, Wolkonski and Rozanov's absolute regularity coefficient in [36], $\beta(\mathcal{U}, \mathcal{V})$, Kolmogorov and Rozanov' maximal correlation coefficient $\rho(\mathcal{U}, \mathcal{V})$ [18], and $\phi(\mathcal{U}, \mathcal{V})$ the uniform mixing coefficient from Ibragimov [17].
Let $X = (X_n)_{n\in\mathbb{Z}}$ be a discrete time stationary process. We set $X_A = \{X_t; t \in A\}$ if $A \subset \mathbb{Z}$ for the A-marginal of X. At last $\sigma(Z)$ will denote the sigma-algebra generated by a random variable Z.
For any coefficient previously defined, say $c(\cdot, \cdot)$, we shall call the process $\mathbf{X}$ a c-mixing process if $\lim_{k\to\infty} c_{\mathbf{X},k} = 0$ where $c_{\mathbf{X},k} = c(\sigma(X_{]-\infty,0]}), \sigma(X_{[k,+\infty[})$.

$$\phi-mixing \Rightarrow \rho-mixing \Rightarrow \alpha-mixing, \quad \text{and} \quad \phi-\text{mixing} \Rightarrow \beta-\text{mixing} \Rightarrow \alpha-\text{mixing}$$

and no other implication holds.

3 Models

3.1 Markovian Models

Let $(\xi_n)_{n\in\mathbb{Z}}$ be an i.i.d. sequence and let M be some measurable function. We turn our attention to models driven by the equation

$$X_n = M(X_{n-1}, ..., X_{n-p}, \xi_n). \tag{5}$$

In order to justify the title of this section, remark that the vector valued sequence $(X_n^{(p)})_{n\in\mathbb{Z}}$, where $X_n^{(p)} = (X_{n-1}, ..., X_{n-p})$ is Markovian.

Such models are associated with ergodic Markov chains (see [21] for a review of the main models). Thus the stationarity assumption is reachable. An interesting subclass of examples is given by two functions R and S and two mutually independent i.i.d. sequences (ξ_n) and (η_n)

$$X_n = R(X_{n-1}, ..., X_{n-p}, \eta_n) + S(X_{n-1}, ..., X_{n-p})\xi_n.$$

Here the function S satisfies $S(x_1, ..., x_p) \geq s > 0$ for some $s \in \mathbb{R}$, and any real numbers $x_1, ..., x_p$ and the functions R and S essentially satisfy contraction assumptions (see [11], [14], [2] or [34] for developments).

For instance $ARMA(p, q)$ processes

$$Y_n = \sum_{i=1}^{p} a_i Y_{n-i} + \xi_n + \sum_{j=1}^{q} b_j \xi_{n-j},$$

have such a Markov representation. Indeed $X_n = (Y_n, Y_{n,n-1} \ldots, Y_{n,n-\ell})$, where $Y_{n,j} = \mathbb{E}\left[Y_n \mid Y_{n+i} : i \leq n\right]$, and $\ell = \max\{p, q+1\}$, is a Markov process. The mixing properties of these models are developed in [22].

Lipschitzian models are shown to be exponentially $\mathcal{L}$-weakly dependent.

Proposition 3.1 ([14]) . *Let a vector valued Markov model be defined through the recurrence relation (5). Assume that*

$$\mathbb{E}\|M(x, \xi_n) - M(y, \xi_n)\|^S \leq a\|x - y\|^S \quad \textit{and} \quad \mathbb{E}\|M(0, \xi_n)\|^S < \infty,$$

for some $0 \leq a < 1$ *and* $S \geq 1$.
Then the sequence $(X_n)_{n\in\mathbb{Z}}$ *is* $(\theta, \mathcal{L}, \psi)$*-weakly dependent with* $\theta_r = \mathcal{O}(a^r)$ *and* $\psi(h, k, u, v) = v\mathrm{Lip}k$.

Only recall that here stationarity is not required.

More general $AR(p)$ nonlinear models, $X_n = M(X_{n-1}, \ldots, X_{n-p}; \xi_n)$, have the same properties: if, for example, $\mathbb{E}|M(0; \xi_n)| < \infty$ and, for some constants $a_j \geq 0, 1 \leq j \leq p$ with $\sum_{j=1}^{p} a_j < 1$,

$$\mathbb{E}|M(x_1, \ldots, x_p; \xi_n) - M(y_1, \ldots, y_p; \xi_n)| \leq \sum_{j=1}^{p} a_j |x_j - y_j|.$$

This model is geometrically $(\theta, \mathcal{L}, \psi)$-weakly dependent with $\psi(h, k, u, v) = \min\{u\mathrm{Lip}h, v\mathrm{Lip}k\}$ as proved in [13].

Remark. Under the assumption that ξ_0's distribution has an almost sure non vanishing density f and is integrable the additive model

$$M(X_{n-1}, \ldots, X_{n-p}, \xi_n) = R(X_{n-1}, \ldots, X_{n-p}) + \xi_n$$

is shown to be ergodic and mixing. More precisely, if the function R is continuous, and $|R(x_1, \ldots, x_p)| \leq A + a_1|x_1| + \cdots + a_p|x_p|$ with $a_1 + \cdots + a_p < 1$, under the invariant initial distribution, the sequence is absolutely regular with $\beta_n = \mathcal{O}(e^{-bn})$ (see [8]).

3.2 Bernoulli shifts

Definition 3.2 *Let $(\xi_i)_{i\in\mathbb{Z}}$ be a sequence of i.i.d. real valued r.v's and the function $H : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ be measurable. The sequence $(X_n)_{n\in\mathbb{Z}}$ is called a Bernoulli shift if it is defined by $X_n = H(\xi_{n-j}, j \in \mathbb{Z})$.*

Causal shifts write as $X_n = H(\xi_n, \xi_{n-1}, \xi_{n-2}, \ldots, \xi_0, \xi_{-1}, \xi_{-2}, \ldots)$, i.e. $H : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$. Such shifts provide natural examples of weakly dependent, but not mixing sequences (see [27]).

Notice that a stationary Markov setting provides such causal sequences. In fact, consider a Markov process driven by the recurrence equation $X_t = M(X_{t-1}, \xi_t)$, for some i.i.d. sequence $(\xi_t)_{t\in\mathbb{Z}}$. Then the function H is defined, *if it exists*, implicitly by the relation $H(x) = M(H(x'), x_0)$, where $x = (x_0, x_1, x_2, \ldots), x' = (x_1, x_2, x_3, \ldots)$. Control theory yields tools to provide explicit Bernoulli shift representations (see [22] and [21]).

3.2.1 Chaotic representations

We now specialize in chaotic expansions associated with the discrete chaos generated by the sequence $(\xi_i)_{i\in\mathbb{Z}}$; in a condensed formulation we write $H(x) = \sum_{k=0}^{\infty} H^{(k)}(x)$, where $H^{(k)}(x)$ denotes the k-th order chaos contribution, $H^{(0)}(x) = a_0^{(0)}$ is only a centering constant, and

$$H^{(k)}(x) = \sum_{j_1=-\infty}^{\infty} \sum_{j_2=-\infty}^{\infty} \cdots \sum_{j_k=-\infty}^{\infty} a^{(k)}_{j_1,\ldots,j_k} x_{j_1} \times x_{j_2} \times \ldots \times x_{j_k}$$

or in short, in a vectorial notation, $H^{(k)}(x) = \sum_{j\in\mathbb{Z}^k} a_j^{(k)} x_j$.
In contrast with mixing conditions, it may be proved that even non causal sequences may admit a $\mathcal{L}$-weakly dependent behaviour. This is a by-product of the proposition 3.4 to follow.

Definition 3.3 *For any integer $k > 0$, we denote δ_k any number such that*

$$\sup_{i\in\mathbb{Z}} \mathbb{E}|H(\xi_{i-j}, j \in \mathbb{Z}) - H(\xi_{i-j}\mathbb{1}_{|j|<k}, j \in \mathbb{Z})| \leq \delta_k.$$

Such sequences $(\delta_k)_{k\in\mathbb{Z}^+}$ are related to the modulus of uniform continuity of H. If, for instance, for positive constants $(a_i)_{i\in\mathbb{Z}}, 0 < b \leq 1$, the inequality $|H(u_i, i \in \mathbb{Z}) - H(v_i, i \in \mathbb{Z})| \leq \sum_{i\in\mathbb{Z}} a_i|u_i - v_i|^b$ holds for any sequence $(u_i), (v_i) \in \mathbb{R}^{\mathbb{Z}}$ and if the sequence $(\xi_i)_{i\in\mathbb{Z}}$ has a finite b-th order moment, then one can choose $\delta_k = \sum_{|i|\geq k} a_i\mathbb{E}|\xi_i|^b$.

Proposition 3.4 ([13]) *Bernoulli shifts are $(\theta, \mathcal{L}, \psi)$-weakly dependent with $\theta_r = 2\delta_{r/2}$, and with the function $\psi(h, k, u, v) = 4\,(u\|k\|_\infty \mathrm{Lip}(h) + v\|h\|_\infty \mathrm{Lip}(k))$. Under causality, this holds with $\theta_r = \delta_r$ and $\psi(h, k, u, v) = 2v\mathrm{Lip}k\|h\|_\infty$.*

Processes associated with a finite number of chaotic terms (i.e. $H^{(k)} = 0$ if $k > k_0$) are also called *Volterra processes.* A suitable bound for δ_r corresponds here with the stationarity condition

$$\delta_r = \sum_{k=0}^{\infty} \left\{ \sum_{j\in\mathbb{Z}^k; \|j\|_\infty > r} \left| a_j^{(k)} \right| \right\} \mathbb{E}|\xi_0|^k < \infty.$$

The first example of such a Volterra process is clearly the class of linear processes $X_t = \sum_{-\infty}^{\infty} a_k \xi_{t-k}$. A suitable sequence is $\theta_r = 2\mathbb{E}|\xi_0| \sum_{|k|>r/2} |a_k|$ with the previous function ψ_1.
The simple bilinear process, $X_t = (a + b\xi_{t-1})X_{t-1} + \xi_t$, is stationary if $c = \mathbb{E}|a + b\xi_0| < 1$ (see [33]). It is a Bernoulli shift with $H(x) = x_0 + \sum_{j=1}^{\infty} x_j \prod_{s=1}^{j} (a + bx_s)$, for $x = (x_i)_{i\in\mathbb{N}}$. We truncate the previous series up to rank r in order to obtain $\delta_r = \theta_r = c^r(r+1)/(1-c)$.
Giraitis, Koul and Surgailis have introduced $ARCH(\infty)-$models (from Giraitis, Koul and Surgailis in [15]. We are given a nonnegative sequence $(b_j)_{j\geq1}$ and a i.i.d. sequence of nonnegative random variables $(\xi_j)_{j\geq0}$. The process (if it exists) is ruled through the recurrence relation

$$X_t = \left[a + \sum_{j=1}^{\infty} b_j X_{t-j} \right] \xi_t.$$

Such models are proved to have a stationary representation with the chaotic expansion

$$X_t = a \sum_{\ell=1}^{\infty} \sum_{j_1=0}^{\infty} \cdots \sum_{j_\ell=0}^{\infty} b_{j_1} \cdots b_{j_\ell} \xi_{t-j_1} \cdots \xi_{t-[j_1+\cdots+j_\ell]}$$

under the simple assumption $\mathbb{E}\xi_0 \sum_{j=1}^{\infty} b_j < 1$. They extend the standard ARCH(p) model ($b_j = 0$ if $j > p$). Here we have $\theta_r = \left(\mathbb{E}\xi_0 \sum_{j=1}^{\infty} b_j \right)^r$.

3.2.2 Association

Definition 3.5 *The sequence $(Z_t)_{t\in\mathbb{Z}}$ is associated, if for all coordinatewise increasing real-valued functions h and k,*

$$\mathrm{Cov}(h(Z_t, t \in A), k(Z_t, t \in B)) \geq 0$$

for all finite subsets A and B of $\mathbb{Z}$.

Associated sequences are $(\theta, \psi_2, \mathcal{L})$-weakly dependent with $\theta_r = \sup_{k\geq r} \mathrm{Cov}(X_0, X_r)$ (see [13]). Note that broad classes of examples of associated processes result from the fact that any independent sequence is associated and that monotonicity preserves association (for this, see [23]).

The case of Gaussian sequences is analogous by setting $\theta_r = \sup_{k\geq r} |\mathrm{Cov}(X_0, X_k)|$. Also, one may consider combinations of sums of Gaussian and associated sequences, or Bernoulli shifts driven by stationary, associated, instead of i.i.d. sequences.

Note that for associated or Gaussian sequences, ψ_2' replaces ψ_2 if $\theta_r = \sup_{k\geq r} |\mathrm{Cov}(X_0, X_k)|$ is replaced by $\theta_r = \sum_{k\geq r} |\mathrm{Cov}(X_0, X_k)|$.

Remark. The causal linear process $X_n = \sum_{t=0}^{\infty} a_t \xi_{n-t}$, satisfies a β-mixing condition if ξ_0's density and (for some $\delta > 0$) the $1+\delta$ order moment exist, together with the condition $\sum_{t=-\infty}^{\infty} |a_t|^{\delta} < \infty$. Then [25] prove that $\beta_n \le C \sum_{l=n}^{\infty} \left(\sum_{k=l}^{\infty} |a_k|\right)^{\delta/1+\delta}$, for some $C > 0$. If, for instance $a_j = \mathcal{O}(j^{-a})$, then under the previous regularity and moment conditions yields if $a > 2 + 1/\delta$, $\beta_n \sim n^{-b}$ with $b = (a-2)\delta/(1+\delta)$. For instance, $\sum_{n=0}^{\infty} \beta_n < \infty$ if $a > 3 + 2/\delta$. If $\delta = 1$ this writes $a > 5$. If $\delta = \infty$ this writes $a > 3$.

4 Algebraic moments of sums (see [13])

Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of centered r.v's and let $S_n = \sum_{i=1}^{n} X_i$. In this section, we give moment bounds for $|\mathbb{E}S_n^q|$, when $q \in \mathbb{N}$ and $q \ge 2$. Let (X_n) be a sequence of centered r.v's and define for positive integer r coefficients of weak dependence as non decreasing sequences $(C_{r,q})_{q\ge 2}$ such that

$$\left|\mathrm{Cov}\left(X_{t_1} \times \cdots \times X_{t_m}, X_{t_{m+1}} \times \cdots \times X_{t_q}\right)\right| \le C_{r,q}, \tag{6}$$

for all the successions $\{t_1, \ldots, t_q\}$ such that $1 \le t_1 \le \ldots \le t_q \le n$ and, for some integer $1 \le m < q$, $t_{m+1} - t_m = r$. Explicit bounds $C_{r,q}$ are provided by [13] in order to build inequalities for the partial sums S_n. Two kinds of bounds are considered, either

$$\left|\mathrm{Cov}\left(X_{t_1} \times \cdots \times X_{t_m}, X_{t_{m+1}} \times \cdots \times X_{t_q}\right)\right| \le c q^{\gamma} M^{q-2} \theta_r \tag{7}$$

or,

$$\left|\mathrm{Cov}\left(X_{t_1} \times \cdots \times X_{t_m}, X_{t_{m+1}} \times \cdots \times X_{t_q}\right)\right| \le c \int_0^{\theta_r} Q_{X_{t_1}}(x) \times \cdots \times Q_{X_{t_q}}(x) dx, \tag{8}$$

where Q_X still denotes the X's quantile function and $c, \gamma \ge 0$ denote real numbers.
In the examples, the previous bound (7) holds for bounded sequences such that $\|X_n\|_\infty \le M$. Under $(\theta, \mathcal{L}, \psi)$-weak dependence, this yields the bound

$$C_{r,q} = \max_{1\le m<q} \psi(j^{\otimes m}, j^{\otimes(q-m)}, m, q-m) M^q \theta_r$$

where $j(x) = x\mathbb{1}_{|x|\le 1} + \mathbb{1}_{x>1} - \mathbb{1}_{x<-1}$.
As in lemma 2.2, we see that under $(\theta, \mathcal{L}, \psi)$-weak dependence with $\psi(h,k,u,v) = c(u,v)\mathrm{Lip}(h)\mathrm{Lip}(k)$ a bound is

$$C_{r,q} = \max_{1\le m<q} c(m, q-m) M^{q-2}\theta_r.$$

For non bounded random variables,

$$C_{r,q} = c \int_0^{\theta_r} Q_{X_{t_1}}(x) \times \cdots \times Q_{X_{t_q}}(x) dx$$

under $(\theta, \mathcal{I}, \psi)$-weak dependence.
An analogous bound is obtained by Rio for strongly mixing sequences (see [26]). The bound (8) holds for more general r.v's, using moment or tail assumptions.
A first consequence of inequality (7) is the following Marcinkiewicz-Zygmund inequality.
Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of centered r.v's satisfying the condition

$$C_{r,q} = \mathcal{O}(r^{-q/2}), \tag{9}$$

then, there is a constant $B > 0$ not depending on n for which

$$|\mathbb{E}S_n^q| \le Bn^{q/2}. \tag{10}$$

The following lemma gives moment inequalities whenever $q \in \{2, 4\}$, it was essentially proved in Billingsley ([5], lemmas 3 and 4, page 172).

