Probab. Theory Relat. Fields 132, 203–236 (2005)
Digital Object Identifier (DOI) 10.1007/s00440-004-0394-3

Jérôme Dedecker · Clémentine Prieur

New dependence coefficients. Examples and applications to statistics

Received: 23 June 2004 / Revised version: 23 September 2004 /
Published online: 27 December 2004 – © Springer-Verlag 2004

Abstract. To measure the dependence between a real-valued random variable X and a σ-algebra $\mathcal{M}$, we consider four distances between the conditional distribution function of X given $\mathcal{M}$ and the distribution function of X. The coefficients obtained are weaker than the corresponding mixing coefficients and may be computed in many situations. In particular, we show that they are well adapted to functions of mixing sequences, iterated random functions and dynamical systems. Starting from a new covariance inequality, we study the mean integrated square error for estimating the unknown marginal density of a stationary sequence. We obtain optimal rates for kernel estimators as well as projection estimators on a well localized basis, under a minimal condition on the coefficients. Using recent results, we show that our coefficients may be also used to obtain various exponential inequalities, a concentration inequality for Lipschitz functions, and a Berry-Esseen type inequality.

1. Introduction and definitions

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X a real-valued random variable with law $\mathbb{P}_X$ and $\mathcal{M}$ a σ-algebra of $\mathcal{A}$. Recall that there exists a function $\mathbb{P}_{X|\mathcal{M}}$ from $\mathcal{B}(\mathbb{R}) \times \Omega$ to $[0, 1]$ such that

1. For any ω in Ω, $\mathbb{P}_{X|\mathcal{M}}(., \omega)$ is a probability measure on $\mathcal{B}(\mathbb{R})$.
2. For any $A \in \mathcal{B}(\mathbb{R})$, $\mathbb{P}_{X|\mathcal{M}}(A, .)$ is a version of $\mathbb{E}(\mathbb{1}_{X\in A}|\mathcal{M})$.

The usual mixing coefficients between $\mathcal{M}$ and $\sigma(X)$, introduced respectively by Rosenblatt (1956), Volkonskii and Rozanov (1959) and Ibragimov (1962), may be defined as follows (see for instance Bradley (2002), Proposition 3.22):

$$\alpha(\mathcal{M}, \sigma(X)) = \sup_{A\in\mathcal{B}(\mathbb{R})} \|\mathbb{P}_{X|\mathcal{M}}(A) - \mathbb{P}_X(A)\|_1$$
$$\beta(\mathcal{M}, \sigma(X)) = \| \sup_{A\in\mathcal{B}(\mathbb{R})} |\mathbb{P}_{X|\mathcal{M}}(A) - \mathbb{P}_X(A)| \,\|_1$$

J. Dedecker: Laboratoire de Statistique Théorique et Appliquée, Université Paris 6, 175 rue du Chevaleret, 75013 Paris, France. e-mail: dedecker@ccr.jussieu.fr

C. Prieur: Laboratoire de Statistique et Probabilités, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse cedex 4, France. e-mail: prieur@cict.fr

Mathematics Subject Classifications (2000): 62G07, 60J10, 60E15, 37C30

Key words or phrases: Dependence coefficients – Mixing – Covariance inequalities – Markov chains – Dynamical systems – Density estimation – Exponential inequalities – Concentration inequalities – Berry-Esseen inequality

$$\phi(\mathcal{M}, \sigma(X)) = \sup_{A \in \mathcal{B}(\mathbb{R})} \|\mathbb{P}_{X|\mathcal{M}}(A) - \mathbb{P}_X(A)\|_\infty .$$

Note that $\alpha(\mathcal{M}, \sigma(X)) = 2 \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|, A \in \mathcal{M}, B \in \sigma(X)\}$, so that our definition differs from that of Rosenblatt (1956) from a factor 2. These coefficients measure the dependence between $\mathcal{M}$ and $\sigma(X)$, and are widely used in the areas of limit theorems and statistics. Due to their importance, the properties of these coefficients have been extensively studied by many authors. For recent and complete works, we mention the monographs by Doukhan (1994), Rio (2000a) and Bradley (2002). One of the most important examples is the following: a stationary, irreducible, aperiodic and positively recurrent Markov chain $(X_i)_{i \geq 0}$ is β-mixing, which means that $\beta(\sigma(X_0), \sigma(X_n))$ tends to zero as n tends to infinity (for more details, see Rio (2000a), inequality (9.22) page 139).

Unfortunately, many simple Markov chains are neither β nor α-mixing. For instance, Andrews (1984) proved that if $(\epsilon_i)_{i \geq 1}$ is iid with marginal $\mathcal{B}(1/2)$, then the stationary solution $(X_i)_{i \geq 0}$ of the equation

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

is not α-mixing (more precisely $\alpha(\sigma(X_0), \sigma(X_n)) = 1/2$ for any n). This example is not an exception: the chain satisfying (1.1) is the Markov chain associated to the dynamical system generated by the map $T(x) = 2x \bmod 1$ on the space $[0, 1]$ equipped with the Lebesgue measure (see Section (4.4) for more details), and it is well known that such dynamical sytems are not α-mixing in the sense that $\alpha(\sigma(T), \sigma(T^n))$ does not tend to zero as n tends to infinity. More precisely, let T be a Borel function preserving a probability μ on $\mathcal{B}(\mathbb{R})$. The sequence $(T^i)_{i \geq 0}$ of random variables from $(\Omega, \mathcal{A}, \mathbb{P}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ to $\mathbb{R}$ is strictly stationary. Since $\sigma(T^n) \subset \sigma(T)$ and since T^n has distribution μ, it follows that $\alpha(\sigma(T), \sigma(T^n)) \geq \alpha(\sigma(T^n), \sigma(T^n)) = \alpha(\mathcal{B}(\mathbb{R}), \mathcal{B}(\mathbb{R}))$, and the later is positive as soon as the probability μ is non trivial. Note that the dynamical system (T^n, μ) is said to be mixing in the ergodic-theoric sense (MES) if for any sets A and B in $\mathcal{B}(\mathbb{R})$, the sequence $D_n(A, B, \mu, T) = |\mu(A \cap T^{-n}(B)) - \mu(A)\mu(B)|$ converges to zero. For such dynamical systems, it is easy to see that strong mixing is a uniform version of MES, since with our definition $\alpha(\sigma(T), \sigma(T^n)) = 2 \sup\{D_{n-1}(A, B, \mu, T), A, B \in \mathcal{B}([0, 1])\}$. Mixing in the ergodic-theoretic sense is an important property which is satisfied for many ergodic dynamical systems. However, since it only gives a non uniform control of $D_n(A, B, \mu, T)$, it is not sufficient in general to obtain functional limit theorems or deviation inequalities for large classes of functions.

Although many dependent processes are not mixing, some of them can be represented as functions of mixing processes, that is $X_n = f((\xi_{n+i})_{i \in \mathbb{Z}})$ where f is a function from $\mathcal{X}^{\mathbb{Z}}$ to $\mathbb{R}$ and $(\xi_i)_{i \in \mathbb{Z}}$ is a mixing sequence. If f is not too bad, this structure of dependence is often sufficient to derive limit theorems for the sequence $(X_i)_{i \in \mathbb{Z}}$. Since the well known results of Billingsley (1968, Section 21), who used this representation to establish limit theorems for the continued-fraction transformation, this approach has proved to be very fruitful. In 1982 Hofbauer and

Keller proved that if T is a nice expanding map preserving a probability μ on [0, 1], with finite partition $\{I_1, \ldots, I_N\}$ of [0, 1] into intervals of continuity and monoticity of T, the label process defined by $\xi_n(x) = i$ if $T^n(x) \in I_i$ is β-mixing with exponential mixing rate, and $T^n = f((\xi_i)_{i \geq n})$ for some measurable f. Using this representation together with a strong invariance principle for functions of β-mixing sequences given in Philipp and Stout (1975), Hofbauer and Keller proved a strong invariance principle for the partial sums $S_n(f) = f \circ T + \cdots + f \circ T^n$, where f is any bounded variation function. Functions of β-mixing processes have been further studied in a recent paper by Borovkova *et al.* (2001), who provided many interesting examples and applications, and showed how the coupling properties of the underlying sequence may be used in such situations. We shall follow a similar approach for the examples of Section 4.1 (see also Rio (1996), Section 1.2 for related results).

Note that, even if one knows that a stationary sequence can be written as a function f of a mixing sequence, one may know nothing about the function f and its properties. This is the case, for instance, in the paper by Hofbauer and Keller (1982), where only the existence of f is proved. This theoretical representation is not sufficient to obtain uniform upper bounds for $|\mathbb{E}(g(X_k)|\mathcal{M}) - \mathbb{E}(g(X_k))|$ over an appropriate class of functions $\mathcal{G}$, which are useful to prove limit theorems for the empirical process indexed by a subset of $\mathcal{G}$ (see Corollary 4, Section 6) as well as deviation inequalities for some functions of the variables (see inequality (1.2) and Theorem 2, Section 7.4). In part (ii) of the proof of Theorem 5 in Hofbauer and Keller, such upper bounds are derived from the properties of the adjoint operator of T, and not from the representation $T^n = f((\xi_i)_{i \geq n})$. Now, as one can see from Theorem 4.4 in Bradley (2002), the control of the conditional expectations $|\mathbb{E}(g(X_k)|\mathcal{M}) - \mathbb{E}(g(X_k))|$ over a class of functions $\mathcal{G}$ is often related to the control of an appropriate dependence coefficient (see also Lemma 1, Section 1.1).

A reasonable question is then: how to weaken the definition of the usual mixing coefficients in order to catch many more examples, without losing too much of their nice properties? A first idea, given by Rosenblatt, is to consider coarser sets than $\mathcal{M}$ or $\mathcal{B}(\mathbb{R})$. In fact changing $\mathcal{M}$ is possible, but the coefficients obtained behave differently from the usual mixing coefficients (see for instance Doukhan and Louhichi (1999)). Another way is to change $\mathcal{B}(\mathbb{R})$ by considering only the coarser set $\{]-\infty, t], t \in \mathbb{R}\}$, as done in Rio (2000a) and Peligrad (2002) for the strong mixing coefficient. The coefficients obtained measure the difference between the conditional distribution function $F_{X|\mathcal{M}}$ of $\mathbb{P}_{X|\mathcal{M}}$ and the distribution function F_X of $\mathbb{P}_X$. More precisely, define the four dependence coefficients

$$\begin{aligned}
\tau(\mathcal{M}, X) &= \int \|F_{X|\mathcal{M}}(t) - F_X(t)\|_1 dt \\
\alpha(\mathcal{M}, X) &= \sup_{t \in \mathbb{R}} \|F_{X|\mathcal{M}}(t) - F_X(t)\|_1 \\
\beta(\mathcal{M}, X) &= \| \sup_{t \in \mathbb{R}} |F_{X|\mathcal{M}}(t) - F_X(t)| \, \|_1 \\
\phi(\mathcal{M}, X) &= \sup_{t \in \mathbb{R}} \|F_{X|\mathcal{M}}(t) - F_X(t)\|_\infty .
\end{aligned}$$

The coefficient $\alpha(\mathcal{M}, X)$ was introduced by Rio (2000a, equation 1.10c) and used by Peligrad (2002), while $\tau(\mathcal{M}, X)$ was introduced by Dedecker and Prieur (2003).

Of course, the coefficients $\alpha(\mathcal{M}, X)$, $\beta(\mathcal{M}, X)$ and $\phi(\mathcal{M}, X)$ are smaller than the corresponding mixing coefficients $\alpha(\mathcal{M}, \sigma(X))$, $\beta(\mathcal{M}, \sigma(X))$ and $\phi(\mathcal{M}, \sigma(X))$. We shall see in Section 4 that these weaker coefficients may be easily computed in many situations, so that our first objective is reached. For instance, if T is a nice piecewise expanding map preserving a probability μ on $[0, 1]$, then $\phi(\sigma(T^n), T)$ decreases geometrically (this works for $T(x) = 2x \bmod 1$, and hence for the model (1.1) the coefficient $\phi(\sigma(X_0), X_n)$ decreases geometrically). The largest classes of examples are obtained for the coefficient τ, which is the easiest to compute.

Among the coefficients described above, some of them have a nice interpretation in terms of coupling. Let us first recall the well known result of Berbee (1979): if Ω is rich enough, there exists a random variable X^* distributed as X and independent of $\mathcal{M}$ such that $\mathbb{P}(X \neq X^*) = \beta(\mathcal{M}, \sigma(X))$. For the mixing coefficient $\alpha(\mathcal{M}, \sigma(X))$, Bradley (1983) proved the following result: if Ω is rich enough, then for each $1 \leq p \leq \infty$ and each $\lambda < \|X\|_p$, there exists X^* distributed as X and independent of $\mathcal{M}$ such that $\mathbb{P}(|X - X^*| \geq \lambda) \leq 18(\|X\|_p/\lambda)^{p/(2p+1)}(\alpha(\mathcal{M}, \sigma(X)))^{2p/(2p+1)}$. For the weaker coefficient $\alpha(\mathcal{M}, X)$, Rio (1995, 2000a) obtained the following upper bound, which is not directly comparable to Bradley's: if X belongs to $[a, b]$ and if Ω is rich enough, there exists X^* independent of $\mathcal{M}$ and distributed as X such that $\|X - X^*\|_1 \leq (b - a)\alpha(\mathcal{M}, X)$. Rio's coupling has been extended by Peligrad (2002) to the case of unbounded variables. Many authors have used these coupling properties to obtain sharp limit theorems as well as sharp exponential bounds (see Merlevède and Peligrad (2002) and the references therein).

Note that the random variable X^* appearing in the results by Rio (1995, 2000a) and Peligrad (2002) is based on Major's quantile transformation (1978). It has the following remarkable property: $\|X - X^*\|_1$ is the infimum of $\|X - Y\|_1$ where Y is independent of $\mathcal{M}$ and distributed as X. Starting from the exact expression of X^*, Dedecker and Prieur (2003) have shown that $\tau(\mathcal{M}, X)$ is the appropriate coefficient for the coupling in $\mathbb{L}^1$: the equality $\|X - X^*\|_1 = \tau(\mathcal{M}, X)$ holds. This property is a useful tool to obtain suitable inequalities and to prove various limit theorems (see Section 7.3). When F_X is regular, it can be used also to obtain upper bounds for $\beta(\mathcal{M}, X)$ (see Proposition 2, Section 3).

We see that both $\beta(\mathcal{M}, \sigma(X))$ and $\tau(\mathcal{M}, X)$ have a property of optimality: they are equal to the infimum of $\mathbb{E}(d_0(X, Y))$ where Y is independent of $\mathcal{M}$ and distributed as X, for the distances $d_0(x, y) = \mathbb{1}_{x \neq y}$ and $d_0(x, y) = |x - y|$ respectively. In fact, these two coefficients belong to the same family, built on the Kantorovitch-Rubinstein distance $K_{d_0}(\mathbb{P}_{X|\mathcal{M}}, \mathbb{P}_X)$ between the probabilities $\mathbb{P}_{X|\mathcal{M}}$ and $\mathbb{P}_X$. We shall be more precise on this subject in Section 7.1.

As made clear by Viennet (1997) in a β-mixing framework, a precise covariance inequality is another useful tool for statistical applications. Using a covariance inequality due to Delyon (1990) Viennet proved that, under a minimal assumption on the β-mixing coefficients, the mean integrated square error (MISE) for the unknown invariant density is of the same order than in the iid case. This result applies to kernel estimators as well as projection estimators. In Proposition 1 of Section 2, we prove an inequality similar to that of Delyon but for $\beta(\mathcal{M}, X)$ instead of $\beta(\mathcal{M}, \sigma(X))$. The main difference is that our inequality is no longer symetric, so that it cannot apply to any projection estimators (see Proposition 4, Section 5.1). Nevertheless, for kernel estimators as well as projection estimators on well localized basis (such as histograms and wavelet basis), we extend Viennet's results to sequences such that $\sum \beta(\sigma(X_0), X_n)$ is finite (see Sections 5.3 and 5.4). Once again the results apply to dynamical systems.

In Proposition 5 of Section 6 we prove an Hoeffding-type inequality for partial sums $S_n(h) = h(X_1) + \cdots + h(X_n)$, where h is a bounded variation function (see Section 1.1 for a Definition). If $\phi(k) = \sup_{i \leq n-k} \phi(\sigma(X_j, 1 \leq j \leq i), X_{k+i})$, the bound is

$$\mathbb{P}(|S_n(h) - \mathbb{E}(S_n(h))| > x) \leq C_1 \exp\left(\frac{-x^2}{nC_2\|dh\|^2(\phi(0) + \cdots + \phi(n-1))}\right), \tag{1.2}$$

for some universal constants C_1 and C_2 (see Proposition 5 for the exact expression). As a byproduct, we obtain an empirical central limit theorem for a class of smooth functions.

To obtain more precise inequalities and limit theorems, it is often necessary to consider the dependence between a past σ-algebra and several points in the future of the sequence. Unfortunately the coefficients we use seem difficult to define in higher dimension, because they are based on distribution functions. Starting from an equivalent definition given in Lemma 1, we see that the difficulty vanishes for $\tau(\mathcal{M}, X)$. The definition of that coefficient can be naturally extended to random variables with values in any Polish space $\mathcal{X}$, without losing the coupling property (see Section 7.1). Following Rio (1996), we can also define the uniform version $\varphi(\mathcal{M}, X)$ of $\tau(\mathcal{M}, X)$.

We shall see in Section 7 that the coefficients $\tau(\mathcal{M}, (X_i, \ldots, X_{i+m}))$ and their uniform version $\varphi(\mathcal{M}, (X_i, \ldots, X_{i+m}))$ are still easy to compute for the examples given in Section 4, and that their asymptotic behavior is the same as when considering only a single point in the future (this is mainly due to the underlying Markov structure of these examples). Then, using recent results of Rio (1996, 2000b), Collet *et al.* (2002) and Dedecker and Prieur (2003), we obtain a Berry-Esseen bound, a concentration inequality for Lipschitz functions, and a functional law of the iterated logarithm for partial sums.

1.1. Equivalent definitions

Definition 1. *A σ-finite signed measure is the difference of two positive σ-finite measures, one of them at least being finite. We say that a function h from $\mathbb{R}$ to $\mathbb{R}$ is σ-BV if there exists a σ-finite signed measure dh such that $h(x) = h(0) + dh([0, x[)$ if $x \geq 0$ and $h(x) = h(0) - dh([x, 0[)$ if $x \leq 0$ (h is left continuous). The function h is BV if the signed measure dh is finite. Recall also the Hahn-Jordan decomposition: for any σ-finite signed measure μ, there is a set D such that $\mu_+(A) = \mu(A \cap D) \geq 0$ and $-\mu_-(A) = \mu(A \backslash D) \leq 0$. μ_+ and μ_- are singular, one of them at least is* *finite and* $\mu = \mu_+ - \mu_-$. *The measure* $|\mu| = \mu_+ + \mu_-$ *is called the total variation measure for* μ. *Denote by* $\|\mu\| = |\mu|(\mathbb{R})$.

As for other measures of dependence, we can define $\tau(\mathcal{M}, X)$, $\alpha(\mathcal{M}, X)$, $\beta(\mathcal{M}, X)$ and $\phi(\mathcal{M}, X)$ as a supremum over some family of functions (compare to Theorem 4.4 in Bradley (2002) for usual mixing coefficients).

Lemma 1. *Let* $(\Omega, \mathcal{A}, \mathbb{P})$ *be a probability space,* X *a real-valued random variable and* $\mathcal{M}$ *a* σ*-algebra of* $\mathcal{A}$. *Let* Λ_1 *be the space of* 1*-Lipschitz functions from* $\mathbb{R}$ *to* $\mathbb{R}$, *and* BV_1 *be the space of BV functions* h *such that* $\|dh\| \le 1$. *We have*

1. $\tau(\mathcal{M}, X) = \Big\|\sup\Big\{\Big|\int f(x)\mathbb{P}_{X|\mathcal{M}}(dx) - \int f(x)\mathbb{P}_X(dx)\Big|, f \in \Lambda_1\Big\}\Big\|_1$.
2. $\alpha(\mathcal{M}, X) = \sup\{\|\mathbb{E}(f(X)|\mathcal{M}) - \mathbb{E}(f(X))\|_1, f \in BV_1\}$.
3. $\beta(\mathcal{M}, X) = \Big\|\sup\Big\{\Big|\int f(x)\mathbb{P}_{X|\mathcal{M}}(dx) - \int f(x)\mathbb{P}_X(dx)\Big|, f \in BV_1\Big\}\Big\|_1$.
4. $\phi(\mathcal{M}, X) = \sup\{\|\mathbb{E}(f(X)|\mathcal{M}) - \mathbb{E}(f(X))\|_\infty, f \in BV_1\}$.

Proof. In Dedecker and Prieur (2003), the equality in 1 is given as a definition of $\tau(\mathcal{M}, X)$. The fact that the right hand side in 1 is equal to $\int \|F_{X|\mathcal{M}}(t) - F_X(t)\|_1 dt$ follows from the equalities (2.8), (2.9) and (2.10) of the proof of Lemma 5 in Dedecker and Prieur (2003).

It remains to prove 2, 3 and 4. Without loss of generality, assume that f in BV_1 is such that $f(-\infty) = 0$. Hence,

$$\begin{aligned}&\int f(x)\mathbb{P}_{X|\mathcal{M}}(dx) - \int f(x)\mathbb{P}_X(dx)\\ &\quad= \int\Big(\int \mathbb{1}_{x>t}\, df(t)\Big)\mathbb{P}_{X|\mathcal{M}}(dx) - \int\Big(\int \mathbb{1}_{x>t}\, df(t)\Big)\mathbb{P}_X(dx).\end{aligned}$$

Applying Fubini, we obtain that

$$\begin{aligned}\Big|\int f(x)\mathbb{P}_{X|\mathcal{M}}(dx) - \int f(x)\mathbb{P}(dx)\Big| &= \Big|\int \big(F_{X|\mathcal{M}}(t) - F_X(t)\big)\, df(t)\Big|\\ &\le \int |F_{X|\mathcal{M}}(t) - F_X(t)||df|(t)\,. \quad (1.3)\end{aligned}$$

From (1.3) we easily infer that

$$\begin{aligned}\sup\{\|\mathbb{E}(f(X)|\mathcal{M}) - \mathbb{E}(f(X))\|_1, f \in BV_1\} &\le \alpha(\mathcal{M}, X)\\ \Big\|\sup\Big|\int f(x)\mathbb{P}_{X|\mathcal{M}}(dx) - \int f(x)\mathbb{P}(dx)\Big|, f \in BV_1\Big\|_1 &\le \beta(\mathcal{M}, X)\\ \sup\{\|\mathbb{E}(f(X)|\mathcal{M}) - \mathbb{E}(f(X))\|_\infty, f \in BV_1\} &\le \phi(\mathcal{M}, X)\,,\end{aligned}$$

and the converse inequalities follow by noting that the function $\mathbb{1}_{]-\infty,t]}$ belongs to BV_1. □

2. Covariance Inequalities

Proposition 1. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let X and Y be two real-valued random variables and h be a σ-BV function. If Y, $h(X)$ and $Yh(X)$ are integrable, then*

$$\mathrm{Cov}(Y, h(X)) = -\int \mathrm{Cov}(Y, \mathbb{1}_{X\le t})dh(t)\,. \tag{2.1}$$

Let $\mathcal{M}$ be a σ-algebra of $\mathcal{A}$, and $b(\mathcal{M}, X) = \sup_{t\in\mathbb{R}} |F_{X|\mathcal{M}}(t) - F_X(t)|$. If Y is $\mathcal{M}$-measurable, we have the inequalities

1. $|\mathrm{Cov}(Y, h(X))| \le \|Y\|_\infty\Big(\int \|F_{X|\mathcal{M}}(t) - F_X(t)\|_1 |dh|(t)\Big)$.
2. $|\mathrm{Cov}(Y, h(X))| \le \|dh\|\,\mathbb{E}(|Y|b(\mathcal{M}, X)) \le \|dh\|\,\|Y\|_1\,\phi(\mathcal{M}, X)$.

Remark 1. The first inequality in item 2 is comparable to that of Delyon (1990) (see also Viennet (1997), Lemma 4.1) in which appear two variables $b_1(\mathcal{M}, X)$ and $b_2(\mathcal{M}, X)$ each having mean $\beta(\mathcal{M}, \sigma(X))$. The main difference is that our inequality is not symetric, because the coefficient $\beta(\mathcal{M}, X)$ is not.