Lemma 4.1 ([13]) *If $(X_n)_{n\in\mathbb{N}}$ is a sequence of centered r.v's, then*

$$\mathbb{E}S_n^2 \le 2n\sum_{r=0}^{n-1} C_{r,2}, \quad \mathbb{E}S_n^4 \le 4!\left\{\left(n\sum_{r=0}^{n-1} C_{r,2}\right)^2 + n\sum_{r=0}^{n-1}(r+1)^2 C_{r,4}\right\}. \tag{11}$$

The following Theorems deal with higher order moments.

Theorem 4.2 ([13]) *Let $q \ge 2$ be some integer. Suppose that dependence coefficients $C_{r,p}$ associated to the sequence (X_n) satisfy for all integers $0 < p \le q$ and for some positive constants M, γ, C*

$$C_{r,p} = Ce^{\gamma p} M^{p-2}\theta_r, \tag{$\mathcal{H}_1$}$$

then, for any integer $n \ge 2$

$$|\mathbb{E}S_n^q| \le \frac{(2q-2)!}{(q-1)!}e^{q\gamma}\left\{\left(Cn\sum_{r=0}^{n-1}\theta_r\right)^{q/2} \vee \left(CM^{q-2}n\sum_{r=0}^{n-1}(r+1)^{q-2}\theta_r\right)\right\}. \tag{12}$$

Theorem 4.2 is adapted to work with bounded sequences. Define the class $\mathcal{I}$

$$\mathcal{I} = \left\{\bigotimes\nolimits_{i=1}^{u} g_{x_i}; x_i \in \mathbb{R}_+^*, u \in \mathbb{N}^*\right\}, \text{ where } g_x(y) = \mathbb{1}_{x\le y} - \mathbb{1}_{y\le -x}, \forall x \in \mathbb{R}_+^*. \tag{13}$$

In order to consider the unbounded case, we shall consider $(\theta, \mathcal{I}, \psi)$-weak dependence where ψ writes $\psi(h, k, u, v) = c(u, v)$.

Theorem 4.3 ([13]) *If $(X_n)_{n\in\mathbb{N}}$ is a centered and $(\theta, \mathcal{I}, \psi)$-weakly dependent sequence, set $C_q = (\max_{u+v\le q} c(u, v)) \vee 2$. Then*

$$|\mathbb{E}S_n^q| \le \frac{(2q-2)!}{(q-1)!}\left(C_q\sum_{i=1}^{n}\int_0^1 (\theta^{-1}(u)\wedge n)^{q-1}Q_i^q(u)du \vee \left(C_2\sum_{i=1}^{n}\int_0^1 (\theta^{-1}(u)\wedge n)Q_i^2(u)du\right)^{q/2}\right).$$

In the special case of strongly mixing and stationary sequences, this is Theorem 1 in [26]. The restriction of working with even integer exponents finds its compensation in the explicit form of the constants.

Remark. Exponential inequalities may be built by using Theorem 4.2. Define

$$M_{q,n} = n\sum_{r=0}^{n-1}(r+1)^{q-2}C_{r,q}.$$

We first suppose that for all integers $q \ge 2$ and n:

$$C_{r,p} = Ce^{\gamma p}M^{p-2}\theta_r, \quad M_{q,n} \le A_n\frac{q!}{\beta^q},$$

where β is some constant and A_n is a sequence independent of q, then

$$\forall x \in \mathbb{R}_+^*, \qquad \mathbb{P}\left(|S_n| \geq x\sqrt{A_n}\right) \leq A \exp\left(-B\sqrt{\beta x}\right), \qquad (DP)$$

with $A = e^{4+1/12}\sqrt{8\pi}$, and $B = e^{5/2-\gamma/2}$. This exponential inequality is analogous to ??? Let (X_n) be a sequence of centered r.v.'s, this holds if $C_{r,q} = C\sigma^2 M^{q-2} e^{\gamma q} e^{-br}$ for constants $C, \sigma, \gamma, b > 0$, if $\|X_n\|_\infty \leq M$ and $\|X_n\|_2 \leq \sigma$, for any integer $n \geq 0$. In this case $A_n = n\sigma^2$. E.g. this holds under $(\theta, \mathcal{L}, \psi)$-weak dependence if $\theta_r = \mathcal{O}(e^{-br})$ and $\psi(h,k,u,v) \leq e^{\delta(u+v)}\text{Lip}(h)\text{Lip}(k))$ for some $\delta \geq 0$.

The use of combinatorics in those inequalities makes them relatively weak. E.g. Bernstein inequality, valid for independent sequences allows to replace the term $\sqrt{x}$ in the previous inequality by x^2 under the same assumption $n\sigma^2 \geq 1$; in mixing cases analogue inequalities are also obtained by using coupling arguments which are not available here.

5 Limit theorems

5.1 The Donsker line

Consider a stationary sequence $(X_n)_{n\in\mathbb{Z}}$. We assume that this sequence is integrable, centered at expectation, and that

$$\mathbb{E}X_0 = 0.$$

Denote by $[x]$ the integral part of a real number x ($[x] \leq x < [x]+1$), the Donsker line $(D_n(t))_{t\in[0,1]}$ is defined for any sample with size n as the following continuous time process

$$D_n(t) = \sum_{k=1}^{[nt]} X_k + (nt - [nt])X_{[nt]+1}.$$

Let $(W_t)_{t\in[0,1]}$ be a Brownian motion, that is W denotes the centered Gaussian real valued process with covariance

$$\mathbb{E}W_s W_t = \min\{s,t\}.$$

We consider the following convergence result in the space $\mathcal{C}([0,1])$ of continuous functions on the unit interval when the sample size n grows to infinity:

Theorem 5.1 *Suppose that the stationary sequence $(X_n)_{n\in\mathbb{Z}}$ satisfies $\mathbb{E}|X_0|^{4+\delta} < \infty$ for some $\delta > 0$. Assume a $(\theta, \mathcal{I}, \psi)$-weak dependence condition with the function $\psi_1(h,k,u,v)$ (respectively ψ_2) and $\theta_r = \mathcal{O}(r^{-a})$ with $a \geq 2 + 4/\delta$ (respectively $a > 2$).*
Then the following functional convergence holds in the space $\mathcal{C}([0,1])$:

$$\frac{1}{\sqrt{n}} D_n(t) \to \sigma W_t.$$

The following series is assumed to be convergent

$$\sigma^2 = \sum_{-\infty}^{\infty} \text{Cov}(X_0, X_k).$$

The case $\sigma^2 = 0$ is detailed by [24]. We shall assume here that $\sigma^2 \neq 0$.

Remark. Without any regularity condition on innovations, theorem 5.1 holds for a bounded Lipschitz function of a linear process $X_n = \sum_{k=-\infty}^{\infty} a_k \xi_{n-k}$ if $a_k = \mathcal{O}(k^{-D})$ when $D > 3$. The latter doesn't need to be causal while Ho and Hsing need this causality assumption in [16].

Proof of Theorem 5.1. Lemma 4.2 and a maximal inequality by Moricz *et alii* (see [11], page 40) yield

$$\mathbb{E}|S_n|^{2+\delta} = \mathcal{O}\left(n^{1+\delta/2}\right) \tag{14}$$

as soon as for any increasing sequence of integers $0 \le i < j < k \le l$

$$\sum_{m=0}^{\infty} m\mathbb{E}|X_0 X_m| < \infty \text{ and } \mathrm{Cov}\,(X_i X_j, X_k X_l) = \mathcal{O}\left((k-j)^{-2}\right). \tag{15}$$

Moreover, this entails that $\sigma^2 = n^{-1}\mathrm{Var}(S_n) > 0$ and the finite dimensional (fidi) convergence follows. The tightness of the process is a consequence of (15). The first part of (15) follows from the covariance bound $|\mathrm{Cov}\,(X_0, X_r)\,| =) = \mathcal{O}\left(\theta_r^{(2+\delta)/(4+\delta)}\right)$. The latter bound follows from $\mathrm{Cov}\,(X_i X_j, X_k X_l) = \mathcal{O}\left(\theta_{k-j}^{\delta/(4+\delta)}\right)$.

Remark. Let $\delta > 0$. Assume the existence of a moment of order $(2+\delta)$ for X_0. The Donsker functional CLT holds in the strong mixing case if $\sum_{n=0}^{\infty} n^{2/\delta}\alpha_n < \infty$ (see [26]). The condition $\sum_{n=0}^{\infty} \rho(2^n) < \infty$ with $\mathbb{E}X_0^2 < \infty$ implies the functional convergence, as Shao proved [30].

Remark. Here we consider conditions in terms of conditional expectations with respect to an adapted filtration. We first recall that theorem 5.1 holds for martingales with stationary square integrable increments $\mathbb{E}X_0^2 < \infty$ (see [5]). Let $(\mathcal{M}_n)$ be a filtration adapted to the process $(X_n)_{n\in\mathbb{Z}}$: X_n is $\mathcal{M}_n$-measurable for any $n \in \mathbb{Z}$. Dedecker and Rio [9] prove that for a centered, square integrable process, such that the sum $\sum_{n=0}^{\infty} X_0\, \mathbb{E}\,(X_n|\mathcal{M}_0)$ is convergent in $\mathbb{L}^1$, the sequence $\mathbb{E}(X_0^2 + 2X_0 S_n|\mathcal{I})$ converges in $\mathbb{L}^1$ to some random variable η. The σ-field $\mathcal{I}$ is the tail σ-field. Moreover, conditionally to $\mathcal{I}$, $D_n(t)/\sqrt{n}$ converge to a Brownian motion ηW_t.

- This result provides a functional CLT to a limit process which is not Gaussian (see [6] for results related to the latter case).
- Note that ergodicity implies the triviality of the tail σ-field and a standard Donsker theorem follows.
- Standard results prove this theorem under a more restrictive $\mathbb{L}^2$ assumption: both series
$$\sum_{n=0}^{\infty} \mathbb{E}(X_n|\mathcal{M}_0) \text{ and } \sum_{n=0}^{\infty} (X_n - \mathbb{E}(X_n|\mathcal{M}_0))$$
converge.
- As a new result yielded by this theorem, consider a stationary Markov sequence $(\xi_n)_{n\in\mathbb{Z}}$ with stationary distribution μ and transition operator P. Let $X_n = g(\xi_n)$ be centered at expectation, nonlinear functionals of (ξ_n). Then the assumption writes as the convergence of the series $\sum_{n=0}^{\infty} gP^n g$ in $\mathbb{L}^1(\mu)$.

- The previous result concerning strongly mixing sequences appears also as a consequence of this theorem.

5.2 Empirical process

Let us consider a stationary sequence $(X_n)_{n\in\mathbb{Z}}$. We assume without loss of generality that the marginal distribution of this sequence is the uniform law on $[0,1]$. The empirical repartition process of the sequence (X_n) at time n is defined as $\frac{1}{\sqrt{n}}E_n(x)$ where

$$E_n(x)=\sum_{k=1}^n\left(\mathbb{1}_{(X_k\le x)}-\mathbb{P}(X_k\le x)\right).$$

Note that $E_n = n\,(F_n-F)$ if F_n, F respectively denote the empirical d.f. and the marginal d.f. We consider the following convergence result in the Skohorod space $\mathcal{D}(\mathbb{R})$ when the sample size n converges to infinity:

$$\frac{1}{\sqrt{n}}E_n(x)\to\bar{B}(x).$$

Here $(\bar{B}(x))_{x\in\mathbb{R}}$ is the dependent analogue of a Brownian bridge, that is $\bar{B}$ denotes the centered Gaussian process with covariance given by

$$\mathbb{E}\bar{B}(x)\bar{B}(y)=\sum_{k=-\infty}^{\infty}\left(\mathbb{P}(X_0\le x,\ X_k\le y)-\mathbb{P}(X_0\le x)\mathbb{P}(X_k\le y)\right). \tag{16}$$

Note that for independent sequences with a marginal repartition function F, this only writes $\bar{B}(x)=B(F(x))$ for some standard Brownian Bridge B; this justifies the name of Generalized Brownian Bridge.

Let (X_n) be a stationary sequence assumed to satisfy the following weak dependence condition.

$$\sup_{f\in\mathcal{F}}\left|\mathrm{Cov}\left(\prod_{i=1}^{2}f(X_{t_i}),\prod_{i=3}^{4}f(X_{t_i})\right)\right|\le\theta_r, \tag{17}$$

where $\mathcal{F}=\{x\to\mathbb{1}_{s<x\le t},\text{ for } s,t\in[0,1]\}$, $0\le t_1\le t_2\le t_3\le t_4$ and $r=t_3-t_2$ (in this case a weak dependence condition holds for a class of functions $\mathbb{R}^u\to\mathbb{R}$ working only with the values $u=1$ or 2).

Proposition 5.2 *Let (X_n) be a stationary sequence such that (17) holds. Assume that there exists $\nu>0$ such that*

$$\theta_r=\mathcal{O}(r^{-5/2-\nu}). \tag{18}$$

Then the sequence of processes $\left(\frac{1}{\sqrt{n}}\{E_n(t);t\in[0,1]\}\right)_{n>0}$ is tight in the Skohorod space $\mathcal{D}([0,1])$.

Theorem 5.3 *Suppose that the stationary sequence (X_n) is $(\theta,\mathcal{L}_1,\psi_j)$-weakly dependent, with either $j=1$ and $\theta_r=\mathcal{O}(r^{-15/2-\nu})$, or $j=2$ and $\theta_r=\mathcal{O}(r^{-5-\nu})$. Then the following empirical functional convergence holds true in the Skohorod space of numerical càdlàg functions on the real line, $D(\mathbb{R})$:*

$$\frac{1}{\sqrt{n}}E_n(x)\to\bar{B}(x).$$

If the sequence (X_n) is i.i.d., then equation (16) reduces to zero-index term. This covariance has essentially 1/2 order Holder regularity. If the second order marginals (X_0, X_k) have a joint, continuous density, then the $k-$order term is a C^2-function. Hence the regularity of the covariance is driven by the central term. In any case the process $\bar{B}$ is perhaps not derivable in quadratic mean.

Proof of Theorem 5.3. Thanks to the Rosenthal inequality in Theorem 4.2

$$\begin{aligned}\|E_n(t) - E_n(s)\|_4 &\leq C\sqrt{n}\sum_{r=0}^{n-1} \min\left\{r^{-5/2-\nu}, |t-s|\right\} + \left(n\sum_{r=0}^{n-1} (r+1)^2 \theta_r\right)^{1/4} \\ &\leq \sqrt{n}\left(|t-s|^{(a-1)/2a} + n^{(2-a)/4}\right).\end{aligned}$$

The conclusion follows from Shao and Yu' tightness lemma [32]. The fidi convergence is due to a CLT Lemma by blocks by Ibragimov [17].

Remark. If the sequence is strongly mixing, the summability condition $\sum_{n=0}^{\infty} \alpha_n < \infty$ implies fidi convergence. The empirical functional convergence holds if, moreover, for some $a > 1$, $\alpha_n = \mathcal{O}(n^{-a})$ (see [26]). In a absolutely regular framework, Doukhan, Massart and Rio (see [26]) obtain the empirical functional convergence when, for some $a > 2$, $\beta_n = \mathcal{O}(n^{-1}(\log n)^{-a})$. Shao and Yu [32] obtain the empirical functional convergence theorem when the maximal correlation coefficient satisfy the condition $\sum_{n=0}^{\infty} \rho(2^n) < \infty$.

5.3 A triangular CLT: Lindeberg-Rio's method

Let $X_{n,j}, 1 \leq j \leq n$ be a triangular array. We shall omit the first index n when possible without confusion. Consider a sequence $(W_j)_{j \in \mathbb{N}}$ of i.i.d. r.v.'s with standard normal law, and independent from $(X_n)_{t \in \mathbb{Z}}$. Define

$$\begin{aligned}\sigma_n^2 &= \text{Var}\,(S_n) \\ S_k &= \sum_{j=1}^{k} X_j,\ 1 \leq k \leq n,\ \text{and } S_0 = 0, \\ \tau_k &= \sum_{j=k}^{n} \sqrt{v_{n,j}} W_j = \sum_{j=k}^{n} V_j,\ 1 \leq k \leq n,\ \text{and } \tau_{n+1} = 0,\end{aligned}$$

where we assume that $v_{n,j} = \text{Var}\,(S_j) - \text{Var}\,(S_{j-1}) > 0$.

Once one has proved that $\sigma_n^2 \to \sigma^2$, it remains to prove that for any three times differentiable function with bounded derivatives up to order 3, φ say,

$$\Delta_n(\varphi) = \mathbb{E}\varphi\,(S_n) - \mathbb{E}\varphi\,(W_0) \to 0. \tag{19}$$

Consider the following:

$$U_j = S_{j-1} + \tau_{j+1},\ R_j(x) = \varphi\,(U_j + x) - \varphi\,(U_j) - \frac{v_j}{2}\varphi''\,(U_j)\ \ (1 \leq j \leq n).$$

Clearly, we have

$$\Delta_n(\varphi) = \sum_{k=1}^{n} \Delta_{n,k}(\varphi)$$

$$\begin{aligned}\Delta_{n,k}(\varphi) &= \mathbb{E}R_k(X_k) - \mathbb{E}R_k(V_k) \\ &= \Delta_k^1(\varphi) - \Delta_k^2(\varphi).\end{aligned}$$

A Taylor expansion yields

$$\begin{aligned}\Delta_k^2(\varphi) &= \mathbb{E}\left(\varphi\left(U_k+V_k\right) - \varphi\left(U_k\right) - V_k\varphi'\left(U_k\right) - \frac{V_k^2}{2}\varphi''\left(U_k\right)\right) \\ &= \frac{1}{6}\mathbb{E}\left(V_k^3\varphi^{(3)}\left(U_k+\vartheta_k V_k\right)\right), \text{ with } 0<\vartheta_k<1, \\ \left|\Delta_k^2(\varphi)\right| &\leq C\left(v_k\right)^{3/2}.\end{aligned}$$

Moreover,

$$\begin{aligned}\Delta_k^1(\varphi) &= \mathbb{E}\left(\varphi\left(U_k+X_k\right) - \varphi\left(U_k\right) - \frac{v_k}{2}\varphi''\left(U_k\right)\right) \\ &= \mathbb{E}\left(X_k\varphi'\left(U_k\right) + \frac{1}{2}\left(X_k^2 - \frac{v_k}{2}\right)\varphi''\left(U_k\right) + \frac{1}{6}X_k^3\varphi^{(3)}\left(U_k+\vartheta_k X_k\right)\right), \text{ with } 0<\vartheta_k<1.\end{aligned}$$

It then follows that

$$\begin{aligned}\sum_{k=1}^n \Delta_k^1(\varphi) &= \sum_{k=1}^n\sum_{j=1}^{k-1}\operatorname{Cov}\left(\varphi''\left(S_{k-1-j}+\tau_{j+1}\right)X_{k-j}, X_k\right) \\ &+ \frac{1}{2}\sum_{k=1}^n\sum_{j=1}^{k-1}\operatorname{Cov}\left(\varphi^{(3)}\left(S_{k-1-j}+\tau_{j+1}+\vartheta_{k-j}X_{k-j}\right)X_{k-j}^2, X_k\right) \\ &+ \sum_{k=1}^n\operatorname{Cov}\left(X_k, \varphi'\left(\tau_{k+1}\right)\right) + \frac{1}{2}\sum_{k=1}^n\mathbb{E}\left(\varphi''\left(U_k\right)\left(X_k^2-\mathbb{E}X_k^2\right)\right) \\ &+ \mathbb{E}\left(\sum_{k=1}^{n-1}\operatorname{Cov}\left(X_0,X_k\right)\sum_{j=k+1}^n\varphi''\left(U_j\right)\right) + \frac{1}{6}\sum_{k=1}^n\mathbb{E}\left(\varphi^{(3)}\left(U_k+\iota_k X_k\right)X_k^3\right) \\ &= E_1+E_2+E_3+E_4+E_5+E_6. \qquad (20)\end{aligned}$$

The term E_3 is null. The other ones are controlled in two ways: by uniform bounds on one hand, by dependence covariance bounds on the other hand. See [7] for further details.

We can apply this method to density estimation on $\mathbb{R}$. Consider a stationary sequence $(X_t)_{t\in\mathbb{Z}}$ with marginal density f. Assume that the densities of the pairs (X_0, X_k), $k\in\mathbb{Z}^+$, exist, and are uniformly bounded : $\sup_{k>0}\|f_{(k)}\|_\infty<\infty$.

Let K be some kernel function with integral 1, Lipschitzian and compactly supported. The kernel density estimator is defined by

$$\hat{f}(x) = \hat{f}_{n,h}(x) = \frac{1}{nh}\sum_{t=1}^n K\left(\frac{x-X_t}{h}\right).$$

Theorem 5.4 *Suppose that the stationary sequence $(X_t)_{t\in\mathbb{Z}}$ is $(\theta,\mathcal{L}_1,\psi_j')$-weakly dependent, with either $j=1,2$ and $\theta_r=\mathcal{O}(r^{-a}), a>2+j$, or $(\theta,\mathcal{L}_1,\psi_j)$-weakly dependent $j=1,2$ and $\theta_r=\mathcal{O}(r^{-a}), a>2+j+1/\delta$, where $n^\delta h\to\infty$ for some $\delta\in]0,1[$. If $f(x)>0$, then*

$$\sqrt{nh}\ \left(\hat{f}(x)-\mathbb{E}\hat{f}(x)\right)\xrightarrow{\mathcal{D}}\mathcal{N}\left(0, f(x)\int K^2(u)du\right)$$

Remark. In order to obtain a CLT for time series, it seems that the first paper using the Bernstein blocking technique in due to Dehling, see [10]. This method is also described in [11] and it is used in [13]. The technique consists in dividing a sample $\{Z_1, \ldots, Z_n\}$ into blocks $U_m = \{Z_k\}_{k \in K_m}$ which are almost independent and such that $\mathrm{Card}\,(\{1, \ldots, n\} \setminus \bigcup_m K_m) = o(n)$ as $n \to \infty$. This condition essentially allows to work with the variables $T_m = \sum_{k \in K_m} Z_k$ (built from those blocks) exactly as in the classical Lindeberg method of the independent case.

Proof of Theorem 5.4. The details of the proof follow [3].

6 Functional estimation

We consider a stationary processes $(Z_t)_{t \in \mathbb{Z}}$ with $Z_t = (X_t, Y_t)$ where $X_t,\ Y_t \in \mathbb{R}$. The quantity of interest is the regression function $r(x) = \mathbb{E}(Y_0 | X_0 = x)$. Let K be some kernel function with integral 1. Assume that K is a Lipschitz function with a compact support. The kernel estimator is defined by

$$\hat{f}(x) = \hat{f}_{n,h}(x) = \frac{1}{nh} \sum_{t=1}^{n} K\left(\frac{x - X_t}{h}\right), \qquad \hat{g}(x) = \hat{g}_{n,h}(x) = \frac{1}{nh} \sum_{t=1}^{n} Y_t K\left(\frac{x - X_t}{h}\right),$$
$$\hat{r}(x) = \hat{r}_{n,h}(x) = \frac{\hat{g}_{n,h}(x)}{\hat{f}_{n,h}(x)} \quad \text{if} \quad \hat{f}_{n,h}(x) \neq 0; \qquad \hat{r}(x) = 0 \text{ otherwise.}$$

Here $h = (h_n)_{n \in \mathbb{N}}$ is a sequence of positive real numbers. We always assume that $h_n \to 0,\ nh_n \to \infty$ as $n \to \infty$.

Definition 6.1 *Let $\rho = a + b$ with $(a, b) \in \mathbb{N} \times]0; 1]$. Define the set of ρ-regular functions $\mathcal{C}_\rho$ by*

$$\mathcal{C}_\rho = \Big\{ u : \mathbb{R} \to \mathbb{R};\ u \in \mathcal{C}_a \ \ \textit{and}\ \exists A \geq 0\ |u^{(a)}(x) - u^{(a)}(y)| \leq A|x - y|^b \\ \textit{for all } x, y \textit{ in any compact subset} \Big\}.$$

Here, $\mathcal{C}_a$ is the set of a-times continuously differentiable functions.

If $g \in \mathcal{C}_\rho$, one can choose a kernel function K of order ρ such that uniformly on any compact subset of $\mathbb{R}$

$$b_h(x) = \mathbb{E}(\hat{g}(x)) - g(x) = \mathcal{O}(h^\rho)$$

uniformly on any compact subset of $\mathbb{R}$. In view of asymptotic analysis we assume that the marginal density $f(.)$ of X_0 exists and is continuous. Moreover $f(x) > 0$ for any point x of interest and the regression function $r(.) = \mathbb{E}(Y_0 | X_0 = .)$ exists, is continuous. At last, for some $p \geq 1$, $x \to g_p(x) = f(x)\mathbb{E}(|Y_0|^p | X_0 = x)$ exists and is continuous. We set $g = fr$ with obvious short notation. Moreover, assume one of the following moment assumptions. Either,

$$\mathbb{E}|Y_0|^S < \infty, \qquad \text{for some} \qquad S \geq p \tag{21}$$

or

$$\mathbb{E} \exp\left(|Y_0|\right) < \infty. \tag{22}$$

6.1 Second Order Properties

We consider first the properties of $\hat{g}(x)$. We also consider the following conditionally centered equivalent of g_2 appearing in the asymptotic variance of the estimator $\hat{r}$,

$$G_2(x) = f(x)\text{Var }(Y_0|X_0 = x) = g_2(x) - f(x)r^2(x).$$

Assume that the densities of the pairs (X_0, X_k), $k \in \mathbb{Z}^+$, exist, and are uniformly bounded : $\sup_{k>0} \|f_{(k)}\|_\infty < \infty$. Moreover, uniformly over all $k \in \mathbb{Z}^+$, the functions

$$r_{(k)}(x, x') = \mathbb{E}\left(|Y_0 Y_k| \Big| X_0 = x, X_k = x'\right) \tag{23}$$

are continuous. Under these assumptions, the functions $g_{(k)} = f_{(k)} r_{(k)}$ are locally bounded.

Theorem 6.2 *Suppose that the stationary sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta, \psi_j, \mathcal{L})$-weakly dependent with $\theta_r = \mathcal{O}(r^{-a})$ and $a > 2 + j$, for $j = 1$ or $j = 2$. Assume that it satisfies the conditions (22) and (23) with $p = 2$. Suppose that $n^\delta h \to \infty$ for some $\delta \in]0, 1[$. Then uniformly in x belonging to any compact subset of $\mathbb{R}$,*

$$\text{Var }(\hat{g}(x)) = \frac{1}{nh}\, g_2(x) \int K^2(u)du \; + \; o\left(\frac{1}{nh}\right)$$

and

$$\text{Var }\left(\hat{g}(x) - r(x)\hat{f}(x)\right) = \frac{1}{nh} G_2(x) \int K^2(u)du + o\left(\frac{1}{nh}\right).$$

6.2 Central limit theorems

We first consider the density estimator.

Theorem 6.3 *Suppose that the stationary sequence $(X_t)_{t\in\mathbb{Z}}$ is either $(\theta, \mathcal{L}_1, \psi'_j)$-weakly dependent, with $\theta_r = \mathcal{O}(r^{-a}), a > 2 + j$, or $(\theta, \mathcal{L}_1, \psi_1)$-weakly dependent with $\theta_r = \mathcal{O}(r^{-a}), a > 3(1 + \sqrt{5})/2$, or $(\theta, \mathcal{L}_1, \psi_2)$-weakly dependent with $\theta_r = \mathcal{O}(r^{-a}), a > 6$. If where $nh \to \infty$ and $f(x) > 0$, then*

$$\sqrt{nh}\ \left(\hat{f}(x) - \mathbb{E}\hat{f}(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, f(x) \int K^2(u)du\right).$$

Proof of Theorem 6.3. The CLT is obtained by Lindeberg Rio method here above described and by Bernstein's blocking technique described in [17]. The Theorem follows by comparing the rates obtained by both methods. See [3] for more details.

Theorem 6.4 *Assume that the stationary sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta, \psi_j, \mathcal{L})$-weakly dependent with $\theta_r = \mathcal{O}(r^{-a})$ with $a > \min\left(\max(2 + j, 3(2 + j)\delta), \max\left(2 + j + \frac{1}{\delta}, \frac{2+2(2+j)\delta}{1+\delta}\right)\right)$, for $j = 1$ or $j = 2$. Assume moreover that it satisfies the conditions (22) and (23) with $p = 2$. Consider a positive kernel K. Let $f, g \in \mathcal{C}_\rho$ for some $\rho \in]0, 2]$, and $nh^{1+2\rho} \to 0$. Then, for all x belonging to any compact subset of $\mathbb{R}$,*

$$\sqrt{nh}\ \left(\hat{r}(x) - r(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\ \frac{G_2(x)}{f^2(x)} \int K^2(u)du\right).$$

Proof of Theorem 6.4. See [3]. They also prove the following

Proposition 6.5 *Assume that the stationary sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta,\psi_j,\mathcal{L})$-weakly dependent with $\theta_r=\mathcal{O}(r^{-a})$ with $a>9$ for $j=1$ and $a>12$ for $j=2$. Suppose that the stationary sequence $(Z_t)_{t\in\mathbb{Z}}$ satisfies the conditions (22) and (23) with $p=2$. Consider a positive kernel K. Let $f,g\in\mathcal{C}_\rho$ for some $\rho\in]0,2]$, and $nh\to\infty$. Then, for all x belonging to any compact subset of $\mathbb{R}$,*

$$\sqrt{nh}\,\left(\hat{r}(x)-\mathbb{E}\hat{r}(x)\right)\xrightarrow{\mathcal{D}}\mathcal{N}\left(0,\ \frac{G_2(x)}{f^2(x)}\int K^2(u)du\right).$$

Remark. The CLT convergence theorem 6.4 for the regression function holds true for strongly mixing sequences, under the moment assumption (21) if $h_n\to0$, $nh_n/\log(n)\to\infty$, $\alpha_n=\mathcal{O}(n^{-a})$ with $\alpha>2S/(S-2)$ ([27]). No positivity assumption on the kernel K is required.

6.3 Almost sure convergence

For the sake of simplicity, we only consider the geometrically dependent case.