Proof. We proceed as in Theorem 2.3 in Yu (1993). Let X^* be a random variable distributed as X and independent of Y. We have the equalities

$$\mathrm{Cov}(Y, h(X)) = \mathbb{E}(Y(h(X) - h(X^*))) = \mathbb{E}\Big(Y\int (\mathbb{1}_{X^*\le t} - \mathbb{1}_{X\le t})\,dh(t)\Big)\,. \tag{2.2}$$

To apply Fubini, it is sufficient to check that

$$\mathbb{E}\Big(|Y|\int |\mathbb{1}_{X^*\le t} - \mathbb{1}_{X\le t}|\,|dh|(t)\Big) < \infty\,. \tag{2.3}$$

Define the function $\overline{h}$ by $\overline{h}(x) = |dh|([0, x[)$ if $x \ge 0$ and $\overline{h}(x) = |dh|([x, 0[)$ if $x \le 0$. With this definition, we have that

$$\begin{aligned}\int |\mathbb{1}_{X^*\le t} - \mathbb{1}_{X\le t}|\,|dh|(t) &\le \int \big(|\mathbb{1}_{X^*\le t} - \mathbb{1}_{0\le t}| + |\mathbb{1}_{X\le t} - \mathbb{1}_{0\le t}|\big)\,|dh|(t)\\ &= \overline{h}(X^*) + \overline{h}(X)\,,\end{aligned} \tag{2.4}$$

Now $dh = \mu_+ - \mu_-$ where μ_- for instance is finite. Define the two functions G_+ and G_- by $G_+(x) = \mu_+([0, x[)$ and $G_-(x) = \mu_-([0, x[)$ if $x \ge 0$ and $G_+(x) = -\mu_+([x, 0[)$ and $G_-(x) = -\mu_-([x, 0[)$ if $x \le 0$. Clearly $h(x)-h(0) = G_+(x) - G_-(x)$. Since $Yh(X)$ is integrable and $|G_-(X)|$ is bounded we infer that $YG_+(X)$ is integrable. It follows that $|Y|(|G_+(X)| + |G_-(X)|)$ is integrable. In the same way, $|G_+|(X) + |G_-|(X)$ is integrable. Since $|dh| = \mu_+ + \mu_-$, we have that $\overline{h} = |G_+| + |G_-|$, and consequently both $Y\overline{h}(X)$ and $Y\overline{h}(X^*)$ are integrable. From (2.4) we infer that (2.3) holds. Now applying Fubini in (2.2), we obtain (2.1). To prove inequalities 1 and 2, note that

$$|\mathrm{Cov}(Y, \mathbb{1}_{X\le t})| \le \mathbb{E}(|Y|\,|F_{X|\mathcal{M}}(t) - F_X(t)|)\,.$$

Consequently

$$|\mathrm{Cov}(Y, h(X))| \le \mathbb{E}\Big(|Y| \int |F_{X|\mathcal{M}}(t) - F_X(t)|\, |dh|(t)\Big). \tag{2.5}$$

Inequalities 1 and 2 follow from (2.5). □

3. Comparison of coefficients

The following Lemma will be very useful to obtain upper bounds for $\tau(\mathcal{M}, X)$, $\alpha(\mathcal{M}, X)$, $\beta(\mathcal{M}, X)$ and $\phi(\mathcal{M}, X)$.

Lemma 2. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X a real-valued random variable and $\mathcal{M}$ a σ-algebra of $\mathcal{A}$. If X^* is a random variable distributed as X and independent of $\mathcal{M}$ then*

1. *$\tau(\mathcal{M}, X) \le \|X - X^*\|_1$. Moreover, if Ω is rich enough, one can choose X^* such that $\tau(\mathcal{M}, X) = \|X - X^*\|_1$.*
2. *Assume that X has a continuous distribution function F. For any $y \in [0, 1]$, we have that*
$$\beta(\mathcal{M}, X) \le y + \mathbb{P}(|F(X) - F(X^*)| > y)\,.$$
3. *Assume that X has a continuous distribution function F. For any $y \in [0, 1]$, we have that*
$$\phi(\mathcal{M}, X) \le y + \|\mathbb{E}(\mathbb{1}_{|F(X)-F(X^*)|>y}|\mathcal{M})\|_\infty\,.$$
In particular, taking $y = \|F(X) - F(X^*)\|_\infty$ in the previous inequality, we obtain that $\phi(\mathcal{M}, X) \le \|F(X) - F(X^*)\|_\infty$.

Using this Lemma, we can now compare $\tau(\mathcal{M}, X)$, $\alpha(\mathcal{M}, X)$, $\beta(\mathcal{M}, X)$ and $\phi(\mathcal{M}, X)$.

Proposition 2. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X a real-valued random variable and $\mathcal{M}$ a σ-algebra of $\mathcal{A}$.*

1. *We have the inequalities $\alpha(\mathcal{M}, X) \le \beta(\mathcal{M}, X) \le \phi(\mathcal{M}, X)$.*
2. *Let Q_X be the generalized inverse of the tail function $t \to \mathbb{P}(|X|>t)$: if $u \in]0, 1[$, $Q_X(u) = \inf\{t \in \mathbb{R} : \mathbb{P}(|X| > t) \le u\}$. We have the inequality*
$$\tau(\mathcal{M}, X) \le 2\int_0^{\alpha(\mathcal{M},X)} Q_X(u)du\,.$$
3. *Assume moreover that X has a continuous distribution function F with modulus of continuity w. Define the function g by $g(x) = xw(x)$. Then*
$$\beta(\mathcal{M}, X) \le \frac{2\tau(\mathcal{M}, X)}{g^{-1}(\tau(\mathcal{M}, X))}\,. \tag{3.1}$$
In particular, if F is Hölder, that is $|F(x) - F(y)| \le C|x - y|^\alpha$ for $\alpha \in]0, 1]$ and $C > 0$, then
$$\beta(\mathcal{M}, X) \le 2C^{1/(\alpha+1)}\,(\tau(\mathcal{M}, X))^{\alpha/(\alpha+1)}\,.$$

If X has a density bounded by K, we obtain the bound

$$\beta(\mathcal{M}, X) \leq 2\sqrt{K\tau(\mathcal{M}, X)}\,. \tag{3.2}$$

Proof of Lemma 2. Item 1. has been proved in Dedecker and Prieur (2003). It remains to prove 2. and 3. □

Proof of 2. We shall use the following lemma, which gives the hereditary properties of $\alpha(\mathcal{M}, X)$, $\beta(\mathcal{M}, X)$ and $\phi(\mathcal{M}, X)$. □

Lemma 3. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X a real-valued random variable and $\mathcal{M}$ a σ-algebra of $\mathcal{A}$. If g is any nondecreasing function, then we have the inequalities $\alpha(\mathcal{M}, g(X)) \leq \alpha(\mathcal{M}, X)$, $\beta(\mathcal{M}, g(X)) \leq \beta(\mathcal{M}, X)$ and $\phi(\mathcal{M}, g(X)) \leq \phi(\mathcal{M}, X)$. In particular, for the distribution function F of X, we have $\alpha(\mathcal{M}, F(X)) = \alpha(\mathcal{M}, X)$, $\beta(\mathcal{M}, F(X)) = \beta(\mathcal{M}, X)$ and $\phi(\mathcal{M}, F(X)) = \phi(\mathcal{M}, X)$.*

Let $Y = F(X)$ and $Y^* = F(X^*)$. Clearly Y^* is independent of $\mathcal{M}$ and distributed as Y. According to Lemma 3, we have that $\beta(\mathcal{M}, X) = \beta(\mathcal{M}, Y)$. Hence, it suffices to prove the result for Y. Let $\mathbb{P}_{Y,Y^*|\mathcal{M}}$ be a conditional distribution of (Y, Y^*) given $\mathcal{M}$ (see Dudley (1989) Theorem 10.2.2 for the existence). Since F is continuous, $\mathbb{P}_{Y,Y^*|\mathcal{M}}$ has marginals $\mathbb{P}_{Y|\mathcal{M}}$ and $\mathbb{P}_{Y^*|\mathcal{M}} = \lambda$, where λ is the Lebesgue measure over $[0, 1]$. For any t, y in $[0, 1]$,

$$\begin{aligned} F_{Y|\mathcal{M}}(t) &= \int \mathbb{1}_{v+u-v\leq t}\mathbb{P}_{Y,Y^*|\mathcal{M}}(du, dv) \\ &\leq \int \mathbb{1}_{v\leq t+y}\mathbb{P}_{Y^*|\mathcal{M}}(dv) + \int \mathbb{1}_{v-u>y}\mathbb{P}_{Y,Y^*|\mathcal{M}}(du, dv) \\ &\leq t + y + \int \mathbb{1}_{v-u>y}\mathbb{P}_{Y,Y^*|\mathcal{M}}(du, dv)\,. \end{aligned}$$

In the same way,

$$1 - F_{Y|\mathcal{M}}(t) \leq 1 - (t - y) + \int \mathbb{1}_{u-v>y}\mathbb{P}_{Y,Y^*|\mathcal{M}}(du, dv)\,.$$

Consequently,

$$\begin{aligned} |F_{Y|\mathcal{M}}(t) - t| &\leq \max\left(F_{Y|\mathcal{M}}(t) - t, 1 - F_{Y|\mathcal{M}}(t) - (1 - t)\right) \\ &\leq y + \int \mathbb{1}_{|u-v|>y}\mathbb{P}_{Y,Y^*|\mathcal{M}}(du, dv)\,, \end{aligned} \tag{3.3}$$

and the result follows from (3.3) by taking the supremum in t and the expectation.

Proof of 3. The result also follows from (3.3). □

Proof of Lemma 3. Note first that, for any ω in Ω,

$$\sup_{t\in\mathbb{R}} |\mathbb{P}_{X|\mathcal{M}}(]-\infty, t]) - \mathbb{P}_X(]-\infty, t])| = \sup_{t\in\mathbb{R}} |\mathbb{P}_{X|\mathcal{M}}(]-\infty, t[) - \mathbb{P}_X(]-\infty, t[)|\,,$$

so that the definition of $\alpha(\mathcal{M}, X)$, $\beta(\mathcal{M}, X)$ and $\phi(\mathcal{M}, X)$ remains unchanged by taking the sets $]-\infty, t[$ instead of $]-\infty, t]$. Now if g is nondecreasing the set $\{x : g(x) \leq t\}$ is one of the sets $\emptyset$, $\mathbb{R}$, $]-\infty, a]$ or $]-\infty, a[$, a in $\mathbb{R}$. From this and the preceding remark, the first point follows. It remains to prove the second point. From the first point, we know that $\alpha(\mathcal{M}, F(X)) \leq \alpha(\mathcal{M}, X)$. Applying again the first point to the generalized inverse F^{-1} of F, we obtain that $\alpha(\mathcal{M}, F^{-1}(F(X))) \leq \alpha(\mathcal{M}, F(X))$. Since $F^{-1}(F(X)) = X$ almost surely (if F is constant on $[a, b]$ (resp. $[a, b[$), that equality may be false on the set $X^{-1}(]a, b])$ (resp. $X^{-1}(]a, b[)$) of probability 0), the result follows. The same arguments apply to $\beta(\mathcal{M}, X)$ and $\phi(\mathcal{M}, X)$. □

Proof of Proposition 2. Item 1. follows from the definition of $\alpha(\mathcal{M}, X)$, $\beta(\mathcal{M}, X)$ and $\phi(\mathcal{M}, X)$. Item 2. has been proved in Lemma 6 of Dedecker and Prieur (2003) and is based on a recent result by Peligrad (2002) (note that in Dedecker and Prieur $\alpha(\mathcal{M}, X)$ is one half of the coefficient $\alpha(\mathcal{M}, X)$ we use here). It remains to prove 3. Enlarging Ω if necessary, we know from Lemma 2 that there exists X^* independent of $\mathcal{M}$ and distributed as X such that $\|X - X^*\|_1 = \tau(\mathcal{M}, X)$. Since $|F(X) - F(X^*)| \leq w(|X - X^*|)$, we obtain from 2 of Lemma 2 (with $y = w(x)$)

$$\beta(\mathcal{M}, X) \leq w(x) + \mathbb{P}(w(|X - X^*|) > w(x)) \leq w(x) + \mathbb{P}(|X - X^*| > x)\,. \tag{3.4}$$

Applying Markov in (3.4), we get that

$$\beta(\mathcal{M}, X) \leq w(x) + \frac{\tau(\mathcal{M}, X)}{x}\,.$$

Inequality (3.1) follows by noting that $xw(x) = \tau(\mathcal{M}, X)$ for $x = g^{-1}(\tau(\mathcal{M}, X))$. □

4. Examples

We first define the coefficients $\tau(i)$, $\alpha(i)$, $\beta(i)$ and $\phi(i)$ of a sequence of real-valued random variables.

Definition 2. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let $(X_i)_{i\geq 0}$ be a sequence of integrable real-valued random variables and $(\mathcal{M}_i)_{i\geq 0}$ be a sequence of σ-algebras of $\mathcal{A}$. The sequence of coefficients $\tau(i)$ is then defined by*

$$\tau(i) = \sup_{k\geq 0} \tau(\mathcal{M}_k, X_{i+k})\,. \tag{4.1}$$

The coefficients $\alpha(i)$, $\beta(i)$ and $\phi(i)$ are defined in the same way.

Remark 2. One can also define the mixing coefficients $\alpha'(i)$, $\beta'(i)$ and $\phi'(i)$ as in (4.1), by taking $\sigma(X_{i+k})$ instead of X_{i+k}. It is clear from the definition that $\alpha(i) \leq \alpha'(i)$, $\beta(i) \leq \beta'(i)$ and $\phi(i) \leq \phi'(i)$.