Theorem 6.6 *Let $(Z_t)_{t\in\mathbb{Z}}$ be a stationary sequence satisfying the conditions (22), (23) with $p=2$, and that it is either $(\theta,\psi_1,\mathcal{L})$- or $(\theta,\psi_2,\mathcal{L})$-weakly dependent with $\theta_r\le a^r$ for some $0<a<1$.*

(i) *If $nh/\log^4(n)\to\infty$, then for any $M>0$, almost surely,*

$$\sup_{|x|\le M}|\hat{g}(x)-\mathbb{E}\hat{g}(x)|=O\left(\frac{\log^2(n)}{\sqrt{nh}}\right).$$

(ii) *For any $M>0$, if $f,g\in\mathcal{C}_\rho$ for some $\rho\in]0,\infty[$, $h\sim\left(\frac{\log^4(n)}{n}\right)^{1/(1+2\rho)}$ and $\inf_{|x|\le M}f(x)>0$, then, almost surely,*

$$\sup_{|x|\le M}|\hat{r}(x)-r(x)|=O\left\{\left(\frac{\log^4(n)}{n}\right)^{\rho/(1+2\rho)}\right\}.$$

Remark. Liebscher [19] proves the uniform almost sure convergence in a strongly mixing framework, at the optimal rate $\mathcal{O}\left(\left(\frac{\log(n)}{n}\right)^{\rho/(1+2\rho)}\right)$, if $\alpha_r=\mathcal{O}(r^{-a})$, with $a>4+3/\rho$.

References

[1] P. Ango Nze: Critères d'ergodicité de modèles markoviens. Estimation non paramétrique sous des hypothèses de dépendance. *Ph. D. Thesis: Université Dauphine, Paris,* 1995.

[2] P. Ango Nze : Critères d'ergodicité géométrique ou arithmétique de modèles linéaires perturbés à représentation markovienne. *Comptes Rendus de l'Académie des Sciences . Paris* **t. 326, Série 1** (1998), 371–376.

[3] P. Ango Nze, P. Bühlman, P. and P. Doukhan : Nonparametric Regression under Non-Mixing Weak Dependence. Technical Report, Department of Statistics, ETH Zürich available under http://www.stat.math.eth.ch/ buhlmann/bibliog.html

[4] P. Bickel and P. Bühlmann : Mixing property and functional central limit theorems for a sieve bootstrap in time series. Technical Report 440, Department of Statistics, University of California Berkeley,(1995).

[5] P. Billingsley : *Convergence of probability measures.* Wiley, New-York, 1968.

[6] P. Breuer and P. Major: Central limit theorems for non-linear functionals of gaussian random fields. *Journal of Multivariate Analysis* **13**(1983), 425–441.

[7] C. Coulon-Prieur and P. Doukhan: A triangular CLT for weakly dependent sequences. *Statistics and Probability Letters* **47** (2000), 61–68.

[8] Y. Davydov: Mixing conditions for Markov chains. *Theory of Probability and Applications* **18** (1973), 313–328.

[9] J. Dedecker and E. Rio: On the functional central limit theorem for stationary processes. *Annales de l'institut Henri Poincaré* **série B, 36** (2000), 1–34.

[10] H. Dehling: *Grenzwertsätze für summen schwach abhändiger vektor-wertiger zufallvariablen.* Doctoral dissertation Göttingen, 1981.

[11] P. Doukhan: *Mixing: properties and examples.* Lecture Notes in Statistics 85, Springer Verlag, 1994.

[12] P. Doukhan: Limit theorems for stationary sequences. Preprint: University of Cergy Pontoise, 1998.

[13] P. Doukhan and S. Louhichi: A new weak dependence condition and application to moment inequalities. *Stochastic Processes and their Applications* **84** (1999), 313–343.

[14] M. Duflo: *Méthodes récursives aléatoires.* Masson, Paris (English Edition, Springer, 1996).

[15] L. Giraitis and D. Surgailis: Multivariate Appell polynomials and the central limit theorem. In: E. Eberlein and M. Taqqu (Eds.) *Dependence in Probability and Statistics.* pp. 27–51, Birkhäuser, Boston, 1986.

[16] H.-C. Ho and T. Hsing: Limit theorems for functionals of moving averages. *Annals of Probability* **25-4** (1997), 1636–1669.

[17] I. A. Ibragimov: Some limit theorems for stationary processes. *Theory of Probability and Applications* **7** (1962), 349–382.

[18] A. N. Kolmogorov and Y. A. Rozanov: On the strong mixing conditions for stationary gaussian sequences. *Theory of Probability and Applications* **5** (1960), 204–207.

[19] E. Liebscher: Strong convergence of sums of α- mixing random variables with applications to density estimation. *Stochastic Processes and their Applications* **65** (1996), 69–98.

[20] S. Louhichi: Weak convergence for empirical processes of associated sequences. Preprint 98-36, University of Paris Sud, Orsay, 1998.

[21] S. P. Meyn and R. L. Tweedie: *Markov chains and stochastic stability.* Communications in control and Engineering, Springer Verlag, 1993.

[22] A. Mokkadem: Propriétés de mélange des modèles autorégressifs polynômiaux. *Annales de l'institut Henri Poincaré* **série B, 26-2** (1990), 219–260.

[23] C. M. Newman: Asymptotic independence and limit theorems for positively and negatively dependent random variables. In: Y. L. Tong (Ed.), *Inequalities in Statistics and Probability*, IMS Lecture Notes-Monograph Series 5, 127–140, 1984.

[24] M. Peligrad: Recent advances in the central limit theorem and its weak invarianace principle for mixing sequences of random variables. In E. Eberlein and M. S. Taqqu (Eds.) *Dependence in Probability and Statistics, a survey of recent results.* Oberwolfach, Birkhäuser, Boston, 1986.

[25] T. D. Pham and L. T. Tran: Some mixing properties of time series models. *Stochastic Processes and their Applications* **19** (1985), 297–303.

[26] E. Rio: *Théorie asymptotique des processus aléatoires faiblement dépendants.* Mathématiques et applications 31, Springer Verlag, 2000.

[27] P. M. Robinson: Nonparametric estimators for time series. *Journal of Time Series Analysis* **4** (1983), 185–207.

[28] M. Rosenblatt: A central limit theorem and a strong mixing condition. *Proceedings of National Academy of Sciences, U.S.A.*, **42** (1956), 43–47.

[29] M. Rosenblatt: *Stationary sequences and random fields.* Birkhäuser, Boston, 1985.

[30] Q. M. Shao: Remark on the invariance principle for ρ-mixing sequences of random variables. *Chinese Annals of Mathematics* **9A:4** (1988), 409–412.

[31] Q. M. Shao: Maximal inequalities for partial sums of ρ-mixing sequences. *Annals of Probability* **23-2** (1995), 948–965.

[32] Q. M. Shao and H. Yu: Weak convergence for weighted empirical processes of dependent sequences. *Annals of Probability* **24-4** (1996), 2052–2078.

[33] H. Tong: A note on a Markov bilinear stochastic process in discrete time. *Journal of Time Series Analysis* **2-4** (1981), 279–284.

[34] A. Yu. Veretennikov: On polynomial mixing and convergence rate for stochastic difference and differential equation. Theory of Probability & Applications, **44-2**, (2000), 361–374.

[35] C. S. Withers: Central limit theorem for dependent variables. *Zeitung Wahrscheinlichkeit und verwandte Gebiete* **57** (1981), 509–534.

[36] V. A. Wolkonski and Y. A. Rozanov: Some limit theorems for random functions, *Part I. Theory of Probability and Applications* **4** (1959), 178–197; *Part II. Theory of Probability and Applications* **6** (1961), 186–198.

[37] H. Yu: A Glivenko-Cantelli Lemma and weak convergence for empirical processes of associated sequences. *Probability Theory and Related Fields* **95** (1993), 357–370.

Weak dependence : models and applications.

Patrick Ango Nze and Paul Doukhan

Abstract We are aimed to develop a systematic introduction to a new weak dependence condition. We show that some popular models hold this property : stationary Markov models, bilinear models, and more generally, Bernoulli shifts. In some cases no mixing properties can be expected without additional regularity assumption of innovation's distribution where as weak dependence conditions can be easily derived. We also develop some of its standard applications. First, probabilistic results : weak invariance principle for Donsker line and empirical process. We also present a statistical application to kernel estimates for density and regression functions.

Résumé Nous développons ici une introduction systematique à une nouvelle condition de dépendance faible. Nous montrons en particulier que quelques modèles très populaires en séries chronologiques vérifient cette condition : des chaînes de Markov stationnaires, des modèles bilinéaires ou plus généralement des schémas de Bernoulli. On ne peut, en général, espérer que ces modèles satisfassent à des conditions de mélange.

Quelques applications classiques sont aussi envisagées. Nous prouvons d'abord des théorèmes limite probabilistes : le principe de Donsker et le théorème de limite centrale empirique. Nous obtenons aussi le comportement asymptotique d'estimations à noyau d'une densité et d'une fonction de régression; leur convergence uniforme presque sûre ainsi que leur comportement asymptotiquement gaussien sont prouvés comme dans le cas de suites indépendantes.

Patrick Ango Nze
Université Lille 3, UFR AES
BP 149,
F-59653 Villeneuve d'Ascq Cedex
France
angonze@univ-lille3.fr
and
Paul Doukhan
UPRESA 8088 CNRS-Mathématiques
Université de Cergy Pontoise
Bât. Les Chênes, 33, Bd. du Port
F-95011 Cergy-Pontoise Cedex and CREST-LS, timbre J340,
3, Av. Pierre Larousse, F-92245 Malakoff
France
Paul.Doukhan@math.u-cergy.fr

1 Introduction

In this paper we are aimed to provide a real alternative to the standard mixing conditions (see e.g. [24]) for modelling some of the stationary sequences commonly used in statistics.

We deal with sequences of random variables. We do not assume independence. Asymptotic results obtained do not need to be analogous to that of independent sequences: this is a first but complicated way to define weak dependence.

We are thus turning to much simpler conditions:

– The first preoccupation we bear in mind is to obtain conditions, simple enough to check on the really used models; our experience (see [11] and [2]) proves the real difficulty to handle strong mixing or absolute regularity conditions.

– The second condition expected from a tractable weak dependence condition is that the planned applications may ensue from such condition.

Our first concern was to establish two probabilistic limit theorems: the Donsker invariance principle and the empirical CLT. They are used in statistical applications: to determine change point estimation in the mean of a stationary sequence or to obtain Kolmogorov-Smirnov tests for distributions. Kernel functional estimation problems provide other statistical applications.

This paper is organized as follows.

A first section is devoted to the definitions our new version of a weak dependence conditions. The main source is the seminal works [13] and [4].

A second section provides the weak dependence properties (in this sense) of some of the classical stationary sequences used in statistics: Markovian models, associated sequences, and Bernoulli shifts. As a by-product, this new dependence condition gives access to some classes of models which could not be studied with former tools (mixing, for instance).

After this we turn to some moment inequalities to be used in the sequel. This is a first application of those weak dependence conditions. Indeed, a moment inequality for sums may be proved by combinatorial arguments. This condition also yields some sub-exponential inequalities described in the application to functional estimation.

Another technical trick is the Lindeberg-Rio (see [26]) method, easily handled in this weak dependence frame. The latter method is compared with Bernstein's blocking technique (see [17]). Each method provides a CLT. The dependence rates for CLT convergence are compared later on in the functional estimation frame. Surprisingly, there is not an uniformly best method.

We then turn to the Donsker invariance principle and the empirical CLT.

Finally we consider functional estimation. Besides the convergence results which are of a proper interest, this frame allows to compare

– the weak dependence conditions with the standard mixing.

– the methods for proving CLTs.

– the obstructions in view of optimal almost sure convergence of kernel estimators (that is the i.i.d. rates).

The number of applications is relatively limited due to the fact that the conditions are quite recent. We hope that this panorama of statistical applications may be enlarged in the years to come.

2 Independence and dependence

2.1 Independence and correlation

Recall that random variables (r.v.), taking values in $\mathbb{R}^d$ are independent in case

$$\forall f, g \ \text{ bounded measurable} : \mathrm{Cov}\,(f(X), g(Y)) = 0.$$

It is enough to consider classes of continuous or more regular functions.

Mixing conditions (recalled hereafter) were introduced by weakening such conditions: the null, right hand side term above is made "small".

Note that in some cases, orthogonality yields independence.

- a) Bernoulli r.v's (that is r.v.'s with two points support.)
- b) Gaussian vectors : $(X, Y) \in \mathbb{R}^{a+b}$.
- c) Associated random vectors.

The first case yields only pairwise independence but not independence, in the case of more than two r.v.'s; it follows from orthogonality of the four couples (X, Y), $(X, 1-Y)$, $(1-X, Y)$ and $(1-X, 1-Y)$ if the couple (X, Y) is orthogonal and the variables (X, Y) are both supported in $\{0, 1\}$.

Except for the first case, properties inherited from covariances are not easy to translate into properties of subjacent sigma-fields.

In the two latter cases b) and c), additional inequalities may be proved for Lipschitz functions. They take the form

$$|\mathrm{Cov}\,(f(X), g(Y))\,| \le c(f, g) \sum_{i,j} |\mathrm{Cov}\,(X_i, Y_j))\,|.$$

2.2 The weak dependence conditions

As the covariances of the initial r.v.'s are much easier to to compute than mixing coefficients, we introduce dependence properties (rather than sigma-fields) for a process $(X_n)_{n\in\mathbb{Z}}$. Set $\mathbb{L}^\infty$ for the set of numerical bounded measurable functions on some space $\mathbb{R}^u$ and $\|\cdot\|_\infty$ the corresponding norm. We define the Lipschitz modulus of a function $h : \mathbb{R}^u \to \mathbb{R}$

$$\mathrm{Lip}(h) = \sup_{x\neq y} \frac{|h(x) - h(y)|}{\|x - y\|_1}, \qquad \text{where } \|(z_1, \ldots, z_u)\|_1 = |z_1| + \cdots + |z_u|.$$

Define

$$\mathcal{L} = \{h \in \mathbb{L}^\infty; \|h\|_\infty \le 1, \mathrm{Lip}(h) < \infty\}\,. \tag{1}$$

Definition 2.1 (Doukhan and Louhichi [13]) *The sequence $(X_n)_{n\in\mathbb{Z}}$ is called $(\theta, \mathcal{L}, \psi)$-weakly dependent if there exists a sequence $\theta = (\theta_r)_{r\in\mathbb{N}}$ decreasing to zero at infinity, and a function ψ with arguments $(h, k, u, v) \in \mathcal{L}^2 \times \mathbb{N}^2$ such that for any u-tuple $(i_1, \ldots, i_u)$ and any v-tuple $(j_1, \ldots, j_v)$ with $i_1 \le \cdots \le i_u < i_u + r \le j_1 \le \cdots \le j_v$, one has*

$$|\mathrm{Cov}(h(X_{i_1}, \ldots, X_{i_u}), k(X_{j_1}, \ldots, X_{j_v}))| \le \psi(h, k, u, v)\theta_r \tag{2}$$

if the functions h and k are defined respectively on $\mathbb{R}^u$ and on $\mathbb{R}^v$.

The examples of interest involve the function $\psi'_1(h,k,u,v) = v\text{Lip}(k)$ (*e.g.* in causal linear processes), $\psi_1(h,k,u,v) = u\text{Lip}(h) + v\text{Lip}(k)$, (*e.g.* in non causal linear processes), $\psi_2(h,k,u,v) = uv\text{Lip}(h)\text{Lip}(k)$ (*e.g.* in associated processes), and $\psi'_2(h,k,u,v) = v\text{Lip}(h)\text{Lip}(k)$.

Notice that the sequence θ depends on both the class $\mathcal{L}$ and the function ψ. The function ψ may really depend on all its arguments, in contrast with the case of bounded mixing sequences. This definition is hereditary through images by convenient functions, as it was noticed by [35] in view to check more easily a weak dependence property.

The following example, due to Rosenblatt ([29]) describes the case where mixing fails to hold. This is the (Markov) $AR(1)$-process with binomial innovations ($\mathbb{P}(\xi_0 = 0) = \mathbb{P}(\xi_0 = 1) = \frac{1}{2}$) :

$$X_n = \frac{1}{2}(X_{n-1} + \xi_n).$$

This is also the Bernoulli shift (see definition below) $X_n = H(\xi_n)$ with $H(x) = \sum_{k=0}^{\infty} 2^{-(k+1)} x_k$. This model has the stationary uniform distribution on the interval but it satisfies no mixing condition (the past may entirely be recovered from present). For a more rigorous proof see [17], page 375.

Define also the class,

$$\mathcal{I} = \left\{ \bigotimes_{i=1}^{u} g_{x_i}; x_i \in \mathbb{R}_+^*, u \in \mathbb{N}^* \right\}, \text{ where } g_x(y) = \mathbb{1}_{x \le y} - \mathbb{1}_{y \le -x}, \forall x \in \mathbb{R}_+^*. \quad (3)$$

The following lemma links $\mathcal{I}$-weak dependence with $\mathcal{L}$-weak dependence. Indeed, examples are proved to satisfy a weak dependence condition w.r.t. the class $\mathcal{L}$. Consider the weaker $\mathcal{L}_0 \cap \mathcal{C}_b^1$-weak dependence condition defined with

$$\mathcal{L}_0 = \left\{ \bigotimes_{i=1}^{u} f_i;\ f_i \in \mathcal{L},\ f_i : \mathbb{R} \to \mathbb{R}, i = 1, \ldots, u,\ u \in \mathbb{N}^* \right\}.$$

Here $\mathcal{C}_b^1$ stands for the set of partially differentiable functions with bounded partial derivatives and the function ψ is either ψ_1 or ψ_2. This latter condition will thus imply $\mathcal{I}$-weak dependence under concentration assumptions.

Lemma 2.2 *Let (X_n) be a sequence of r.v's. Suppose that, for some real $\alpha > 0$*

$$C(\lambda) = \sup_{x \in \mathbb{R}} \sup_i \mathbb{P}\left(x \le X_i \le x + \lambda\right) \le C\lambda^{\alpha}. \quad (4)$$

Suppose that the sequence (X_n) is

- *$(\theta, \mathcal{L}_0 \cap \mathcal{C}_b^1, \psi_1)$-weakly dependent, then it is $(\theta_{\mathcal{I}}, \mathcal{I}, \psi)$-weakly dependent with*

$$\theta_{\mathcal{I},r} = \theta_r^{\frac{\alpha}{1+\alpha}} \text{ and } \psi(h,k,u,v) = 2(8C)^{\frac{1}{1+\alpha}}(u+v).$$

- *$(\theta, \mathcal{L}_0 \cap \mathcal{C}_b^1, \psi_2)$-weakly dependent, then it is $(\theta_{\mathcal{I}}, \mathcal{I}, \psi)$-weakly dependent with*

$$\theta_{\mathcal{I},r} = \theta_r^{\frac{\alpha}{2+\alpha}} \text{ and } \psi(h,k,u,v) = (8C)^{\frac{2}{2+\alpha}}(u+v)^{\frac{2(1+\alpha)}{2+\alpha}}.$$

In order to understand better the mechanism of proofs we recall the proof of this lemma.

Proof of Lemma 2.2. First recall the following useful inequality, valid for real numbers $0 \le x_i, y_i \le 1$, $|x_1 \dots x_m - y_1 \dots y_m| \le \sum_{i=1}^m |x_i - y_i|$. Let $g, f \in \mathcal{I}$, then $g(y_1, \dots, y_u) = g_{x_1}(y_1) \dots g_{x_u}(y_u)$, and $f(y_1, \dots, y_v) = g_{x'_1}(y_1) \dots g_{x'_v}(y_v)$ for some $u, v \in \mathbb{N}^*$ and $x_i, x'_j \ge 0$. For fixed $x > 0$ and $a > 0$ let

$$f_x(y) = \mathbb{1}_{y>x} - \mathbb{1}_{y\le -x} + \left(\frac{y}{a} - \frac{x}{a} + 1\right)\mathbb{1}_{x-a<y<x} + \left(\frac{y}{a} + \frac{x}{a} - 1\right)\mathbb{1}_{-x<y<-x+a}.$$

Therefore $\mathrm{Lip}(f_x) = a^{-1}$ and $\|f_x\|_\infty = 1$ and $\mathrm{Lip}(h) \le a^{-1}$, $\mathrm{Lip}(k) \le a^{-1}$ if we set

$$h(y_1, \dots, y_u) = f_{x_1}(y_1) \dots f_{x_u}(y_u), \qquad k(y_1, \dots, y_v) = f_{x'_1}(y_1) \dots f_{x'_v}(y_v).$$

Consider $i_1 \le \dots \le i_u \le i_u + r \le j_1 \le \dots \le j_v$ and set

$$\mathrm{Cov}(h, k) := \mathrm{Cov}(h(X_{i_1}, \dots, X_{i_u}), k(X_{j_1}, \dots, X_{j_v})).$$

• $(\theta, \mathcal{L}_0 \cap \mathcal{C}_b^1, \psi_1)$ (resp. $(\theta, \mathcal{L}_0 \cap \mathcal{C}_b^1, \psi_2)$)-weak dependence implies, with either $c(u, v) = (u + v)$ or $c(u, v) = (u + v)^2$,

$$|\mathrm{Cov}(h, k)| \le \frac{1}{a} c(u, v)\theta_r \qquad \left(\text{resp. } |\mathrm{Cov}(h, k)| \le \frac{1}{a^2} c(u, v)\theta_r\right).$$

• Inequality (4) yields $|\mathrm{Cov}(g, f) - \mathrm{Cov}(h, k)| \le 8Ca^\alpha(u + v)$. Hence $|\mathrm{Cov}(g, f)| \le 8Ca^\alpha(u + v) + \frac{1}{a}c(u, v)\theta_r$ (resp. $|\mathrm{Cov}(g, f)| \le 8Ca^\alpha(u + v) + \frac{1}{a^2}c(u, v)\theta_r$).
The Lemma follows by setting

$$a = \left[\frac{c(u, v)\theta_r}{8C(u + v)}\right]^{1/(1+\alpha)} \qquad \left(\text{resp. } a = \left[\frac{c(u, v)\theta_r}{8C(u + v)}\right]^{1/(2+\alpha)}\right).$$

2.3 Mixing

For the completeness sake, we recall here the definitions of the main mixing coefficients. For more details, the reader is deferred to Doukhan [11].

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let $\mathcal{U}, \mathcal{V}$ be two sub σ-algebras of $\mathcal{A}$, various measures of dependence between $\mathcal{U}$ and $\mathcal{V}$ have been introduced; among them let us recall

$$\begin{aligned}
\alpha(\mathcal{U}, \mathcal{V}) &= \sup\{|\mathbb{P}(U \cap V) - \mathbb{P}(U)\mathbb{P}(V)|; U \in \mathcal{U}, V \in \mathcal{V}\}, \\
\beta(\mathcal{U}, \mathcal{V}) &= \mathbb{E}\sup\{|\mathbb{P}(V|\mathcal{U}) - \mathbb{P}(V)|; V \in \mathcal{V}\}, \\
\rho(\mathcal{U}, \mathcal{V}) &= \sup\{|\mathrm{Corr}(u, v)|; u \in \mathbb{L}^2(\mathcal{U}), v \in \mathbb{L}^2(\mathcal{V})\}, \\
\phi(\mathcal{U}, \mathcal{V}) &= \sup\{|\mathbb{P}(V|U) - \mathbb{P}(V)|; U \in \mathcal{U}, V \in \mathcal{V}\}.
\end{aligned}$$

Those coefficients are respectively Rosenblatt's [28] strong mixing coefficient, $\alpha(\mathcal{U}, \mathcal{V})$, Wolkonski and Rozanov's absolute regularity coefficient in [36], $\beta(\mathcal{U}, \mathcal{V})$, Kolmogorov and Rozanov' maximal correlation coefficient $\rho(\mathcal{U}, \mathcal{V})$ [18], and $\phi(\mathcal{U}, \mathcal{V})$ the uniform mixing coefficient from Ibragimov [17].
Let $X = (X_n)_{n\in\mathbb{Z}}$ be a discrete time stationary process. We set $X_A = \{X_t; t \in A\}$ if $A \subset \mathbb{Z}$ for the A-marginal of X. At last $\sigma(Z)$ will denote the sigma-algebra generated by a random variable Z.
For any coefficient previously defined, say $c(\cdot, \cdot)$, we shall call the process $\mathbf{X}$ a c-mixing process if $\lim_{k\to\infty} c_{\mathbf{X},k} = 0$ where $c_{\mathbf{X},k} = c(\sigma(X_{]-\infty,0]}), \sigma(X_{[k,+\infty[})$.

$$\phi-mixing \Rightarrow \rho-mixing \Rightarrow \alpha-mixing, \quad \text{and} \quad \phi-\text{mixing} \Rightarrow \beta-\text{mixing} \Rightarrow \alpha-\text{mixing}$$

and no other implication holds.

3 Models

3.1 Markovian Models

Let $(\xi_n)_{n\in\mathbb{Z}}$ be an i.i.d. sequence and let M be some measurable function. We turn our attention to models driven by the equation

$$X_n = M(X_{n-1}, ..., X_{n-p}, \xi_n). \tag{5}$$

In order to justify the title of this section, remark that the vector valued sequence $(X_n^{(p)})_{n\in\mathbb{Z}}$, where $X_n^{(p)} = (X_{n-1}, ..., X_{n-p})$ is Markovian.

Such models are associated with ergodic Markov chains (see [21] for a review of the main models). Thus the stationarity assumption is reachable. An interesting subclass of examples is given by two functions R and S and two mutually independent i.i.d. sequences (ξ_n) and (η_n)

$$X_n = R(X_{n-1}, ..., X_{n-p}, \eta_n) + S(X_{n-1}, ..., X_{n-p})\xi_n.$$

Here the function S satisfies $S(x_1, ..., x_p) \geq s > 0$ for some $s \in \mathbb{R}$, and any real numbers $x_1, ..., x_p$ and the functions R and S essentially satisfy contraction assumptions (see [11], [14], [2] or [34] for developments).

For instance $ARMA(p, q)$ processes

$$Y_n = \sum_{i=1}^{p} a_i Y_{n-i} + \xi_n + \sum_{j=1}^{q} b_j \xi_{n-j},$$

have such a Markov representation. Indeed $X_n = (Y_n, Y_{n,n-1} \ldots, Y_{n,n-\ell})$, where $Y_{n,j} = \mathbb{E}\left[Y_n \mid Y_{n+i} : i \leq n\right]$, and $\ell = \max\{p, q+1\}$, is a Markov process. The mixing properties of these models are developed in [22].

Lipschitzian models are shown to be exponentially $\mathcal{L}$-weakly dependent.

Proposition 3.1 ([14]) . *Let a vector valued Markov model be defined through the recurrence relation (5). Assume that*

$$\mathbb{E}\|M(x, \xi_n) - M(y, \xi_n)\|^S \leq a\|x - y\|^S \quad \textit{and} \quad \mathbb{E}\|M(0, \xi_n)\|^S < \infty,$$

for some $0 \leq a < 1$ *and* $S \geq 1$.
Then the sequence $(X_n)_{n\in\mathbb{Z}}$ *is* $(\theta, \mathcal{L}, \psi)$*-weakly dependent with* $\theta_r = \mathcal{O}(a^r)$ *and* $\psi(h, k, u, v) = v\mathrm{Lip}k$.

Only recall that here stationarity is not required.

More general $AR(p)$ nonlinear models, $X_n = M(X_{n-1}, \ldots, X_{n-p}; \xi_n)$, have the same properties: if, for example, $\mathbb{E}|M(0; \xi_n)| < \infty$ and, for some constants $a_j \geq 0, 1 \leq j \leq p$ with $\sum_{j=1}^{p} a_j < 1$,

$$\mathbb{E}|M(x_1, \ldots, x_p; \xi_n) - M(y_1, \ldots, y_p; \xi_n)| \leq \textstyle\sum_{j=1}^{p} a_j|x_j - y_j|.$$

This model is geometrically $(\theta, \mathcal{L}, \psi)$-weakly dependent with $\psi(h, k, u, v) = \min\{u\mathrm{Lip}h, v\mathrm{Lip}k\}$ as proved in [13].

Remark. Under the assumption that ξ_0's distribution has an almost sure non vanishing density f and is integrable the additive model

$$M(X_{n-1},\dots,X_{n-p},\xi_n)=R(X_{n-1},\dots,X_{n-p})+\xi_n$$

is shown to be ergodic and mixing. More precisely, if the function R is continuous, and $|R(x_1,\dots,x_p)|\le A+a_1|x_1|+\dots+a_p|x_p|$ with $a_1+\dots+a_p<1$, under the invariant initial distribution, the sequence is absolutely regular with $\beta_n=\mathcal{O}(e^{-bn})$ (see [8]).

3.2 Bernoulli shifts

Definition 3.2 *Let $(\xi_i)_{i\in\mathbb{Z}}$ be a sequence of i.i.d. real valued r.v's and the function $H:\mathbb{R}^{\mathbb{Z}}\to\mathbb{R}$ be measurable. The sequence $(X_n)_{n\in\mathbb{Z}}$ is called a Bernoulli shift if it is defined by $X_n=H(\xi_{n-j},j\in\mathbb{Z})$.*

Causal shifts write as $X_n=H(\xi_n,\xi_{n-1},\xi_{n-2},\dots,\xi_0,\xi_{-1},\xi_{-2},\dots)$, i.e. $H:\mathbb{R}^{\mathbb{N}}\to\mathbb{R}$. Such shifts provide natural examples of weakly dependent, but not mixing sequences (see [27]).

Notice that a stationary Markov setting provides such causal sequences. In fact, consider a Markov process driven by the recurrence equation $X_t=M(X_{t-1},\xi_t)$, for some i.i.d. sequence $(\xi_t)_{t\in\mathbb{Z}}$. Then the function H is defined, *if it exists*, implicitly by the relation $H(x)=M(H(x'),x_0)$, where $x=(x_0,x_1,x_2,\dots),x'=(x_1,x_2,x_3,\dots)$. Control theory yields tools to provide explicit Bernoulli shift representations (see [22] and [21]).

3.2.1 Chaotic representations

We now specialize in chaotic expansions associated with the discrete chaos generated by the sequence $(\xi_i)_{i\in\mathbb{Z}}$; in a condensed formulation we write $H(x)=\sum_{k=0}^{\infty}H^{(k)}(x)$, where $H^{(k)}(x)$ denotes the k-th order chaos contribution, $H^{(0)}(x)=a_0^{(0)}$ is only a centering constant, and

$$H^{(k)}(x)=\sum_{j_1=-\infty}^{\infty}\sum_{j_2=-\infty}^{\infty}\cdots\sum_{j_k=-\infty}^{\infty}a^{(k)}_{j_1,\dots,j_k}x_{j_1}\times x_{j_2}\times\dots\times x_{j_k}$$

or in short, in a vectorial notation, $H^{(k)}(x)=\sum_{j\in\mathbb{Z}^k}a_j^{(k)}x_j$.
In contrast with mixing conditions, it may be proved that even non causal sequences may admit a $\mathcal{L}$-weakly dependent behaviour. This is a by-product of the proposition 3.4 to follow.