In this section, we present four classes of examples for which we can compute upper bounds for the coefficients $\tau(i)$, $\alpha(i)$, $\beta(i)$ and $\phi(i)$. Among these examples, many are not mixing, in the sense that $\alpha'(i)$ does not even tends to zero. Some of the examples of Sections 4.1 and 4.2 have been also studied in Rio (1996, Section 1.2), Doukhan and Louhichi (1999) and Borovkova *et al.* (2001), but these authors do not provide any bounds for the coefficients we are interested in. In Section 4.1 and 4.2 we construct a sequence (X_k^*) coupled with (X_k), and we derive upper bounds for the coefficients by applying Lemma 2. In the context of functions of stationary sequences (Section 4.1), our approach is similar to that of Borovkova *et al.* (2001, Section 2), who used the coupling properties of the underlying sequence to obtain informations on the sequence $(X_k)_{k\geq 0}$. Some of the bounds for $\tau(i)$ in examples 4.1, 4.2 and 4.3 were given in Dedecker and Prieur (2003).

4.1. *Example 1: causal functions of stationary sequences*

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

Assume that there exists a stationary sequence $(\xi_i')_{i\in\mathbb{Z}}$ distributed as $(\xi_i)_{i\in\mathbb{Z}}$ and independent of $(\xi_i)_{i\leq 0}$. Define $X_n^* = H(\xi_n', \xi_{n-1}', \xi_{n-2}', \ldots)$. Clearly X_n^* is independent of $\sigma(X_i, i \leq 0)$ and distributed as X_n. For any $p \geq 1$ (p may be infinite) define the sequence $(\delta_{i,p})_{i>0}$ by

$$\|X_i - X_i^*\|_p = \delta_{i,p}\,. \tag{4.2}$$

Let $\mathcal{M}_i = \sigma(X_j, j \leq i)$. Arguing as in Lemma 2 and Proposition 2, we can easily prove that the coefficients τ, β and ϕ of the sequence $(X_n)_{n\geq 0}$ satisfy

1. $\tau(i) \leq \delta_{i,1}$.
2. Assume that X_0 has a continuous distribution function with modulus of continuity w. Define the function g_p by $g_p(y) = y(w(y))^{1/p}$. Then for any $1 \leq p < \infty$ we have
$$\alpha(i) \leq \beta(i) \leq 2\left(\frac{\delta_{i,p}}{g_p^{-1}(\delta_{i,p})}\right)^p\,.$$
In particular, if X_0 has a density bounded by K, we obtain that $\beta(i) \leq 2(K\delta_{i,p})^{\frac{p}{p+1}}$.
3. Assume that X_0 has a continuous distribution function with modulus of continuity w. Then $\alpha(i) \leq \beta(i) \leq \phi(i) \leq w(\delta_{i,\infty})$.

For $\phi(i)$, it is sometimes interesting to start from the first inequality in Lemma 2 item 3. For $1 \leq p < \infty$ define

$$\delta_{i,p}' = \|\mathbb{E}(|X_i - X_i^*|^p|\mathcal{M}_0)\|_\infty^{1/p}\,. \tag{4.3}$$

4. With the same notations as in item 2, we have $\phi(i) \leq 2\left(\dfrac{\delta'_{i,p}}{g_p^{-1}(\delta'_{i,p})}\right)^p$.

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

Application: causal linear processes. In that case $X_n = \sum_{j\geq 0} a_j \xi_{n-j}$. For any $p \geq 1$, we have that

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

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

$$\delta_{i,p} \leq \|\xi_0 - \xi'_0\|_p \sum_{j\geq i} |a_j| + \sum_{j=0}^{i-1} |a_j| \left(2^p \int_0^{\beta(\sigma(\xi_k, k\leq 0), \sigma(\xi_k, k\geq i-j))} Q_{\xi_0}^p(u)\right)^{1/p} du .$$

where Q_{ξ_0} is defined in Proposition 2 (note that in Merlevède and Peligrad the constant in front of the integral is 2^{p+2}. In fact it works with the constant 2^p).

If the sequence $(\xi_i)_{i\in\mathbb{Z}}$ is iid, it follows that $\delta_{i,p} \leq \|\xi_0 - \xi'_0\|_p \sum_{j\geq i} |a_j|$. Moreover, for $p = 2$ we have exactly $\delta_{i,2} = (2\text{Var}(\xi_0) \sum_{j\geq i} a_j^2)^{1/2}$. For instance, if $a_i = 2^{-i-1}$ and $\xi_0 \sim \mathcal{B}(1/2)$, then $\delta_{i,\infty} \leq 2^{-i}$. Since X_0 is uniformly distributed over [0, 1], we have $\phi(i) \leq 2^{-i}$. Recall that this sequence is not strongly mixing (see Andrews (1984)). More precisely, the coefficient $\alpha'(i)$ defined in Remark 2 is equal to 1/2.

4.2. Example 2: iterated random functions

Let $(X_n)_{n\geq 0}$ be a real-valued stationary Markov chain, such that $X_n = F(X_{n-1}, \xi_n)$ for some measurable function F and some i.i.d. sequence $(\xi_i)_{i>0}$ independent of X_0. Let X_0^* be a random variable distributed as X_0 and independent of $(X_0, (\xi_i)_{i>0})$. Define $X_n^* = F(X_{n-1}^*, \xi_n)$. The sequence $(X_n^*)_{n\geq 0}$ is distributed as $(X_n)_{n\geq 0}$ and independent of X_0. Let $\mathcal{M}_i = \sigma(X_j, 0 \leq j \leq i)$. As in Example 1, define the sequence $(\delta_{i,p})_{i>0}$ and $(\delta'_{i,p})_{i>0}$ by (4.2) and (4.3) respectively. The coefficients τ, β and ϕ of the sequence $(X_n)_{n\geq 0}$ satisfy 1, 2, 3 and 4 of Example 1.

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

$$\delta_{i,p}^p = \iint \|X_i^x - X_i^y\|_p^p \mu(dx)\mu(dy)$$
$$(\delta'_{i,p})^p = \inf\Big\{M : \mu\Big(\int \|X_i^x - X_i^y\|_p^p \mu(dy) > M\Big) = 0\Big\} .$$

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

$$\|X_i^x - X_i^y\|_p \leq d_{i,p}|x-y|\,,$$

then $\delta_{i,p} \leq d_{i,p}\|X_0 - X_0^*\|_p$ and $\delta'_{i,p} \leq d_{i,p}\|X_0 - X_0^*\|_\infty$. For instance, in the usual case where $\|F(x,\xi_0) - F(y,\xi_0)\|_p \leq \kappa|x-y|$ for some $\kappa < 1$, we can take $d_{i,p} = \kappa^i$.

An important example is $X_n = f(X_{n-1}) + \xi_n$ for some κ-lipschitz function f. If X_0 has a moment of order p, then $\delta_{i,p} \leq \kappa^i\|X_0 - X_0^*\|_p$. In particular, if X_0 is bounded and has a density bounded by K then $\phi(i) \leq 2K\|X_0\|_\infty \kappa^i$.

We refer to the nice review paper by Diaconis and Freedman (1999) for various examples of iterative random maps.

4.3. Example 3: Markov kernels.

Let P be a Markov kernel defined on a measurable subset $\mathcal{X}$ of $\mathbb{R}$. For any continuous bounded function f from $\mathcal{X}$ to $\mathbb{R}$ we have $P(f)(x) = \int_{\mathcal{X}} f(z)P(x,dz)$. Let $\Lambda_m(\mathcal{X})$ be the set of functions f from $\mathcal{X}$ to $\mathbb{R}$ such that $|f(x) - f(y)| \leq m|x-y|$. We make the following assumptions on P

H For some $0 < \kappa < 1$, P maps $\Lambda_1(\mathcal{X})$ to $\Lambda_\kappa(\mathcal{X})$.

Let $(X_n)_{n\geq 0}$ be a stationary Markov chain with values in $\mathcal{X}$, with marginal distribution μ and transition kernel P satisfying H. Let $\mathcal{M}_i = \sigma(X_j, 0 \leq j \leq i)$. By stationarity and the Markov property, we have that $\tau(i) = \tau(\sigma(X_0), X_i)$. Clearly the function $f_i = \mathbb{E}(f(X_i)|X_0 = x)$ belongs to $\Lambda_{\kappa^i}(\mathcal{X})$. Since

$$\tau(\sigma(X_0), X_i) \leq \iint \sup_{f\in\Lambda_1(\mathcal{X})} |f_i(x) - f_i(y)|\mu(dx)\mu(dy),$$

we infer that $\tau(i) \leq \kappa^i\|X_0 - X_0^*\|_1$ where X_0^* is independent and distributed as X_0. If furthermore X_0 has a density bounded by K, we infer from (3.2) of Proposition 2 that $\beta(i) \leq 2\sqrt{K\|X_0 - X_0^*\|_1\kappa^i}$.

In the case of iterated random maps (Example 2 above) the map F is a measurable function from $\mathcal{X} \times \mathcal{Y}$ to $\mathcal{X}$, and the kernel P has the form $P(f)(x) = \int_{\mathcal{Y}} f(F(x,z))\nu(dz)$ for some probability measure ν on $\mathcal{Y}$. Assumption H is satisfied as soon as

$$\int |F(x,z) - F(y,z)|\nu(dz) \leq \kappa|x-y|\,,$$

which was the condition previously found.

4.4. Example 4: dynamical systems on [0, 1].

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

For any finite measure ν on I, we use the notations $\nu(h) = \int_I h(x)\nu(dx)$. For any finite signed measure ν on I, let $\|\nu\| = |\nu|(I)$ be the total variation of ν. Denote by $\|g\|_{1,\lambda}$ the $\mathbb{L}^1$-norm with respect to the Lebesgue measure λ on I.

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

$$|\mathrm{Cov}(h(X_0), k(X_n))| \le a_n \|k(X_n)\|_1 (\|h\|_{1,\lambda} + \|dh\|) , \tag{4.4}$$

for some nonincreasing sequence a_n tending to zero as n tends to infinity.

Note that if (4.4) holds, then

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

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

$$|\mathrm{Cov}(h(X_0), k(X_n))| \le 2a_n \|k(X_n)\|_1 \|dh\| . \tag{4.5}$$

Inequality (4.5) is similar to the second inequality in Proposition 1 item 2, with $X = X_0$ and $Y = k(X_n)$, and one can wonder if $\phi(\sigma(X_n), X_0) \le 2a_n$. The answer is positive, due to the following Lemma.

Lemma 4. *Let* $(\Omega, \mathcal{A}, \mathbb{P})$ *be a probability space, X a real-valued random variable and* $\mathcal{M}$ *a* σ*-algebra of* $\mathcal{A}$*. We have the equality*

$$\phi(\mathcal{M}, X) = \sup\{|\mathrm{Cov}(Y, h(X))| : Y \textit{ is } \mathcal{M}\textit{-measurable, } \|Y\|_1 \le 1 \textit{ and } h \in BV_1\} .$$

Hence, we obtain an easy way to prove that a dynamical system $(T^i)_{i\ge 0}$ is ϕ-dependent:

$$\text{If (4.4) holds, then } \phi(\sigma(X_n), X_0) \le 2a_n. \tag{4.6}$$

In many cases, Inequality (4.4) follows from the spectral properties of the Markov operator associated to T. In these cases, due to the underlying Markovian structure, (4.6) holds with $\mathcal{M}_n = \sigma(X_i, i \ge n)$ instead of $\sigma(X_n)$.