Definition 3.3 *For any integer $k>0$, we denote δ_k any number such that*

$$\sup_{i\in\mathbb{Z}}\mathbb{E}|H(\xi_{i-j},j\in\mathbb{Z})-H(\xi_{i-j}\mathbb{I}_{|j|<k},j\in\mathbb{Z})|\le\delta_k.$$

Such sequences $(\delta_k)_{k\in\mathbb{Z}^+}$ are related to the modulus of uniform continuity of H. If, for instance, for positive constants $(a_i)_{i\in\mathbb{Z}},0<b\le1$, the inequality $|H(u_i,i\in\mathbb{Z})-H(v_i,i\in\mathbb{Z})|\le\sum_{i\in\mathbb{Z}}a_i|u_i-v_i|^b$ holds for any sequence $(u_i),(v_i)\in\mathbb{R}^{\mathbb{Z}}$ and if the sequence $(\xi_i)_{i\in\mathbb{Z}}$ has a finite b-th order moment, then one can choose $\delta_k=\sum_{|i|\ge k}a_i\mathbb{E}|\xi_i|^b$.

Proposition 3.4 ([13]) *Bernoulli shifts are $(\theta,\mathcal{L},\psi)$-weakly dependent with $\theta_r=2\delta_{r/2}$, and with the function $\psi(h,k,u,v)=4\,(u\|k\|_\infty\mathrm{Lip}(h)+v\|h\|_\infty\mathrm{Lip}(k))$. Under causality, this holds with $\theta_r=\delta_r$ and $\psi(h,k,u,v)=2v\mathrm{Lip}k\|h\|_\infty$.*

Processes associated with a finite number of chaotic terms (i.e. $H^{(k)} = 0$ if $k > k_0$) are also called *Volterra processes.* A suitable bound for δ_r corresponds here with the stationarity condition

$$\delta_r = \sum_{k=0}^{\infty} \left\{ \sum_{j\in\mathbb{Z}^k; \|j\|_\infty > r} \left| a_j^{(k)} \right| \right\} \mathbb{E}|\xi_0|^k < \infty.$$

The first example of such a Volterra process is clearly the class of linear processes $X_t = \sum_{-\infty}^{\infty} a_k \xi_{t-k}$. A suitable sequence is $\theta_r = 2\mathbb{E}|\xi_0| \sum_{|k|>r/2} |a_k|$ with the previous function ψ_1.
The simple bilinear process, $X_t = (a + b\xi_{t-1})X_{t-1} + \xi_t$, is stationary if $c = \mathbb{E}|a + b\xi_0| < 1$ (see [33]). It is a Bernoulli shift with $H(x) = x_0 + \sum_{j=1}^{\infty} x_j \prod_{s=1}^{j} (a + bx_s)$, for $x = (x_i)_{i\in\mathbb{N}}$. We truncate the previous series up to rank r in order to obtain $\delta_r = \theta_r = c^r(r+1)/(1-c)$.
Giraitis, Koul and Surgailis have introduced $ARCH(\infty)-$models (from Giraitis, Koul and Surgailis in [15]. We are given a nonnegative sequence $(b_j)_{j\geq 1}$ and a i.i.d. sequence of nonnegative random variables $(\xi_j)_{j\geq 0}$. The process (if it exists) is ruled through the recurrence relation

$$X_t = \left[a + \sum_{j=1}^{\infty} b_j X_{t-j} \right] \xi_t.$$

Such models are proved to have a stationary representation with the chaotic expansion

$$X_t = a \sum_{\ell=1}^{\infty} \sum_{j_1=0}^{\infty} \cdots \sum_{j_\ell=0}^{\infty} b_{j_1} \cdots b_{j_\ell} \xi_{t-j_1} \cdots \xi_{t-[j_1+\cdots+j_\ell]}$$

under the simple assumption $\mathbb{E}\xi_0 \sum_{j=1}^{\infty} b_j < 1$. They extend the standard ARCH(p) model ($b_j = 0$ if $j > p$). Here we have $\theta_r = \left(\mathbb{E}\xi_0 \sum_{j=1}^{\infty} b_j \right)^r$.

3.2.2 Association

Definition 3.5 *The sequence $(Z_t)_{t\in\mathbb{Z}}$ is associated, if for all coordinatewise increasing real-valued functions h and k,*

$$\mathrm{Cov}(h(Z_t, t \in A), k(Z_t, t \in B)) \geq 0$$

for all finite subsets A and B of $\mathbb{Z}$.

Associated sequences are $(\theta, \psi_2, \mathcal{L})$-weakly dependent with $\theta_r = \sup_{k\geq r} \mathrm{Cov}(X_0, X_r)$ (see [13]). Note that broad classes of examples of associated processes result from the fact that any independent sequence is associated and that monotonicity preserves association (for this, see [23]).

The case of Gaussian sequences is analogous by setting $\theta_r = \sup_{k\geq r} |\mathrm{Cov}(X_0, X_k)|$. Also, one may consider combinations of sums of Gaussian and associated sequences, or Bernoulli shifts driven by stationary, associated, instead of i.i.d. sequences.

Note that for associated or Gaussian sequences, ψ_2' replaces ψ_2 if $\theta_r = \sup_{k\geq r} |\mathrm{Cov}(X_0, X_k)|$ is replaced by $\theta_r = \sum_{k\geq r} |\mathrm{Cov}(X_0, X_k)|$.

Remark. The causal linear process $X_n = \sum_{t=0}^{\infty} a_t \xi_{n-t}$, satisfies a β-mixing condition if ξ_0's density and (for some $\delta > 0$) the $1+\delta$ order moment exist, together with the condition $\sum_{t=-\infty}^{\infty} |a_t|^{\delta} < \infty$. Then [25] prove that $\beta_n \leq C \sum_{l=n}^{\infty} \left(\sum_{k=l}^{\infty} |a_k|\right)^{\delta/1+\delta}$, for some $C > 0$. If, for instance $a_j = \mathcal{O}(j^{-a})$, then under the previous regularity and moment conditions yields if $a > 2 + 1/\delta$, $\beta_n \sim n^{-b}$ with $b = (a-2)\delta/(1+\delta)$. For instance, $\sum_{n=0}^{\infty} \beta_n < \infty$ if $a > 3 + 2/\delta$. If $\delta = 1$ this writes $a > 5$. If $\delta = \infty$ this writes $a > 3$.

4 Algebraic moments of sums (see [13])

Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of centered r.v's and let $S_n = \sum_{i=1}^{n} X_i$. In this section, we give moment bounds for $|\mathbb{E}S_n^q|$, when $q \in \mathbb{N}$ and $q \geq 2$. Let (X_n) be a sequence of centered r.v's and define for positive integer r coefficients of weak dependence as non decreasing sequences $(C_{r,q})_{q\geq 2}$ such that

$$\left|\mathrm{Cov}\left(X_{t_1} \times \cdots \times X_{t_m}, X_{t_{m+1}} \times \cdots \times X_{t_q}\right)\right| \leq C_{r,q}, \tag{6}$$

for all the successions $\{t_1, \ldots, t_q\}$ such that $1 \leq t_1 \leq \ldots \leq t_q \leq n$ and, for some integer $1 \leq m < q$, $t_{m+1} - t_m = r$. Explicit bounds $C_{r,q}$ are provided by [13] in order to build inequalities for the partial sums S_n. Two kinds of bounds are considered, either

$$\left|\mathrm{Cov}\left(X_{t_1} \times \cdots \times X_{t_m}, X_{t_{m+1}} \times \cdots \times X_{t_q}\right)\right| \leq c q^{\gamma} M^{q-2} \theta_r \tag{7}$$

or,

$$\left|\mathrm{Cov}\left(X_{t_1} \times \cdots \times X_{t_m}, X_{t_{m+1}} \times \cdots \times X_{t_q}\right)\right| \leq c \int_0^{\theta_r} Q_{X_{t_1}}(x) \times \cdots \times Q_{X_{t_q}}(x) dx, \tag{8}$$

where Q_X still denotes the X's quantile function and $c, \gamma \geq 0$ denote real numbers.
In the examples, the previous bound (7) holds for bounded sequences such that $\|X_n\|_\infty \leq M$. Under $(\theta, \mathcal{L}, \psi)$-weak dependence, this yields the bound

$$C_{r,q} = \max_{1 \leq m < q} \psi(j^{\otimes m}, j^{\otimes(q-m)}, m, q-m) M^q \theta_r$$

where $j(x) = x\mathbb{1}_{|x|\leq 1} + \mathbb{1}_{x>1} - \mathbb{1}_{x<-1}$.
As in lemma 2.2, we see that under $(\theta, \mathcal{L}, \psi)$-weak dependence with $\psi(h,k,u,v) = c(u,v)\mathrm{Lip}(h)\mathrm{Lip}(k)$ a bound is

$$C_{r,q} = \max\nolimits_{1\leq m<q} c(m, q-m) M^{q-2} \theta_r.$$

For non bounded random variables,

$$C_{r,q} = c \int_0^{\theta_r} Q_{X_{t_1}}(x) \times \cdots \times Q_{X_{t_q}}(x) dx$$

under $(\theta, \mathcal{I}, \psi)$-weak dependence.
An analogous bound is obtained by Rio for strongly mixing sequences (see [26]). The bound (8) holds for more general r.v's, using moment or tail assumptions.
A first consequence of inequality (7) is the following Marcinkiewicz-Zygmund inequality.
Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of centered r.v's satisfying the condition

$$C_{r,q} = \mathcal{O}(r^{-q/2}), \tag{9}$$

then, there is a constant $B > 0$ not depending on n for which

$$|\mathbb{E}S_n^q| \le Bn^{q/2}. \tag{10}$$

The following lemma gives moment inequalities whenever $q \in \{2, 4\}$, it was essentially proved in Billingsley ([5], lemmas 3 and 4, page 172).

Lemma 4.1 ([13]) *If $(X_n)_{n\in\mathbb{N}}$ is a sequence of centered r.v's, then*

$$\mathbb{E}S_n^2 \le 2n\sum_{r=0}^{n-1} C_{r,2}, \quad \mathbb{E}S_n^4 \le 4!\left\{\left(n\sum_{r=0}^{n-1} C_{r,2}\right)^2 + n\sum_{r=0}^{n-1}(r+1)^2 C_{r,4}\right\}. \tag{11}$$

The following Theorems deal with higher order moments.

Theorem 4.2 ([13]) *Let $q \ge 2$ be some integer. Suppose that dependence coefficients $C_{r,p}$ associated to the sequence (X_n) satisfy for all integers $0 < p \le q$ and for some positive constants M, γ, C*

$$C_{r,p} = Ce^{\gamma p}M^{p-2}\theta_r, \tag{$\mathcal{H}_1$}$$

then, for any integer $n \ge 2$

$$|\mathbb{E}S_n^q| \le \frac{(2q-2)!}{(q-1)!}e^{q\gamma}\left\{\left(Cn\sum_{r=0}^{n-1}\theta_r\right)^{q/2} \vee \left(CM^{q-2}n\sum_{r=0}^{n-1}(r+1)^{q-2}\theta_r\right)\right\}. \tag{12}$$

Theorem 4.2 is adapted to work with bounded sequences. Define the class $\mathcal{I}$

$$\mathcal{I} = \left\{\bigotimes_{i=1}^{u} g_{x_i}; x_i \in \mathbb{R}_+^*, u \in \mathbb{N}^*\right\}, \text{ where } g_x(y) = \mathbb{1}_{x\le y} - \mathbb{1}_{y \le -x}, \forall x \in \mathbb{R}_+^*. \tag{13}$$

In order to consider the unbounded case, we shall consider $(\theta, \mathcal{I}, \psi)$-weak dependence where ψ writes $\psi(h, k, u, v) = c(u, v)$.

Theorem 4.3 ([13]) *If $(X_n)_{n\in\mathbb{N}}$ is a centered and $(\theta, \mathcal{I}, \psi)$-weakly dependent sequence, set $C_q = (\max_{u+v\le q} c(u,v)) \vee 2$. Then*

$$|\mathbb{E}S_n^q| \le \frac{(2q-2)!}{(q-1)!}\left(C_q\sum_{i=1}^{n}\int_0^1(\theta^{-1}(u)\wedge n)^{q-1}Q_i^q(u)du \vee \left(C_2\sum_{i=1}^{n}\int_0^1(\theta^{-1}(u)\wedge n)Q_i^2(u)du\right)^{q/2}\right).$$

In the special case of strongly mixing and stationary sequences, this is Theorem 1 in [26]. The restriction of working with even integer exponents finds its compensation in the explicit form of the constants.

Remark. Exponential inequalities may be built by using Theorem 4.2. Define

$$M_{q,n} = n\sum_{r=0}^{n-1}(r+1)^{q-2}C_{r,q}.$$

We first suppose that for all integers $q \ge 2$ and n:

$$C_{r,p} = Ce^{\gamma p}M^{p-2}\theta_r, \quad M_{q,n} \le A_n\frac{q!}{\beta^q},$$

where β is some constant and A_n is a sequence independent of q, then

$$\forall x \in \mathbb{R}_+^*, \qquad \mathbb{P}\left(|S_n| \geq x\sqrt{A_n}\right) \leq A\exp\left(-B\sqrt{\beta x}\right), \qquad (DP)$$

with $A = e^{4+1/12}\sqrt{8\pi}$, and $B = e^{5/2-\gamma/2}$. This exponential inequality is analogous to ??? Let (X_n) be a sequence of centered r.v.'s, this holds if $C_{r,q} = C\sigma^2 M^{q-2} e^{\gamma q} e^{-br}$ for constants $C, \sigma, \gamma, b > 0$, if $\|X_n\|_\infty \leq M$ and $\|X_n\|_2 \leq \sigma$, for any integer $n \geq 0$. In this case $A_n = n\sigma^2$. E.g. this holds under $(\theta, \mathcal{L}, \psi)$-weak dependence if $\theta_r = \mathcal{O}(e^{-br})$ and $\psi(h,k,u,v) \leq e^{\delta(u+v)}\mathrm{Lip}(h)\mathrm{Lip}(k))$ for some $\delta \geq 0$.

The use of combinatorics in those inequalities makes them relatively weak. E.g. Bernstein inequality, valid for independent sequences allows to replace the term $\sqrt{x}$ in the previous inequality by x^2 under the same assumption $n\sigma^2 \geq 1$; in mixing cases analogue inequalities are also obtained by using coupling arguments which are not available here.

5 Limit theorems

5.1 The Donsker line

Consider a stationary sequence $(X_n)_{n\in\mathbb{Z}}$. We assume that this sequence is integrable, centered at expectation, and that

$$\mathbb{E}X_0 = 0.$$

Denote by $[x]$ the integral part of a real number x ($[x] \leq x < [x]+1$), the Donsker line $(D_n(t))_{t\in[0,1]}$ is defined for any sample with size n as the following continuous time process

$$D_n(t) = \sum_{k=1}^{[nt]} X_k + (nt - [nt])X_{[nt]+1}.$$

Let $(W_t)_{t\in[0,1]}$ be a Brownian motion, that is W denotes the centered Gaussian real valued process with covariance

$$\mathbb{E}W_s W_t = \min\{s,t\}.$$

We consider the following convergence result in the space $\mathcal{C}([0,1])$ of continuous functions on the unit interval when the sample size n grows to infinity:

Theorem 5.1 *Suppose that the stationary sequence $(X_n)_{n\in\mathbb{Z}}$ satisfies $\mathbb{E}|X_0|^{4+\delta} < \infty$ for some $\delta > 0$. Assume a $(\theta, \mathcal{I}, \psi)$-weak dependence condition with the function $\psi_1(h,k,u,v)$ (respectively ψ_2) and $\theta_r = \mathcal{O}(r^{-a})$ with $a \geq 2 + 4/\delta$ (respectively $a > 2$).*
Then the following functional convergence holds in the space $\mathcal{C}([0,1])$:

$$\frac{1}{\sqrt{n}} D_n(t) \to \sigma W_t.$$

The following series is assumed to be convergent

$$\sigma^2 = \sum_{-\infty}^{\infty} \mathrm{Cov}(X_0, X_k).$$

The case $\sigma^2 = 0$ is detailed by [24]. We shall assume here that $\sigma^2 \neq 0$.