Proof of Lemma 4. Write first $|\mathrm{Cov}(Y, h(X))| = |\mathbb{E}(Y(\mathbb{E}(h(X)|\mathcal{M}) - \mathbb{E}(h(X))))|$. For any positive ε, there exists A_ε in $\mathcal{M}$ such that $\mathbb{P}(A_\varepsilon) > 0$ and for any ω in A_ε,

$$|\mathbb{E}(h(X)|\mathcal{M})(\omega) - \mathbb{E}(h(X))| > \|\mathbb{E}(h(X)|\mathcal{M}) - \mathbb{E}(h(X))\|_\infty - \varepsilon.$$

Define the random variable

$$Y_\varepsilon := \frac{\mathbb{1}_{A_\varepsilon}}{\mathbb{P}(A_\varepsilon)} \operatorname{sign}\left(\mathbb{E}(h(X)|\mathcal{M}) - \mathbb{E}(h(X))\right) .$$

Y_ε is $\mathcal{M}$-measurable, $\mathbb{E}|Y_\varepsilon| = 1$ and $|\mathrm{Cov}(Y_\varepsilon, h(X))| \ge \|\mathbb{E}(h(X)|\mathcal{M}) - \mathbb{E}(h(X))\|_\infty - \varepsilon$. This being true for any positive ε, we infer from Lemma 1 that

$$\phi(\mathcal{M}, X) \le \sup\{|\mathrm{Cov}(Y, h(X))| : \text{Y is } \mathcal{M}\text{-measurable, } \|Y\|_1 \le 1 \text{ and } h \in BV_1\} .$$

The converse inequality follows straightforwardly from Lemma 1. □

Spectral gap. Define the operator $\mathcal{L}$ from $\mathbb{L}^1(I,\lambda)$ to $\mathbb{L}^1(I,\lambda)$ *via* the equality

$$\int_0^1 \mathcal{L}(h)(x)k(x)d\lambda(x) = \int_0^1 h(x)(k\circ T)(x)d\lambda(x) \quad \text{where } h\in\mathbb{L}^1(I,\lambda) \text{ and } k\in\mathbb{L}^\infty(I,\lambda).$$

The operator $\mathcal{L}$ is called the Perron-Frobenius operator of T. In many interesting cases, the spectral analysis of $\mathcal{L}$ in the Banach space of BV-functions equiped with the norm $\|h\|_v = \|dh\| + \|h\|_{1,\lambda}$ can be done by using the Theorem of Ionescu-Tulcea and Marinescu (see Lasota and Yorke (1974) and Hofbauer and Keller (1982)). Assume that 1 is a simple eigenvalue of $\mathcal{L}$ and that the rest of the spectrum is contained in a closed disk of radius strictly smaller than one. Then there exists a unique T-invariant absolutely continuous probability μ whose density f_μ is BV, and

$$\mathcal{L}^n(h) = \lambda(h)f_\mu + \Psi^n(h) \quad \text{with} \quad \|\Psi^n(h)\|_v \le K\rho^n\|h\|_v. \tag{4.7}$$

for some $0\le\rho<1$ and $K>0$. Assume moreover that:

$$I_* = \{f_\mu \neq 0\} \text{ is an interval, and there exists } \gamma>0 \text{ such that } f_\mu > \gamma^{-1} \text{ on } I_*. \tag{4.8}$$

Without loss of generality assume that $I_* = I$ (otherwise, take the restriction to I_* in what follows). Define now the Markov kernel associated to T by

$$P(h)(x) = \frac{\mathcal{L}(f_\mu h)(x)}{f_\mu(x)}. \tag{4.9}$$

It is easy to check (see for instance Barbour *et al.* (2000)) that $(X_0, X_1, \dots, X_n)$ has the same distribution as $(Y_n, Y_{n-1}, \dots, Y_0)$ where $(Y_i)_{i\ge0}$ is a stationary Markov chain with invariant distribution μ and transition kernel P. Since $\|fg\|_\infty \le \|fg\|_v \le 2\|f\|_v\|g\|_v$, we infer that, taking $C = 2K\gamma(\|df_\mu\|+1)$,

$$P^n(h) = \mu(h) + g_n \quad \text{with} \quad \|g_n\|_\infty \le C\rho^n\|h\|_v. \tag{4.10}$$

This estimate implies (4.4) with $a_n = C\rho^n$. Indeed,

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

Collecting the above facts, we infer that $\phi(\sigma(X_n), X_0) \le 2C\rho^n$. Moreover, using the Markov property we obtain that

$$\begin{aligned}
\phi(\sigma(X_n,\dots,X_{m+n}), X_0) &= \phi(\sigma(Y_0,\dots Y_m), Y_{n+m}) \\
&= \phi(\sigma(Y_m), Y_{n+m}) = \phi(\sigma(X_n), X_0).
\end{aligned}$$

This being true for any integer m, it holds for $\mathcal{M}_n = \sigma(X_i, i \geq n)$. We conclude that if (4.7) and (4.8) hold then there exists $C > 0$ and $0 \leq \rho < 1$ such that

$$\phi(\sigma(X_i, i \geq n), X_0) \leq 2C\rho^n \,. \tag{4.11}$$

Application: Expanding maps. Let $([a_i, a_{i+1}[)_{1 \leq i \leq N}$ be a finite partition of $[0, 1[$. We make the same assumptions on T as in Collet *et al* (2002).

1. For each $1 \leq j \leq N$, the restriction T_j of T to $]a_j, a_{j+1}[$ is strictly monotonic and can be extented to a function $\overline{T}_j$ belonging to $C^2([a_j, a_{j+1}])$.
2. Let I_n be the set where $(T^n)'$ is defined. There exists $A > 0$ and $s > 1$ such that $\inf_{x \in I_n} |(T^n)'(x)| > As^n$.
3. The map T is topologically mixing: for any two nonempty open sets U, V, there exists $n_0 \geq 1$ such that $T^{-n}(U) \cap V \neq \emptyset$ for all $n \geq n_0$.

If T satisfies 1. 2. and 3. then (4.7) holds. If furhtermore (4.8) holds (see Morita (1994) for sufficient conditions), then (4.11) holds.

Remark 3. The spectral analysis may be done under weaker assumptions on T (see Morita (1994) and Broise (1996)). In particular, the partition need not necessarily be finite: the gauss map $T(x) = x - [x]$ satisfies also (4.11). We have chosen this class of examples because it is easy to describe, and because we can go further in the analysis of the associated Markov chain $(Y_i)_{i \geq 0}$ by using a recent result of Collet *et al.* (2002) (see Example 4, Section 7.2).

5. MISE for $\boldsymbol{\beta}$-dependent sequences.

We consider the problem of estimating the unknown marginal density f from the observations $(X_1, \ldots, X_n)$ of a stationary sequence $(X_i)_{i \geq 0}$. In this context, Viennet (1997) obtained optimal results for the MISE under the condition $\sum_{k>0} \beta(\sigma(X_0), \sigma(X_k)) < \infty$. We wish to extend Viennet's results to sequences satisfying only

$$\sum_{k>0} \beta(\sigma(X_0), X_k) < \infty \,. \tag{5.1}$$

For kernel density estimators, this can be done by assuming only that the kernel K is BV and Lebesgue integrable. For projection estimators, it works only if the basis is well localized, because our variance inequality is less precise than that of Viennet. Note that Condition (5.1) is much less restrictive than Viennet's, for it contains many non mixing examples. In particular, since f is supposed to be square integrable with respect to the Lebesgue measure, the distribution function F of X_0 is 1/2-Hölder. Hence, we infer from point 3 of Proposition 2 that (5.1) holds as soon as $\sum_{k>0}(\tau(\sigma(X_0), X_k))^{1/3} < \infty$. If f is bounded (5.1) holds as soon as $\sum_{k>0}(\tau(\sigma(X_0), X_k))^{1/2} < \infty$.

5.1. Variance inequalities

According to Definition 2 and to the stationarity of $(X_i)_{i\geq 0}$, we set $\beta(i) = \beta(\sigma(X_0), X_i)$. The main results of this section are the following upper bounds (compare to Theorems 1.2 and 1.3(a) in Rio (2000a) for the mixing coefficients $\alpha(\sigma(X_0), \sigma(X_i))$).

Proposition 3. *Let K be any BV function such that $\int |K(x)|dx$ is finite. Let $(X_i)_{i\geq 0}$ be a stationary sequence, and define*

$$Y_{k,n} = h^{-1}K(h^{-1}(x - X_k)) \quad \text{and} \quad f_n(x) = \frac{1}{n}\sum_{k=1}^{n} Y_{k,n}\,. \tag{5.2}$$

The following inequality holds

$$nh\int \mathrm{Var}(f_n(x))dx \leq \int (K(x))^2 dx + 2\Big(\sum_{k=1}^{n-1}\beta(k)\Big)\|dK\|\int |K(x)|dx\,.$$

Proposition 4. *Let $(\varphi_i)_{1\leq i\leq n}$ be an orthonormal system of $\mathbb{L}^2(\mathbb{R}, \lambda)$ (λ is the Lebesgue measure) and assume that each φ_i is BV. Let $(X_i)_{i\geq 0}$ be a stationary sequence, and define*

$$Y_{j,n} = \frac{1}{n}\sum_{k=1}^{n}\varphi_j(X_k) \quad \text{and} \quad f_n = \sum_{j=1}^{m} Y_{j,n}\varphi_j\,. \tag{5.3}$$

The following inequality holds

$$n\int \mathrm{Var}(f_n(x))dx \leq \sup_{x\in\mathbb{R}}\Big(\sum_{j=1}^{m}\varphi_j^2(x)\Big) + 2\Big(\sum_{k=1}^{n-1}\beta(k)\Big)\sup_{x\in\mathbb{R}}\Big(\sum_{j=1}^{m}\|d\varphi_j\|\,|\varphi_j(x)|\Big)\,.$$

Remark 4. Since $\beta(\mathcal{M}, X) \leq \phi(\mathcal{M}, X)$, Propositions 3 and 4 apply to dynamical systems satisfying (4.4) with $2\sum_{i=1}^{n-1} a_k$ instead of $\sum_{i=1}^{n-1}\beta(k)$. For kernel estimators this can be also deduced from a variance estimate given in Prieur (2001).

Proof of Proposition 3. We start from the elementary inequality

$$\mathrm{Var}(f_n(x)) \leq \frac{1}{n}\|Y_{0,n}\|_2^2 + \frac{2}{n}\sum_{i=1}^{n-1}|\mathrm{Cov}(Y_{0,n}, Y_{i,n})|\,.$$

Now $h\int \|Y_{0,n}\|_2^2(x)dx = \int (K(x))^2 dx$. To complete the proof, we apply Proposition 1:

$$\begin{aligned} h\int |\mathrm{Cov}(Y_{0,n}, Y_{i,n})|(x)dx &\leq \|dK\|\mathbb{E}\Big(b(\sigma(X_0), X_i)\int |Y_{0,n}(x)|dx\Big) \\ &\leq \beta(i)\|dK\|\int |K(x)|dx\,. \end{aligned}$$

□

Proof of Proposition 4. Since $(\varphi_i)_{1\leq i\leq n}$ is an orthonormal system of $\mathbb{L}^2(\mathbb{R}, \lambda)$ we have that

$$\int \mathrm{Var}(f_n(x))dx = \sum_{j=1}^{m} \mathrm{Var}(Y_{j,n}) .$$

Applying Proposition 1, we obtain that

$$\begin{aligned}\mathrm{Var}(Y_{j,n}) &\leq \frac{1}{n}\|\varphi_j(X_0)\|_2^2 + \frac{2}{n}\sum_{k=1}^{n-1} |\mathrm{Cov}(\varphi_j(X_0), \varphi_j(X_k))| \\ &\leq \frac{1}{n}\|\varphi_j(X_0)\|_2^2 + \frac{2}{n}\sum_{k=1}^{n-1} \|d\varphi_j\|\mathbb{E}(|\varphi_j(X_0)|b(\sigma(X_0), X_k)) .\end{aligned}$$

To complete the proof we sum in j:

$$\begin{aligned}n\int \mathrm{Var}(f_n(x))dx &\leq \mathbb{E}\Big(\sum_{j=1}^{m} \varphi_j^2(X_0)\Big) \\ &\quad +2\sum_{k=1}^{n-1} \mathbb{E}\Big(b(\sigma(X_0), X_k) \sum_{j=1}^{m} \|d\varphi_j\|\, |\varphi_j(X_0)|\Big) .\end{aligned}$$

□

5.2. *Some function spaces*

In this section we recall the definition of the spaces $\mathrm{Lip}^*(s, 2, I)$, where I is either $\mathbb{R}$ or some compact interval $[a, b]$ (see DeVore and Lorentz (1993), Chapter 2). Let $I_{rh} = \mathbb{R}$ if $I = \mathbb{R}$ and $I_{rh} = [a, b - rh]$ otherwise. For any $h \geq 0$, let T_h be the translation operator $T_h(f, x) = f(x+h)$ and $\Delta_h = T_h - T_0$ be the difference operator. By induction, define the operators $\Delta_h^r = \Delta_h \circ \Delta_h^{r-1}$. Let λ be the Lebesgue measure on I and $\|.\|_{2,\lambda}$ the usual norm on $\mathbb{L}^2(I, \lambda)$. The modulus of smoothness of order r of a function f in $\mathbb{L}^2(I, \lambda)$ is defined by

$$\omega_r(f, t)_2 = \sup_{0\leq h\leq t} \|\Delta_h^r(f, .)\mathbb{1}_{I_{rh}}\|_{2,\lambda} ,$$

For $s > 0$, $\mathrm{Lip}^*(s, 2, I)$ is the space of functions f in $\mathbb{L}^2(I, \lambda)$ such that

$$\|f\|_{s,2,I} = \|f\|_{2,\lambda} + \sup_{t>0} \frac{\omega_{[s]+1}(f, t)_2}{t^s} < \infty .$$

These spaces are Banach spaces with respect to the norm $\|.\|_{s,2,I}$. Recall that $\mathrm{Lip}^*(s, 2, I)$ is a particular case of Besov spaces (precisely $\mathrm{Lip}^*(s, 2, I) = B_{s,2,\infty}(I)$) and that it contains Sobolev spaces $W_s(I) = B_{s,2,2}(I)$. Recall that, if s is an integer, the space $W_s(I)$ is the space of functions for which $f^{(s-1)}$ is absolutely continuous with almost everywhere derivative $f^{(s)}$ belonging to $\mathbb{L}^2(I, \lambda)$.