Remark. Without any regularity condition on innovations, theorem 5.1 holds for a bounded Lipschitz function of a linear process $X_n = \sum_{k=-\infty}^{\infty} a_k \xi_{n-k}$ if $a_k = \mathcal{O}(k^{-D})$ when $D > 3$. The latter doesn't need to be causal while Ho and Hsing need this causality assumption in [16].

Proof of Theorem 5.1. Lemma 4.2 and a maximal inequality by Moricz *et alii* (see [11], page 40) yield

$$\mathbb{E}|S_n|^{2+\delta} = \mathcal{O}\left(n^{1+\delta/2}\right) \tag{14}$$

as soon as for any increasing sequence of integers $0 \leq i < j < k \leq l$

$$\sum_{m=0}^{\infty} m\mathbb{E}|X_0 X_m| < \infty \text{ and } \mathrm{Cov}\,(X_i X_j, X_k X_l) = \mathcal{O}\left((k-j)^{-2}\right). \tag{15}$$

Moreover, this entails that $\sigma^2 = n^{-1}\mathrm{Var}(S_n) > 0$ and the finite dimensional (fidi) convergence follows. The tightness of the process is a consequence of (15). The first part of (15) follows from the covariance bound $|\mathrm{Cov}\,(X_0, X_r)\,| =) = \mathcal{O}\left(\theta_r^{(2+\delta)/(4+\delta)}\right)$. The latter bound follows from $\mathrm{Cov}\,(X_i X_j, X_k X_l) = \mathcal{O}\left(\theta_{k-j}^{\delta/(4+\delta)}\right)$.

Remark. Let $\delta > 0$. Assume the existence of a moment of order $(2+\delta)$ for X_0. The Donsker functional CLT holds in the strong mixing case if $\sum_{n=0}^{\infty} n^{2/\delta}\alpha_n < \infty$ (see [26]). The condition $\sum_{n=0}^{\infty} \rho(2^n) < \infty$ with $\mathbb{E}X_0^2 < \infty$ implies the functional convergence, as Shao proved [30].

Remark. Here we consider conditions in terms of conditional expectations with respect to an adapted filtration. We first recall that theorem 5.1 holds for martingales with stationary square integrable increments $\mathbb{E}X_0^2 < \infty$ (see [5]). Let $(\mathcal{M}_n)$ be a filtration adapted to the process $(X_n)_{n\in\mathbb{Z}}$: X_n is $\mathcal{M}_n$-measurable for any $n \in \mathbb{Z}$. Dedecker and Rio [9] prove that for a centered, square integrable process, such that the sum $\sum_{n=0}^{\infty} X_0\, \mathbb{E}\,(X_n|\mathcal{M}_0)$ is convergent in $\mathbb{L}^1$, the sequence $\mathbb{E}(X_0^2 + 2X_0 S_n|\mathcal{I})$ converges in $\mathbb{L}^1$ to some random variable η. The σ-field $\mathcal{I}$ is the tail σ-field. Moreover, conditionally to $\mathcal{I}$, $D_n(t)/\sqrt{n}$ converge to a Brownian motion ηW_t.

- This result provides a functional CLT to a limit process which is not Gaussian (see [6] for results related to the latter case).
- Note that ergodicity implies the triviality of the tail σ-field and a standard Donsker theorem follows.
- Standard results prove this theorem under a more restrictive $\mathbb{L}^2$ assumption: both series

$$\sum_{n=0}^{\infty} \mathbb{E}(X_n|\mathcal{M}_0) \text{ and } \sum_{n=0}^{\infty} (X_n - \mathbb{E}(X_n|\mathcal{M}_0))$$

  converge.
- As a new result yielded by this theorem, consider a stationary Markov sequence $(\xi_n)_{n\in\mathbb{Z}}$ with stationary distribution μ and transition operator P. Let $X_n = g(\xi_n)$ be centered at expectation, nonlinear functionals of (ξ_n). Then the assumption writes as the convergence of the series $\sum_{n=0}^{\infty} gP^n g$ in $\mathbb{L}^1(\mu)$.

- The previous result concerning strongly mixing sequences appears also as a consequence of this theorem.

5.2 Empirical process

Let us consider a stationary sequence $(X_n)_{n\in\mathbb{Z}}$. We assume without loss of generality that the marginal distribution of this sequence is the uniform law on $[0,1]$. The empirical repartition process of the sequence (X_n) at time n is defined as $\frac{1}{\sqrt{n}}E_n(x)$ where

$$E_n(x) = \sum_{k=1}^n \left(\mathbb{1}_{(X_k\le x)} - \mathbb{P}(X_k \le x)\right).$$

Note that $E_n = n\,(F_n - F)$ if F_n, F respectively denote the empirical d.f. and the marginal d.f. We consider the following convergence result in the Skohorod space $\mathcal{D}(\mathbb{R})$ when the sample size n converges to infinity:

$$\frac{1}{\sqrt{n}}E_n(x) \to \bar{B}(x).$$

Here $(\bar{B}(x))_{x\in\mathbb{R}}$ is the dependent analogue of a Brownian bridge, that is $\bar{B}$ denotes the centered Gaussian process with covariance given by

$$\mathbb{E}\bar{B}(x)\bar{B}(y) = \sum_{k=-\infty}^{\infty} \left(\mathbb{P}(X_0 \le x,\ X_k \le y) - \mathbb{P}(X_0 \le x)\mathbb{P}(X_k \le y)\right). \tag{16}$$

Note that for independent sequences with a marginal repartition function F, this only writes $\bar{B}(x) = B(F(x))$ for some standard Brownian Bridge B; this justifies the name of Generalized Brownian Bridge.

Let (X_n) be a stationary sequence assumed to satisfy the following weak dependence condition.

$$\sup_{f\in\mathcal{F}} \left|\mathrm{Cov}\left(\prod_{i=1}^{2} f(X_{t_i}), \prod_{i=3}^{4} f(X_{t_i})\right)\right| \le \theta_r, \tag{17}$$

where $\mathcal{F} = \{x \to \mathbb{1}_{s<x\le t},\ \text{for } s,t \in [0,1]\}$, $0 \le t_1 \le t_2 \le t_3 \le t_4$ and $r = t_3 - t_2$ (in this case a weak dependence condition holds for a class of functions $\mathbb{R}^u \to \mathbb{R}$ working only with the values $u = 1$ or 2).

Proposition 5.2 *Let (X_n) be a stationary sequence such that (17) holds. Assume that there exists $\nu > 0$ such that*

$$\theta_r = \mathcal{O}(r^{-5/2-\nu}). \tag{18}$$

Then the sequence of processes $\left(\frac{1}{\sqrt{n}}\{E_n(t); t\in[0,1]\}\right)_{n>0}$ is tight in the Skohorod space $\mathcal{D}([0,1])$.

Theorem 5.3 *Suppose that the stationary sequence (X_n) is $(\theta, \mathcal{L}_1, \psi_j)$-weakly dependent, with either $j=1$ and $\theta_r = \mathcal{O}(r^{-15/2-\nu})$, or $j=2$ and $\theta_r = \mathcal{O}(r^{-5-\nu})$. Then the following empirical functional convergence holds true in the Skohorod space of numerical càdlàg functions on the real line, $D(\mathbb{R})$:*

$$\frac{1}{\sqrt{n}}E_n(x) \to \bar{B}(x).$$

If the sequence (X_n) is i.i.d., then equation (16) reduces to zero-index term. This covariance has essentially $1/2$ order Holder regularity. If the second order marginals (X_0, X_k) have a joint, continuous density, then the $k-$order term is a C^2-function. Hence the regularity of the covariance is driven by the central term. In any case the process $\bar{B}$ is perhaps not derivable in quadratic mean.

Proof of Theorem 5.3. Thanks to the Rosenthal inequality in Theorem 4.2

$$\begin{aligned}\|E_n(t)-E_n(s)\|_4 &\leq C\sqrt{n}\sum_{r=0}^{n-1}\min\left\{r^{-5/2-\nu},|t-s|\right\}+\left(n\sum_{r=0}^{n-1}(r+1)^2\theta_r\right)^{1/4}\\ &\leq \sqrt{n}\left(|t-s|^{(a-1)/2a}+n^{(2-a)/4}\right).\end{aligned}$$

The conclusion follows from Shao and Yu' tightness lemma [32]. The fidi convergence is due to a CLT Lemma by blocks by Ibragimov [17].

Remark. If the sequence is strongly mixing, the summability condition $\sum_{n=0}^{\infty}\alpha_n<\infty$ implies fidi convergence. The empirical functional convergence holds if, moreover, for some $a>1$, $\alpha_n=\mathcal{O}(n^{-a})$ (see [26]). In a absolutely regular framework, Doukhan, Massart and Rio (see [26]) obtain the empirical functional convergence when, for some $a>2$, $\beta_n=\mathcal{O}(n^{-1}(\log n)^{-a})$. Shao and Yu [32] obtain the empirical functional convergence theorem when the maximal correlation coefficient satisfy the condition $\sum_{n=0}^{\infty}\rho(2^n)<\infty$.

5.3 A triangular CLT: Lindeberg-Rio's method

Let $X_{n,j}, 1\leq j\leq n$ be a triangular array. We shall omit the first index n when possible without confusion. Consider a sequence $(W_j)_{j\in\mathbb{N}}$ of i.i.d. r.v.'s with standard normal law, and independent from $(X_n)_{t\in\mathbb{Z}}$. Define

$$\begin{aligned}\sigma_n^2 &= \text{Var }(S_n)\\ S_k &= \sum_{j=1}^{k}X_j,\ 1\leq k\leq n,\ \text{and } S_0=0,\\ \tau_k &= \sum_{j=k}^{n}\sqrt{v_{n,j}}W_j=\sum_{j=k}^{n}V_j,\ 1\leq k\leq n,\ \text{and } \tau_{n+1}=0,\end{aligned}$$

where we assume that $v_{n,j}=\text{Var }(S_j)-\text{Var }(S_{j-1})>0$.

Once one has proved that $\sigma_n^2\to\sigma^2$, it remains to prove that for any three times differentiable function with bounded derivatives up to order 3, φ say,

$$\Delta_n(\varphi)=\mathbb{E}\varphi(S_n)-\mathbb{E}\varphi(W_0)\to 0. \tag{19}$$

Consider the following:

$$U_j=S_{j-1}+\tau_{j+1},\ R_j(x)=\varphi(U_j+x)-\varphi(U_j)-\frac{v_j}{2}\varphi''(U_j)\ \ (1\leq j\leq n).$$

Clearly, we have

$$\Delta_n(\varphi) = \sum_{k=1}^{n}\Delta_{n,k}(\varphi)$$

$$\begin{aligned}\Delta_{n,k}(\varphi) &= \mathbb{E}R_k(X_k) - \mathbb{E}R_k(V_k)\\ &= \Delta_k^1(\varphi) - \Delta_k^2(\varphi).\end{aligned}$$

A Taylor expansion yields

$$\begin{aligned}\Delta_k^2(\varphi) &= \mathbb{E}\left(\varphi\left(U_k+V_k\right)-\varphi\left(U_k\right)-V_k\varphi'\left(U_k\right)-\frac{V_k^2}{2}\varphi''\left(U_k\right)\right)\\ &= \frac{1}{6}\mathbb{E}\left(V_k^3\varphi^{(3)}\left(U_k+\vartheta_kV_k\right)\right),\ \text{with } 0<\vartheta_k<1,\\ \left|\Delta_k^2(\varphi)\right| &\le C\left(v_k\right)^{3/2}.\end{aligned}$$

Moreover,

$$\begin{aligned}\Delta_k^1(\varphi) &= \mathbb{E}\left(\varphi\left(U_k+X_k\right)-\varphi\left(U_k\right)-\frac{v_k}{2}\varphi''\left(U_k\right)\right)\\ &= \mathbb{E}\left(X_k\varphi'\left(U_k\right)+\frac{1}{2}\left(X_k^2-\frac{v_k}{2}\right)\varphi''\left(U_k\right)+\frac{1}{6}X_k^3\varphi^{(3)}\left(U_k+\vartheta_kX_k\right)\right),\ \text{with } 0<\vartheta_k<1.\end{aligned}$$

It then follows that

$$\begin{aligned}\sum_{k=1}^n\Delta_k^1(\varphi) &= \sum_{k=1}^n\sum_{j=1}^{k-1}\mathrm{Cov}\left(\varphi''\left(S_{k-1-j}+\tau_{j+1}\right)X_{k-j},X_k\right)\\ &+ \frac{1}{2}\sum_{k=1}^n\sum_{j=1}^{k-1}\mathrm{Cov}\left(\varphi^{(3)}\left(S_{k-1-j}+\tau_{j+1}+\vartheta_{k-j}X_{k-j}\right)X_{k-j}^2,X_k\right)\\ &+ \sum_{k=1}^n\mathrm{Cov}\left(X_k,\varphi'\left(\tau_{k+1}\right)\right)+\frac{1}{2}\sum_{k=1}^n\mathbb{E}\left(\varphi''\left(U_k\right)\left(X_k^2-\mathbb{E}X_k^2\right)\right)\\ &+ \mathbb{E}\left(\sum_{k=1}^{n-1}\mathrm{Cov}\left(X_0,X_k\right)\sum_{j=k+1}^n\varphi''\left(U_j\right)\right)+\frac{1}{6}\sum_{k=1}^n\mathbb{E}\left(\varphi^{(3)}\left(U_k+\iota_kX_k\right)X_k^3\right)\\ &= E_1+E_2+E_3+E_4+E_5+E_6. \qquad (20)\end{aligned}$$

The term E_3 is null. The other ones are controlled in two ways: by uniform bounds on one hand, by dependence covariance bounds on the other hand. See [7] for further details.

We can apply this method to density estimation on $\mathbb{R}$. Consider a stationary sequence $(X_t)_{t\in\mathbb{Z}}$ with marginal density f. Assume that the densities of the pairs (X_0, X_k), $k\in\mathbb{Z}^+$, exist, and are uniformly bounded : $\sup_{k>0}\|f_{(k)}\|_\infty<\infty$.

Let K be some kernel function with integral 1, Lipschitzian and compactly supported. The kernel density estimator is defined by

$$\hat f(x) = \hat f_{n,h}(x) = \frac{1}{nh}\sum_{t=1}^n K\left(\frac{x-X_t}{h}\right).$$

Theorem 5.4 *Suppose that the stationary sequence $(X_t)_{t\in\mathbb{Z}}$ is $(\theta,\mathcal{L}_1,\psi_j')$-weakly dependent, with either $j=1,2$ and $\theta_r=\mathcal{O}(r^{-a}), a>2+j$, or $(\theta,\mathcal{L}_1,\psi_j)$-weakly dependent $j=1,2$ and $\theta_r=\mathcal{O}(r^{-a}), a>2+j+1/\delta$, where $n^\delta h\to\infty$ for some $\delta\in]0,1[$. If $f(x)>0$, then*

$$\sqrt{nh}\ \left(\hat f(x)-\mathbb{E}\hat f(x)\right)\xrightarrow{\mathcal{D}}\mathcal{N}\left(0,f(x)\int K^2(u)du\right)$$

Remark. In order to obtain a CLT for time series, it seems that the first paper using the Bernstein blocking technique in due to Dehling, see [10]. This method is also described in [11] and it is used in [13]. The technique consists in dividing a sample $\{Z_1, \ldots, Z_n\}$ into blocks $U_m = \{Z_k\}_{k \in K_m}$ which are almost independent and such that $\mathrm{Card}\,(\{1, \ldots, n\} \setminus \bigcup_m K_m) = o(n)$ as $n \to \infty$. This condition essentially allows to work with the variables $T_m = \sum_{k \in K_m} Z_k$ (built from those blocks) exactly as in the classical Lindeberg method of the independent case.

Proof of Theorem 5.4. The details of the proof follow [3].