5.3. Application to Kernel estimators

If f_n is defined by (5.2), set $f_h = \mathbb{E}(f_n)$. Let r be some positive integer, and assume that the kernel K is such that: for any f belonging to the Sobolev space $W_r(\mathbb{R})$ we have

$$\int (f(x) - f_h(x))^2 dx \le M_1 h^{2r} \|f^{(r)}\|_2^2 , \tag{5.4}$$

for some constant M_1 depending only on r. From (5.4) and Theorem 5.2 page 217 in DeVore and Lorentz (1993), we infer that, for any f in $\mathbb{L}^2(\mathbb{R}, \lambda)$,

$$\int (f(x) - f_h(x))^2 dx \le M_2 (w_r(f, h)_2)^2 ,$$

for some constant M_2 depending only on r. This last inequality imply that, if f belongs to $\text{Lip}^*(s, 2, \mathbb{R})$ for $r - 1 \le s < r$, then

$$\int (f(x) - f_h(x))^2 dx \le M_2 h^{2s} \|f\|_{s,2,\mathbb{R}}^2 .$$

This evaluation of the bias together with Proposition 3 leads to the following Corollary.

Corollary 1. *Let r be some positive integer. Let $(X_i)_{i\ge 1}$ be a stationary sequence with common marginal density f belonging to* $\text{Lip}^*(s, 2, \mathbb{R})$ *with $r - 1 \le s < r$, or to $W_s(\mathbb{R})$ with $s = r$. Let K be a BV function satsisfying (5.4) and such that $\int |K(x)|dx$ is finite. Let f_n be defined by (5.2) with $h = n^{-1/(2s+1)}$. If (5.1) holds, then there exists a constant C such that*

$$\mathbb{E}\Big(\int (f_n(x) - f(x))^2 dx\Big) \le C n^{-2s/(2s+1)} .$$

Here are two well known classes of kernel satisfying (5.4).

Example 1. One says that K is a kernel of order k, if

1. $\int K(x)dx = 1, \quad \int (K(x))^2 dx < \infty \quad \text{and} \quad \int |x|^{k+1} |K(x)| dx < \infty .$
2. $\int x^j K(x) dx = 0$ for $1 \le j \le k$.

If K is a Kernel of order k, then it satisfies (5.4) for any $r \le k + 1$. For instance, the naive kernel $K = (1/2)\mathbb{1}_{]-1,1]}$ is BV and of order 1. Consequently Corollary 1 applies to functions belonging to $\text{Lip}^*(s, 2, \mathbb{R})$ for $s < 2$, or to $W_2(\mathbb{R})$.

Example 2. Assume that the fourier transform K^* of K satisfies $|1 - K^*(x)| \le M|x|^r$ for some positive constant M. Then K satisfies (5.4) for this r. For instance, $K(x) = \sin(x)/(\pi x)$ satisfies (5.4) for any positive integer r. Unfortunately, it is neither BV nor integrable. Another function satisfying (5.4) for any positive integer r is the analogue of the de la vallée-Poussin kernel $V(x) = (\cos(x) - \cos(2x))/\pi x^2$. This function is BV and integrable, so that Corollary 1 apply to any function belonging to $\text{Lip}^*(s, 2, I)$ for $s > 0$.

5.4. Application to unconditional systems.

Proposition 4 is of special interest for orthonormal systems $(\varphi_i)_{i\geq 1}$ satisfying the two conditions:

P1 There exists C_1 independent of m such that $\max_{1\leq i\leq m} \|d\varphi_i\| \leq C_1\sqrt{m}$.

P2 There exists C_2 independent of m such that $\sup_{x\in\mathbb{R}}\Big(\sum_{j=1}^{m} |\varphi_j(x)|\Big) \leq C_2\sqrt{m}$.

An orthonormal system satisfying P2 is called *unconditional*. For such systems, we obtain from Proposition 4 that

$$n\int \mathrm{Var}(f_n(x))dx \leq m\Big(C_2^2 + 2C_1C_2\Big(\sum_{k=1}^{n-1}\beta(k)\Big)\Big)\,. \tag{5.5}$$

Example 1: piecewise polynomials. Let $(Q_i)_{1\leq i\leq r+1}$ be an orthonormal basis of the space of polynomials of order r on $[0, 1]$ and define the function R_i on $\mathbb{R}$ by: $R_i(x) = Q_i(x)$ if x belongs to $]0, 1]$ and 0 otherwise. We consider the regular partition of $]0, 1]$ into k intervals $(](j-1)/k, j/k])_{1\leq j\leq k}$. Define the functions $R_{i,j}(x) = \sqrt{k}R_i(kx-(j-1))$. Clearly the family $(R_{i,k})_{1\leq i\leq r+1}$ is an orthonormal basis of the space of polynomials of order r on the interval $[(j-1)/k, j/k]$. Let $m = k(r+1)$ and $(\varphi_i)_{i\geq 1}$ be any family such that

$$\{\varphi_i, 1\leq i\leq m\} = \{R_{i,j}, 1\leq j\leq k, 1\leq i\leq r+1\}\,. \tag{5.6}$$

The orthonormal system $(\varphi_i)_{i\geq 1}$ satisfies P1 and P2 with

$$C_1 = (r+1)^{-1/2}\max_{1\leq i\leq r+1}\|dR_i\| \quad \text{and} \quad C_2 = (r+1)^{-1/2}\sup_{x\in[0,1]}\Big(\sum_{i=1}^{r+1}|R_i(x)|\Big)\,.$$

The case of histograms corresponds to $r = 0$. In that case $\varphi_j = \sqrt{k}\mathbb{1}_{](j-1)/k, j/k]}$. Clearly $C_2 = 1$ and $\|d\varphi_j\| = 2\sqrt{k}$, so that $C_1 = 2$.

Assume now that X_0 has a density f such that $f\mathbb{1}_{[0,1]}$ belongs to $\mathrm{Lip}^*(s,2,[0,1])$. Suppose that $r > s-1$, and denote by $\bar{f}$ the orthogonal projection of f on the subspace generated by $(\varphi_i)_{1\leq 1\leq m}$. From Lemma 12 in Barron *et al.* (1999) we know that there exists a constant K depending only on s such that

$$\int_0^1 (f(x)-\bar{f}(x))^2dx \leq Km^{-2s}\,. \tag{5.7}$$

Since $\bar{f} = \mathbb{E}(f_n)$, we obtain from (5.5) and (5.7) the following corollary.

Corollary 2. *Let $(X_i)_{i\geq 1}$ be a stationary sequence with common marginal density f such that $f\mathbb{1}_{[0,1]}$ belongs to* $\mathrm{Lip}^*(s, 2, [0, 1])$. *Let r be any nonnegative integer such that $r > s-1$ and $k = [n^{1/(2s+1)}]$. Let $(\varphi_i)_{1\leq i\leq m}$ be defined by (5.6) and f_n be defined by (5.3). If (5.1) holds, then there exists a constant C such that*

$$\mathbb{E}\Big(\int_0^1 (f_n(x)-f(x))^2dx\Big) \leq Cn^{-2s/(2s+1)}\,.$$

Example 2: wavelet basis. Let $\{e_{j,k}, j \geq 0, k \in \mathbb{Z}\}$ be an orthonormal wavelet basis with the following convention: $e_{0,k}$ are translate of the father wavelet and for $j \geq 1$, $e_{j,k} = 2^{j/2}\psi(2^j x - k)$, where ψ is the mother wavelet. Assume that these wavelets are compactly supported and have continuous derivatives up to order r (if $r = 0$, the wavelets are supposed to be BV). Let g be some function with support in $[-A, A]$. Changing the indexation of the basis if necessary, we can write $g = \sum_{j\geq 0}\sum_{k=1}^{2^j M} a_{j,k}e_{j,k}$, where $M \geq 1$ is some finite integer depending on A and on the size of the wavelets supports. Let $m = \sum_{j=0}^{J} 2^j M$ and $(\varphi_i)_{i\geq 1}$ be any family such that

$$\{\varphi_i, 1 \leq i \leq m\} = \{e_{j,k}, 0 \leq j \leq J, 1 \leq k \leq 2^j M\}\,. \tag{5.8}$$

The orthonormal system $(\varphi_i)_{i\geq 1}$ satisfies P1 and P2.

Assume now that X_0 has a density f belonging to $\text{Lip}^*(s, 2, \mathbb{R})$ with compact support in $[-A, A]$. Denote by $\bar{f}$ the orthogonal projection of f on the subspace generated by $(\varphi_i)_{1\leq i\leq m}$. From Lemma 12 in Barron *et al.* (1999) we know that there exist a constant K depending only on s such that

$$\int_0^1 (f - \bar{f}(x))^2 dx \leq K2^{-2Js}\,. \tag{5.9}$$

Since $\bar{f} = \mathbb{E}(f_n)$, we obtain from (5.5) and (5.9) the following corollary.

Corollary 3. *Let $(X_i)_{i\geq 1}$ be a stationary sequence with common marginal density f belonging to* $\text{Lip}^*(s, 2, \mathbb{R})$ *and with compact support in* $[-A, A]$. *Let r be any nonnegative integer such that $r > s - 1$ and J be such that* $J = [\log_2(n^{1/(2s+1)})]$. *Let $(\varphi_i)_{1\leq i\leq m}$ be defined by (5.8) and f_n be defined by (5.3). If (5.1) holds, then there exists a constant C such that*

$$\mathbb{E}\Big(\int (f_n(x) - f)^2 dx\Big) \leq Cn^{-2s/(2s+1)}\,.$$

Remark 5. More generally, if $\sum_{i=1}^{n} \beta(\sigma(X_0), X_i) = O(n^a)$ for some a in $[0, 1[$, we obtain the rate $n^{-2s(1-a)/(2s+1)}$ for the MISE in Corollaries 1, 2 and 3. Note that if (5.1) holds the rate $n^{-2s/(2s+1)}$ is known to be optimal for i.i.d. observations.

6. Exponential inequality for ϕ-dependent sequences

Starting from a moment inequality of Dedecker and Doukhan (2003) (see also Theorem 2.5 in Rio (2000a) for the stationary case) we obtain an Hoeffding-type inequality for partial sums. Given a filtration $\mathcal{M}_i$, the coefficients $\phi(k)$ are defined as in (4.1).

Proposition 5. *Let $(X_i)_{i\geq 0}$ be a sequence of random variables and $\mathcal{M}_i = \sigma(X_j, 1 \leq j \leq i)$. For any BV function h, define*

$$S_n(h) = \sum_{i=1}^{n} h(X_i) \quad \textit{and} \quad b_{i,n} = \Big(\sum_{k=0}^{n-i} \phi(k)\Big)\|dh\|\,\|h(X_i) - \mathbb{E}(h(X_i))\|_{p/2}\,.$$

For any $p \geq 2$ we have the inequality

$$\|S_n(h) - \mathbb{E}(S_n(h))\|_p \leq \Big(2p \sum_{i=1}^{n} b_{i,n}\Big)^{1/2} \leq \|dh\| \Big(2p \sum_{k=0}^{n-1} (n-k)\phi(k)\Big)^{1/2}. \tag{6.1}$$

We also have that

$$\mathbb{P}(|S_n(h) - \mathbb{E}(S_n(h))| > x) \leq e^{1/e} \exp\left(\frac{-x^2}{4e\|dh\|^2 \sum_{k=0}^{n-1}(n-k)\phi(k)}\right). \tag{6.2}$$

Remark 6. Applying the method of martingale differences, as in Deddens, Peligrad and Yang (1987), we can also prove that

$$\mathbb{P}(|S_n(h) - \mathbb{E}(S_n(h))| > x) \leq 2\exp\left(\frac{-x^2}{2\|dh\|^2 \sum_{i=1}^{n}\big(1 + 2\sum_{k=1}^{n-i+1} \phi(k)\big)^2}\right). \tag{6.3}$$

Both (6.2) and (6.3) yield the same kind of inequality provided that $\sum_{k>0} \phi(k)$ is finite. Note that this condition is realized for expanding maps considered in Example 4. We shall see that for such maps we can also obtain a concentration inequality for lipschitz functions (cf. Collet *et al* (2002) and Theorem 2, Section 7.4).

Proof of Proposition 5. Let $Y_i = h(X_i) - \mathbb{E}(h(X_i))$. Applying Proposition 4 in Dedecker and Doukhan (2003), we obtain that

$$\|S_n(h) - \mathbb{E}(S_n(h))\|_p \leq \Big(2p \sum_{i=1}^{n} \max_{i \leq l \leq n} \Big\| Y_i \sum_{k=i}^{l} \mathbb{E}(Y_k|\mathcal{M}_i)\Big\|_{p/2}\Big)^{1/2}. \tag{6.4}$$

From Item 4. of Lemma 1, we infer that

$$\max_{i \leq l \leq n} \Big\| Y_i \sum_{k=i}^{l} \mathbb{E}(Y_k|\mathcal{M}_i)\Big\|_{p/2} \leq \|Y_i\|_{p/2} \sum_{k=i}^{n} \|\mathbb{E}(Y_k|\mathcal{M}_i)\|_\infty \leq b_{i,n}. \tag{6.5}$$

The first inequality in (6.1) follows from (6.4) and (6.5). To prove the second inequality in (6.1), it remains to bound $b_{i,n}$. From Lemma 1, $\|Y_i\|_{p/2} \leq \|Y_i\|_\infty \leq \|dh\|\phi(0) \leq \|dh\|$, so that $b_{i,n} \leq \|dh\|_2^2(\phi(0) + \cdots + \phi(n-i))$ and (6.1) is proved.

To prove (6.2), let $B = \|dh\|^2 \sum_{k=0}^{n-1}(n-k)\phi(k)$. For any $p \geq 2$ we have

$$\begin{aligned}\mathbb{P}(|S_n(h) - \mathbb{E}(S_n(h))| > x) &\leq \min\Big(1, \frac{\mathbb{E}(|S_n(h) - \mathbb{E}(S_n(h))|^p)}{x^p}\Big)\\ &\leq \min\Big(1, \Big(\frac{2pB}{x^2}\Big)^{\frac{p}{2}}\Big).\end{aligned}$$

Obvious computations show that the function $p \to (2pBx^{-2})^{p/2}$ has a unique minimum in $p_0 = (2eB)^{-1}x^2$ and is increasing on the interval $[p_0, +\infty]$. By comparing p_0 and 2, we infer that

$$\mathbb{P}(|S_n(h) - \mathbb{E}(S_n(h))| > x) \leq g\Big(\frac{x^2}{4eB}\Big),$$

where g is the function from $\mathbb{R}_+$ to $\mathbb{R}_+$ defined by

$$g(y) = \mathbb{1}_{y \le e^{-1}} + (ey)^{-1}\mathbb{1}_{e^{-1} < y \le 1} + e^{-y}\mathbb{1}_{y>1}\,.$$

Finally, (6.2) follows by noting that $g(y) \le \exp(-y + e^{-1})$ for any positive y. □

From Proposition 5 we obtain an empirical central limit theorem for classes of BV functions. We need some notations. Let $(X_i)_{i\ge 0}$ be a stationary sequence of real-valued random variables with common marginal distribution P. Denote by P_n the empirical probability measure and by Z_n the centered and normalized empirical measure

$$P_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}, \qquad Z_n = \sqrt{n}(P_n - P)\,.$$

Let $\mathcal{F}$ be a class of measurable functions from $\mathbb{R}$ to $\mathbb{R}$. The space $\ell^\infty(\mathcal{F})$ is the space of all functions z from $\mathcal{F}$ to $\mathbb{R}$ such that $\|z\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} |z(f)|$ is finite. A random variable X with values in $\ell^\infty(\mathcal{F})$ is tight if for any positive ϵ there exists a compact set K_ϵ of $(\ell^\infty(\mathcal{F}), \|.\|_{\mathcal{F}})$ such that $\mathbb{P}(X \in K_\epsilon) \ge 1 - \epsilon$.

For any P-integrable function f, let $P_n(f) = \int f(x)P_n(dx)$, $P(f) = \int f(x)P(dx)$ and $Z_n(f) = \sqrt{n}(P_n(f) - P(f))$. Assume that $P(|f|)$ is finite for any f in $\mathcal{F}$ and that $\sup_{f\in\mathcal{F}} |f(x) - P(f)|$ is finite for every x in $\mathbb{R}$. Under this minimal condition, the empirical process $\{Z_n(f), f \in \mathcal{F}\}$ can be viewed as a variable with values in $\ell^\infty(\mathcal{F})$, altough it may not be measurable with respect to the Borel σ-algebra generated by $\|.\|_{\mathcal{F}}$. Nevertheless, we say that Z_n converges weakly to a $\ell^\infty(\mathcal{F})$-valued random variable Z (i.e. Borel measurable) if, for every continuous bounded function h from $(\ell^\infty(\mathcal{F}), \|.\|_{\mathcal{F}})$ to $\mathbb{R}$, the outer expectation $\mathbb{E}^*(h(Z_n))$ converges to $\mathbb{E}(h(Z))$ (see for instance van der Vaart and Wellner (1996) p. 4 for the definition of outer expectations and measures, and more details about weak convergence for non-measurable maps).

If ρ is a seminorm on $\mathcal{F}$, the metric entropy $H(\varepsilon, \mathcal{F}, \rho)$ is the logarithm of the smallest number of balls with radius ε (with respect to ρ) needed to cover $\mathcal{F}$.

Corollary 4. *Let $(X_i)_{i\in\mathbb{Z}}$ be a stationary and ergodic sequence of real-valued random variables and $\mathcal{M}_i = \sigma(X_j, j \le i)$. Let $\mathcal{F}$ be a class of BV functions. On $\mathcal{F}$ we put the seminorm $|f|_v = \|df\|$. Let $\mathcal{M}_i = \sigma(X_j, j \le i)$ and assume that*

$$\sum_{k=1}^{\infty} \phi(k) < \infty \quad \textit{and} \quad \int_0^1 \sqrt{H(x, \mathcal{F}, |.|_v)}dx < \infty\,.$$

Then Z_n converges weakly in $\ell^\infty(\mathcal{F})$ to a tight gaussian process with covariance function

$$\Gamma(f, g) = \sum_{k\in\mathbb{Z}} \mathrm{Cov}(f(X_0), g(X_k))\,.$$

Application. Assume that X_0 belongs to $[0, 1]$ and that $\mathcal{F}$ is a class of absolutely continuous functions from $[0, 1]$ to $\mathbb{R}$. In that case, $|f|_v = \int_0^1 |f'(t)|dt = \|f'\|_{1,\lambda}$. If $\mathcal{F}' = \{f', f \in \mathcal{F}\}$ then the condition on the entropy can be written as

$$\int_0^1 \sqrt{H(x, \mathcal{F}', \|.\|_{1,\lambda})}dx < \infty\,.$$

For instance, it is satisfied if $\mathcal{F}'$ is the class of increasing functions from [0, 1] to $[-K, K]$, which means that $\mathcal{F}$ is the class of convex and K-lipschitz functions.

Proof of Corollary 4. Let $C(\phi) = \sum_{k=0}^{\infty} \phi(k)$. Applying (6.2) of Proposition 5, we obtain

$$\mathbb{P}(|Z_n(f) - Z_n(g)| > x) \leq e^{1/e} \exp\left(\frac{-x^2}{4e|f - g|_v^2 C(\phi)}\right). \tag{6.6}$$

This means that for each n, the process $\{Z_n(f), f \in \mathcal{F}\}$ is subgaussian (cf. Ledoux and Talagrand (1991), p. 322). We can therefore apply the chaining procedure of Theorem 11.6 in Ledoux and Talagrand (1991) (with the outer expectation $\mathbb{E}^*$ instead of $\mathbb{E}$) to obtain that: for each positive ε there exists a positive real δ, depending only on ε and of the value of the entropy integral, such that $\mathbb{E}^*(\sup_{|f-g|_v<\delta} |Z_n(f) - Z_n(g)|) < \varepsilon$. This prove that Z_n is asymptotically $|.|_v$-equicontinuous.

To complete the proof, it remains to check the finite dimensional convergence of the process Z_n. Let $\mathbf{f} = (f_1, \dots, f_k)$ be an element of $\mathcal{F}^k$ and for any x in $\mathbb{R}^k$ define the function $< x, \mathbf{f} - P(\mathbf{f}) >= x_1(f_1 - P(f_1)) + \cdots + x_k(f_k - P(f_k))$. Define the matrix C by $C_{i,j} = \Gamma(f_i, f_j)$. Since $(X_i)_{i\in\mathbb{Z}}$ is ergodic, we infer from Dedecker and Rio (2000) that the random variable $Z_n(< x, \mathbf{f} - P(\mathbf{f}) >)$ converges in distribution to a mean-zero normal distribution with variance $x^t C x$ as soon as

$$\sum_{k\geq 0} \| < x, \mathbf{f} - P(\mathbf{f}) > (X_0)\mathbb{E}(< x, \mathbf{f} - P(\mathbf{f}) > (X_k)|\mathcal{M}_0)\|_1 < \infty. \tag{6.7}$$

Consequently, if (6.7) holds, the random vector $(Z_n(f_1), \dots, Z_n(f_k))$ converges in distribution to a Gaussian vector with covariance matrix C. Applying Lemma 1, we obtain that

$$\begin{aligned} &\| < x, \mathbf{f} - P(\mathbf{f}) > (X_0)\mathbb{E}(< x, \mathbf{f} - P(\mathbf{f}) > (X_k)|\mathcal{M}_0)\|_1 \\ &\qquad \leq \| < x, \mathbf{f} - P(\mathbf{f}) > (X_0)\|_\infty | < x, \mathbf{f} - P(\mathbf{f}) > |_v\, \alpha(\mathcal{M}_0, X_k), \end{aligned}$$

so that (6.7) holds as soon as $\sum_{k\geq 0} \alpha(k)$ is finite. This completes the proof. □

7. Extension to higher dimension

It seems difficult to extend coefficients based on the conditional distribution function in higher dimension. A way to proceed is to start from the functional definition of the coefficients given in Lemma 1. For $\alpha(\mathcal{M}, X)$, $\beta(\mathcal{M}, X)$ and $\phi(\mathcal{M}, X)$ the extension remains difficult because the notion of bounded variation is rather delicate even in $\mathbb{R}^2$. For τ, the extension is immediate and satisfactory.

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\mathcal{M}$ a σ-agebra of $\mathcal{A}$ and X a random variable with values in a Polish space $(\mathcal{X}, d)$. As in $\mathbb{R}$ there exists a conditional distribution $\mathbb{P}_{X|\mathcal{M}}$ of X given $\mathcal{M}$ (see Dudley (1989), Theorem 10.2.2). Let $\Lambda_1(\mathcal{X})$

be the space of 1-lipschitz functions from $\mathcal{X}$ to $\mathbb{R}$. Assume that $\int d(0,x)\mathbb{P}_X(dx)$ is finite and define

$$\tau(\mathcal{M}, X) = \left\| \sup\left\{ \left| \int f(x)\mathbb{P}_{X|\mathcal{M}}(dx) - \int f(x)\mathbb{P}_X(dx) \right|, f \in \Lambda_1(\mathcal{X}) \right\} \right\|_1 . \tag{7.1}$$

We shall see in section 7.1 that this coefficient has the same coupling property as in the real case. If $d(0, X)$ is bounded, we can define the uniform version of τ, which was first introduced by Rio (1996):

$$\varphi(\mathcal{M}, X) = \sup\{\|\mathbb{E}(f(X)|\mathcal{M}) - \mathbb{E}(f(X))\|_\infty, f \in \Lambda_1(\mathcal{X})\} .$$

Note that this definition slightly differs from Rio's, who takes $\Lambda_1(\mathcal{X})$ as the set of 1-Lipschitz functions from $\mathcal{X}$ to $[0, 1]$. With our definitions, $\tau(\mathcal{M}, X)$ and $\varphi(\mathcal{M}, X)$ have an interpretation in terms of the Kantorovitch-Rubinstein distance (see Section 7.1 below).

The main advantage of such definitions in spaces of higher dimension is that it allows to define the dependence between two sequences $(X_i)_{i\geq 0}$ and $(\mathcal{M}_i)_{i\geq 0}$ by considering k-tuples in the future and not only a single variable. More precisely, put the distance $d_1(x, y) = d(x_1, y_1) + \cdots + d(x_k, y_k)$ on $\mathcal{X}^k$, and define

$$\tau_k(i) = \max_{1\leq l\leq k} \frac{1}{l} \sup\{\tau(\mathcal{M}_p, (X_{j_1}, \ldots, X_{j_l})), p + i \leq j_1 < \cdots < j_l\} \text{ and}$$
$$\tau_\infty(i) = \sup_{k>0} \tau_k(i).$$

The coefficient φ_k and φ_∞ are defined in the same way.