6 Functional estimation

We consider a stationary processes $(Z_t)_{t \in \mathbb{Z}}$ with $Z_t = (X_t, Y_t)$ where $X_t,\ Y_t \in \mathbb{R}$. The quantity of interest is the regression function $r(x) = \mathbb{E}(Y_0 | X_0 = x)$. Let K be some kernel function with integral 1. Assume that K is a Lipschitz function with a compact support. The kernel estimator is defined by

$$\hat{f}(x) = \hat{f}_{n,h}(x) = \frac{1}{nh} \sum_{t=1}^{n} K\left(\frac{x - X_t}{h}\right), \qquad \hat{g}(x) = \hat{g}_{n,h}(x) = \frac{1}{nh} \sum_{t=1}^{n} Y_t K\left(\frac{x - X_t}{h}\right),$$

$$\hat{r}(x) = \hat{r}_{n,h}(x) = \frac{\hat{g}_{n,h}(x)}{\hat{f}_{n,h}(x)} \quad \text{if} \quad \hat{f}_{n,h}(x) \neq 0; \qquad \hat{r}(x) = 0 \text{ otherwise.}$$

Here $h = (h_n)_{n \in \mathbb{N}}$ is a sequence of positive real numbers. We always assume that $h_n \to 0,\ nh_n \to \infty$ as $n \to \infty$.

Definition 6.1 *Let $\rho = a + b$ with $(a, b) \in \mathbb{N} \times]0; 1]$. Define the set of ρ-regular functions $\mathcal{C}_\rho$ by*

$$\mathcal{C}_\rho = \Big\{ u : \mathbb{R} \to \mathbb{R};\ u \in \mathcal{C}_a \ \text{ and } \exists A \geq 0\ |u^{(a)}(x) - u^{(a)}(y)| \leq A |x - y|^b \\ \text{for all } x, y \text{ in any compact subset} \Big\}.$$

Here, $\mathcal{C}_a$ is the set of a-times continuously differentiable functions.

If $g \in \mathcal{C}_\rho$, one can choose a kernel function K of order ρ such that uniformly on any compact subset of $\mathbb{R}$

$$b_h(x) = \mathbb{E}(\hat{g}(x)) - g(x) = \mathcal{O}(h^\rho)$$

uniformly on any compact subset of $\mathbb{R}$. In view of asymptotic analysis we assume that the marginal density $f(.)$ of X_0 exists and is continuous. Moreover $f(x) > 0$ for any point x of interest and the regression function $r(.) = \mathbb{E}(Y_0 | X_0 = .)$ exists, is continuous. At last, for some $p \geq 1$, $x \to g_p(x) = f(x)\mathbb{E}(|Y_0|^p | X_0 = x)$ exists and is continuous. We set $g = fr$ with obvious short notation. Moreover, assume one of the following moment assumptions. Either,

$$\mathbb{E}|Y_0|^S < \infty, \qquad \text{for some} \qquad S \geq p \tag{21}$$

or

$$\mathbb{E} \exp(|Y_0|) < \infty. \tag{22}$$

6.1 Second Order Properties

We consider first the properties of $\hat{g}(x)$. We also consider the following conditionally centered equivalent of g_2 appearing in the asymptotic variance of the estimator $\hat{r}$,

$$G_2(x) = f(x)\text{Var }(Y_0|X_0 = x) = g_2(x) - f(x)r^2(x).$$

Assume that the densities of the pairs (X_0, X_k), $k \in \mathbb{Z}^+$, exist, and are uniformly bounded : $\sup_{k>0} \|f_{(k)}\|_\infty < \infty$. Moreover, uniformly over all $k \in \mathbb{Z}^+$, the functions

$$r_{(k)}(x, x') = \mathbb{E}\left(|Y_0 Y_k|\Big|X_0 = x, X_k = x'\right) \tag{23}$$

are continuous. Under these assumptions, the functions $g_{(k)} = f_{(k)}r_{(k)}$ are locally bounded.

Theorem 6.2 *Suppose that the stationary sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta, \psi_j, \mathcal{L})$-weakly dependent with $\theta_r = \mathcal{O}(r^{-a})$ and $a > 2 + j$, for $j = 1$ or $j = 2$. Assume that it satisfies the conditions (22) and (23) with $p = 2$. Suppose that $n^\delta h \to \infty$ for some $\delta \in]0, 1[$. Then uniformly in x belonging to any compact subset of $\mathbb{R}$,*

$$\text{Var }(\hat{g}(x)) = \frac{1}{nh}\; g_2(x) \int K^2(u)du \;+\; o\left(\frac{1}{nh}\right)$$

and

$$\text{Var }\left(\hat{g}(x) - r(x)\hat{f}(x)\right) = \frac{1}{nh}G_2(x) \int K^2(u)du + o\left(\frac{1}{nh}\right).$$

6.2 Central limit theorems

We first consider the density estimator.

Theorem 6.3 *Suppose that the stationary sequence $(X_t)_{t\in\mathbb{Z}}$ is either $(\theta, \mathcal{L}_1, \psi_j')$-weakly dependent, with $\theta_r = \mathcal{O}(r^{-a}), a > 2 + j$, or $(\theta, \mathcal{L}_1, \psi_1)$-weakly dependent with $\theta_r = \mathcal{O}(r^{-a}), a > 3(1 + \sqrt{5})/2$, or $(\theta, \mathcal{L}_1, \psi_2)$-weakly dependent with $\theta_r = \mathcal{O}(r^{-a}), a > 6$. If where $nh \to \infty$ and $f(x) > 0$, then*

$$\sqrt{nh}\;\left(\hat{f}(x) - \mathbb{E}\hat{f}(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, f(x)\int K^2(u)du\right).$$

Proof of Theorem 6.3. The CLT is obtained by Lindeberg Rio method here above described and by Bernstein's blocking technique described in [17]. The Theorem follows by comparing the rates obtained by both methods. See [3] for more details.

Theorem 6.4 *Assume that the stationary sequence $(Z_t)_{t\in\mathbb{Z}}$ is $(\theta, \psi_j, \mathcal{L})$-weakly dependent with $\theta_r = \mathcal{O}(r^{-a})$ with $a > \min\left(\max(2 + j, 3(2 + j)\delta), \max\left(2 + j + \frac{1}{\delta}, \frac{2+2(2+j)\delta}{1+\delta}\right)\right)$, for $j = 1$ or $j = 2$. Assume moreover that it satisfies the conditions (22) and (23) with $p = 2$. Consider a positive kernel K. Let $f, g \in \mathcal{C}_\rho$ for some $\rho \in]0, 2]$, and $nh^{1+2\rho} \to 0$. Then, for all x belonging to any compact subset of $\mathbb{R}$,*

$$\sqrt{nh}\;\left(\hat{r}(x) - r(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\; \frac{G_2(x)}{f^2(x)}\int K^2(u)du\right).$$

Proof of Theorem 6.4. See [3]. They also prove the following

Proposition 6.5 *Assume that the stationary sequence* $(Z_t)_{t\in\mathbb{Z}}$ *is* $(\theta, \psi_j, \mathcal{L})$*-weakly dependent with* $\theta_r = \mathcal{O}(r^{-a})$ *with* $a > 9$ *for* $j = 1$ *and* $a > 12$ *for* $j = 2$*. Suppose that the stationary sequence* $(Z_t)_{t\in\mathbb{Z}}$ *satisfies the conditions (22) and (23) with* $p = 2$*. Consider a positive kernel* K*. Let* $f, g \in \mathcal{C}_\rho$ *for some* $\rho \in]0, 2]$*, and* $nh \to \infty$*. Then, for all* x *belonging to any compact subset of* $\mathbb{R}$*,*

$$\sqrt{nh}\,\left(\hat{r}(x) - \mathbb{E}\hat{r}(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\ \frac{G_2(x)}{f^2(x)} \int K^2(u)du\right).$$

Remark. The CLT convergence theorem 6.4 for the regression function holds true for strongly mixing sequences, under the moment assumption (21) if $h_n \to 0$, $nh_n/\log(n) \to \infty$, $\alpha_n = \mathcal{O}(n^{-a})$ with $\alpha > 2S/(S-2)$ ([27]). No positivity assumption on the kernel K is required.

6.3 Almost sure convergence

For the sake of simplicity, we only consider the geometrically dependent case.

Theorem 6.6 *Let* $(Z_t)_{t\in\mathbb{Z}}$ *be a stationary sequence satisfying the conditions (22), (23) with* $p = 2$*, and that it is either* $(\theta, \psi_1, \mathcal{L})$*- or* $(\theta, \psi_2, \mathcal{L})$*-weakly dependent with* $\theta_r \le a^r$ *for some* $0 < a < 1$*.*

(i) *If* $nh/\log^4(n) \to \infty$*, then for any* $M > 0$*, almost surely,*

$$\sup_{|x|\le M} |\hat{g}(x) - \mathbb{E}\hat{g}(x)| = O\left(\frac{\log^2(n)}{\sqrt{nh}}\right).$$

(ii) *For any* $M > 0$*, if* $f, g \in \mathcal{C}_\rho$ *for some* $\rho \in]0, \infty[$*,* $h \sim \left(\frac{\log^4(n)}{n}\right)^{1/(1+2\rho)}$ *and* $\inf_{|x|\le M} f(x) > 0$*, then, almost surely,*

$$\sup_{|x|\le M} |\hat{r}(x) - r(x)| = O\left\{\left(\frac{\log^4(n)}{n}\right)^{\rho/(1+2\rho)}\right\}.$$

Remark. Liebscher [19] proves the uniform almost sure convergence in a strongly mixing framework, at the optimal rate $\mathcal{O}\left(\left(\frac{\log(n)}{n}\right)^{\rho/(1+2\rho)}\right)$, if $\alpha_r = \mathcal{O}(r^{-a})$, with $a > 4 + 3/\rho$.

References

[1] P. Ango Nze: Critères d'ergodicité de modèles markoviens. Estimation non paramétrique sous des hypothèses de dépendance. *Ph. D. Thesis: Université Dauphine, Paris,* 1995.

[2] P. Ango Nze : Critères d'ergodicité géométrique ou arithmétique de modèles linéaires perturbés à représentation markovienne. *Comptes Rendus de l'Académie des Sciences . Paris* **t. 326, Série 1** (1998), 371–376.

[3] P. Ango Nze, P. Bühlman, P. and P. Doukhan : Nonparametric Regression under Non-Mixing Weak Dependence. Technical Report, Department of Statistics, ETH Zürich available under http://www.stat.math.eth.ch/ buhlmann/bibliog.html

[4] P. Bickel and P. Bühlmann : Mixing property and functional central limit theorems for a sieve bootstrap in time series. Technical Report 440, Department of Statistics, University of California Berkeley,(1995).

[5] P. Billingsley : *Convergence of probability measures.* Wiley, New-York, 1968.

[6] P. Breuer and P. Major: Central limit theorems for non-linear functionals of gaussian random fields. *Journal of Multivariate Analysis* **13**(1983), 425–441.

[7] C. Coulon-Prieur and P. Doukhan: A triangular CLT for weakly dependent sequences. *Statistics and Probability Letters* **47** (2000), 61–68.

[8] Y. Davydov: Mixing conditions for Markov chains. *Theory of Probability and Applications* **18** (1973), 313–328.

[9] J. Dedecker and E. Rio: On the functional central limit theorem for stationary processes. *Annales de l'institut Henri Poincaré* **série B, 36** (2000), 1–34.

[10] H. Dehling: *Grenzwertsätze für summen schwach abhändiger vektor-wertiger zufallvariablen.* Doctoral dissertation Göttingen, 1981.

[11] P. Doukhan: *Mixing: properties and examples.* Lecture Notes in Statistics 85, Springer Verlag, 1994.

[12] P. Doukhan: Limit theorems for stationary sequences. Preprint: University of Cergy Pontoise, 1998.

[13] P. Doukhan and S. Louhichi: A new weak dependence condition and application to moment inequalities. *Stochastic Processes and their Applications* **84** (1999), 313–343.

[14] M. Duflo: *Méthodes récursives aléatoires.* Masson, Paris (English Edition, Springer, 1996).

[15] L. Giraitis and D. Surgailis: Multivariate Appell polynomials and the central limit theorem. In: E. Eberlein and M. Taqqu (Eds.) *Dependence in Probability and Statistics.* pp. 27–51, Birkhäuser, Boston, 1986.

[16] H.-C. Ho and T. Hsing: Limit theorems for functionals of moving averages. *Annals of Probability* **25-4** (1997), 1636–1669.

[17] I. A. Ibragimov: Some limit theorems for stationary processes. *Theory of Probability and Applications* **7** (1962), 349–382.

[18] A. N. Kolmogorov and Y. A. Rozanov: On the strong mixing conditions for stationary gaussian sequences. *Theory of Probability and Applications* **5** (1960), 204–207.

[19] E. Liebscher: Strong convergence of sums of α- mixing random variables with applications to density estimation. *Stochastic Processes and their Applications* **65** (1996), 69–98.

[20] S. Louhichi: Weak convergence for empirical processes of associated sequences. Preprint 98-36, University of Paris Sud, Orsay, 1998.

[21] S. P. Meyn and R. L. Tweedie: *Markov chains and stochastic stability.* Communications in control and Engineering, Springer Verlag, 1993.

[22] A. Mokkadem: Propriétés de mélange des modèles autorégressifs polynômiaux. *Annales de l'institut Henri Poincaré* **série B, 26-2** (1990), 219–260.

[23] C. M. Newman: Asymptotic independence and limit theorems for positively and negatively dependent random variables. In: Y. L. Tong (Ed.), *Inequalities in Statistics and Probability*, IMS Lecture Notes-Monograph Series 5, 127–140, 1984.

[24] M. Peligrad: Recent advances in the central limit theorem and its weak invarianace principle for mixing sequences of random variables. In E. Eberlein and M. S. Taqqu (Eds.) *Dependence in Probability and Statistics, a survey of recent results.* Oberwolfach, Birkhäuser, Boston, 1986.

[25] T. D. Pham and L. T. Tran: Some mixing properties of time series models. *Stochastic Processes and their Applications* **19** (1985), 297–303.

[26] E. Rio: *Théorie asymptotique des processus aléatoires faiblement dépendants.* Mathématiques et applications 31, Springer Verlag, 2000.

[27] P. M. Robinson: Nonparametric estimators for time series. *Journal of Time Series Analysis* **4** (1983), 185–207.

[28] M. Rosenblatt: A central limit theorem and a strong mixing condition. *Proceedings of National Academy of Sciences, U.S.A.*, **42** (1956), 43–47.

[29] M. Rosenblatt: *Stationary sequences and random fields.* Birkhäuser, Boston, 1985.

[30] Q. M. Shao: Remark on the invariance principle for ρ-mixing sequences of random variables. *Chinese Annals of Mathematics* **9A:4** (1988), 409–412.

[31] Q. M. Shao: Maximal inequalities for partial sums of ρ-mixing sequences. *Annals of Probability* **23-2** (1995), 948–965.

[32] Q. M. Shao and H. Yu: Weak convergence for weighted empirical processes of dependent sequences. *Annals of Probability* **24-4** (1996), 2052–2078.

[33] H. Tong: A note on a Markov bilinear stochastic process in discrete time. *Journal of Time Series Analysis* **2-4** (1981), 279–284.

[34] A. Yu. Veretennikov: On polynomial mixing and convergence rate for stochastic difference and differential equation. Theory of Probability & Applications, **44-2**, (2000), 361–374.

[35] C. S. Withers: Central limit theorem for dependent variables. *Zeitung Wahrscheinlichkeit und verwandte Gebiete* **57** (1981), 509–534.

[36] V. A. Wolkonski and Y. A. Rozanov: Some limit theorems for random functions, *Part I. Theory of Probability and Applications* **4** (1959), 178–197; *Part II. Theory of Probability and Applications* **6** (1961), 186–198.

[37] H. Yu: A Glivenko-Cantelli Lemma and weak convergence for empirical processes of associated sequences. *Probability Theory and Related Fields* **95** (1993), 357–370.