7.1. Coupling

Let P and Q be two probability measures on a Polish space $(\mathcal{X}, d, \mathcal{B}(\mathcal{X}))$. In 1970 Dobrushin proved that there exists a probality measure μ on $\mathcal{X} \times \mathcal{X}$ such that $\mu(\cdot \times \mathcal{X}) = P(\cdot)$, $\mu(\mathcal{X} \times \cdot) = Q(\cdot)$ and

$$\frac{1}{2}\|P - Q\| = \mu(\{x \neq y, (x, y) \in \mathcal{X} \times \mathcal{X}\}) , \tag{7.2}$$

where $\|.\|$ is the variation norm. Starting from (7.2) (cf. Proposition 4.2.1 in Berbee (1979)), Berbee obtained the following coupling result: let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $\mathcal{M}$ a σ-algebra of $\mathcal{A}$ and X a $\mathcal{X}$-valued random variable. If Ω is rich enough, there exists X^* distributed as X and independent of $\mathcal{M}$ such that

$$\frac{1}{2}\|\mathbb{P}_{X|\mathcal{M}} - \mathbb{P}_X\| = \mathbb{E}(\mathbb{1}_{X\neq X^*}|\mathcal{M}) \text{ almost surely} \tag{7.3}$$

From (7.3), it follows that $\beta(\mathcal{M}, \sigma(X)) = \mathbb{P}(X \neq X^*)$.

It is by now well known that Dobrushin's result (7.2) is a particular case of the Monge-Kantorovitch problem (see for instance Rachev and Rüschendorf (1998), page 93). More precisely, let d_0 be the discrete metric $d_0(x, y) = \mathbb{1}_{x\neq y}$ and

$\Lambda_1(\mathcal{X}, d_0)$ be the set of Borel functions from $\mathcal{X}$ to $\mathbb{R}$ such that $|f(x) - f(y)| \leq d_0(x, y)$. Property (7.2) is equivalent to: there exists a probability μ on $\mathcal{X} \times \mathcal{X}$ such that $\mu(\cdot \times \mathcal{X}) = P(\cdot)$, $\mu(\mathcal{X} \times \cdot) = Q(\cdot)$ and

$$K_{d_0}(P, Q) := \sup\Big\{\Big|\int f(x)P(dx) - \int f(x)Q(dx)\Big|, f \in \Lambda_1(\mathcal{X}, d_0)\Big\} = \int d_0(x, y)\mu(dx, dy). \tag{7.4}$$

In fact, (7.4) holds for a wide class of distances (for instance the distances satisfying the equality (4.5.1) in Rachev and Rüschendorf (1998)). In particular it holds for any continuous (with respect to d) distance d_0. In that case, one can prove an analogue of Berbee's result: if Ω is rich enough, there exists X^* distributed as X and independent of $\mathcal{M}$ such that

$$K_{d_0}(\mathbb{P}_{X|\mathcal{M}}, \mathbb{P}_X) = \mathbb{E}(d_0(X, X^*)|\mathcal{M}) \text{ almost surely.} \tag{7.5}$$

If $d_0 = d$ and $\mathcal{M} = \sigma(Z)$ for some random variable Z with values in a Polish space $\mathcal{Z}$, property (7.5) has been proved in Dedecker and Prieur (2004). Our proof is based on a conditional version of the Kantorovitch and Rubinstein theorem (see Proposition 1.2). After this note was published, we read the book by Castaing *et al.* (2004) on Young measures. Using the equality (3.4.4) in Castaing *et al.* (2004) instead of Proposition 1.2 in Dedecker and Prieur (2004), we see that (7.5) is true for any σ-agebra $\mathcal{M}$ of $\mathcal{A}$ and any continuous distance d_0 with respect to d. From (7.5) with $d_0 = d$ we obtain that $\tau(\mathcal{M}, X) = \mathbb{E}(d(X, X^*))$.

Starting from their coupling properties, one can compare the coefficients $\beta(\mathcal{M}, \sigma(X))$ and $\tau(\mathcal{M}, X)$. Following the proof of Proposition 2.3 in Merlevède and Peligrad (2002), we obtain that, for any x in $\mathcal{X}$,

$$\tau(\mathcal{M}, X) \leq 2\int_0^{\beta(\mathcal{M},\sigma(X))} Q_{d(X,x)}(u)du\,, \tag{7.6}$$

where the function $Q_{d(X,x)}$ is defined as in Proposition 2 (note that (7.6) can be deduced from Proposition 2.3 of Merlevède and Peligrad (2002) with a constant 8 instead of 2). If $\mathcal{X} = \mathbb{R}$, we know from item 2 of Proposition 2 that (7.6) holds for $x = 0$ and the weak coefficient $\alpha(\mathcal{M}, X)$. A reasonable question is then: can we obtain a bound similar to (7.6) for any polish space $\mathcal{X}$ with the mixing coefficient $\alpha(\mathcal{M}, \sigma(X))$ instead of $\beta(\mathcal{M}, \sigma(X))$? In fact, this is not true in general, according to a counter-example given by Dehling (1983). In this paper, he constructed an example of a sequence $(X_k)_{k>0}$ with values in the unit ball of ℓ^2, such that $\alpha(\sigma(X_i, 1 \leq i < k), \sigma(X_i, i \geq k))$ converges to 0, and which cannot be approximated by independent random variables Y_k distributed as X_k in such a way that $\|X_k - Y_k\|_{\ell^2}$ converges to 0 in probability. This proves that in infinite dimensional Hilbert spaces the coefficients $\tau(\mathcal{M}, X)$ and $\alpha(\mathcal{M}, \sigma(X))$ cannot be compared.

We shall give in equation (7.16) of Remark 7 an alternative definition for $\tau(\mathcal{M}, X)$ than that given in (7.1). With this other definition, (7.6) is true for $\alpha(\mathcal{M}, \sigma(X))$ instead of $\beta(\mathcal{M}, \sigma(X))$. For this weaker coefficient, according to Dehling's example, the equality $\tau(\mathcal{M}, X) = \mathbb{E}(d(X, X^*))$ does not hold for any Polish space $\mathcal{X}$, although it holds for $\mathbb{R}$.

7.2. Examples

We proceed as in Examples 1, 2, 3 and 4 of Section 4. For φ, the case of causal functions of uniformly mixing sequences has been studied in Rio (1996).

Example 1: causal functions of stationary sequences. $(X_n)_{n\in\mathbb{Z}}$, $(X_n^*)_{n\geq 0}$ and $\mathcal{M}_i$ be defined as in Example 1 of Section 4. Then for $j_k > \cdots > j_1 \geq i$, we have both

$$\tau(\mathcal{M}_0, (X_{j_1}, \ldots, X_{j_k})) \leq \sum_{l=1}^{k} \|X_{j_l} - X_{j_l}^*\|_1 \tag{7.7}$$

$$\varphi(\mathcal{M}_0, (X_{j_1}, \ldots, X_{j_k})) \leq \sum_{l=1}^{k} \|\mathbb{E}(|X_{j_l} - X_{j_l}^*|\,|\mathcal{M}_0)\|_\infty\,. \tag{7.8}$$

Let $(\delta_i)_{i\geq 0}$ and $(\delta_i')_{i\geq 0}$ be two nonincreasing sequence such that $\|X_i - X_i^*\|_1 \leq \delta_i$ and $\|\mathbb{E}(|X_i - X_i^*|\,|\mathcal{M}_0)\|_\infty \leq \delta_i'$ respectively. Then $\tau_\infty(i) \leq \delta_i$ and $\varphi_\infty(i) \leq \delta_i'$. For instance, if $(\xi_i)_{i\in\mathbb{Z}}$ is iid and $X_n = \sum_{j\geq 0} a_j \xi_{n-j}$, we can take $\delta_i = 2\|\xi_0\|_1 \sum_{j\geq i} |a_j|$ and $\delta_i' = 2\|\xi_0\|_\infty \sum_{j\geq i} |a_j|$.

Example 2: iterative random functions. Let $(X_n)_{n\geq 0}$, $(X_n^*)_{n\geq 0}$ and $\mathcal{M}_i$ be defined as in Example 2 of Section 4. Then (7.7) and (7.8) hold. Let$(\delta_i)_{i\geq 0}$ and $(\delta_i')_{i\geq 0}$ be two nonincreasing sequences such that $\|X_i - X_i^*\|_1 \leq \delta_i$ and $\|\mathbb{E}(|X_i - X_i^*|\,|\mathcal{M}_0)\|_\infty \leq \delta_i'$ respectively. Then $\tau_\infty(i) \leq \delta_i$ and $\varphi_\infty(i) \leq \delta_i'$. Denote by $(X_n^x)_{n\geq 0}$ the chain starting from $X_0^x = x$. If $(d_i)_{i\geq 0}$ is some non increasing sequence such that $\|X_i^x - X_i^y\|_1 \leq d_i|x-y|$ then $\delta_i \leq 2\|X_0\|_1 d_i$ and $\delta_i' \leq 2\|X_0\|_\infty d_i$. If $\|F(x,\xi_0) - F(y,\xi_0)\|_\infty \leq \kappa|x-y|$ for some $\kappa < 1$, we can take $d_i = \kappa^i$. For instance, if $X_n = f(X_{n-1}) + \xi_n$ for some κ-lipshitz function f, then $\tau_\infty(i) \leq 2\|X_0\|_1\kappa^i$ and $\varphi_\infty(i) \leq 2\|X_0\|_\infty\kappa^i$.

Example 3: Markov kernels. Let $(X_n)_{n\in\mathbb{N}}$ be a stationary Markov chain with values in $\mathcal{X}$, with marginal distribution μ and transition kernel P satisfying Condition H of Example 3, Section 4.3. Then for $j_k > \cdots > j_1 \geq i$ and f in $\Lambda_1(\mathcal{X}^k)$, the function $\mathbb{E}(f(X_{j_1}, \ldots, X_{j_k})|X_{j_1=x})$ belongs to $\Lambda_{1+\kappa+\cdots+\kappa^{k-1}}(\mathcal{X})$ and consequently the function $f_{j_1,\ldots,j_k}(x) = \mathbb{E}(f(X_{j_1}, \ldots, X_{j_k})|X_0 = x)$ belongs to $\Lambda_{\kappa^i(1+\kappa+\cdots+\kappa^{k-1})}(\mathcal{X})$. One has that

$$\tau(\sigma(X_0), (X_{j_1}, \ldots, X_{j_k})) \leq \iint \sup_{f\in\Lambda_1(\mathcal{X}^k)} |f_{j_1,\ldots,j_k}(x) - f_{j_1,\ldots,j_k}(y)|\mu(dx)\mu(dy)$$

$$\varphi(\sigma(X_0), (X_{j_1}, \ldots, X_{j_k})) \leq \sup_{f\in\Lambda_1(\mathcal{X}^k)} \sup_{(x,y)\in\mathcal{X}^2} |f_{j_1,\ldots,j_k}(x) - f_{j_1,\ldots,j_k}(y)|\,.$$

Consequently, if X_0^* is an independent copy of X_0, we obtain the bounds

$$\tau(\sigma(X_0), (X_{j_1}, \ldots, X_{j_k})) \leq \kappa^i(1+\kappa+\cdots+\kappa^{k-1})\|X_0 - X_0^*\|_1 \tag{7.9}$$

$$\varphi(\sigma(X_0), (X_{j_1}, \ldots, X_{j_k})) \leq \kappa^i(1+\kappa+\cdots+\kappa^{k-1})\|X_0 - X_0^*\|_\infty\,. \tag{7.10}$$

We infer that $\tau_\infty(i) \leq 2\|X_0\|_1\kappa^i$ and $\varphi_\infty(i) \leq 2\|X_0\|_\infty\kappa^i$.

Example 4: Expanding maps. Let T be a map from [0, 1] to [0, 1] satisfying Conditions 1. 2. and 3. of Section 4.4 (see the application). Assume moreover that the density f_μ ot the invariant probability μ satisfies (4.8). Let $X_i = T^i$ and define P as in (4.9). We know from Section (4.4) that on $([0, 1], \mu)$, the sequence $(X_n, \dots, X_0)$ has the same distribution as $(Y_0, \dots, Y_n)$ where $(Y_i)_{i\geq 0}$ is the stationary Markov chain with Markov Kernel P. Consequently

$$\varphi(\sigma(X_j, j \geq i+k), (X_0, \dots, X_k)) = \varphi(\sigma(Y_0), (Y_i, \dots, Y_{i+k})) . \qquad (7.11)$$

To bound $\varphi(\sigma(Y_0), (Y_i, \dots, Y_{i+k}))$, the first step is to compute $\mathbb{E}(f(Y_0, \dots, Y_k)|Y_0 = x)$. As for P, define the operator Q_k by

$$\int_0^1 Q_k(f)(x)g(x)f_\mu(x)dx = \int_0^1 f(T^k(x), \dots, x)g(T^k(x))f_\mu(x)dx .$$

Clearly $\mathbb{E}(f(Y_0, \dots, Y_k)|Y_0 = x) = Q_k(f)(x)$ and by definition

$$\begin{aligned}\varphi(\sigma(Y_0), (Y_i, \dots, Y_{i+k})) &= \sup_{f\in\Lambda_1(\mathbb{R}^{k+1})} \|P^i \circ Q_k(f) - \mu(Q_k(f))\|_\infty \\ &= \sup_{f\in\Lambda_1(\mathbb{R}^{k+1})} \|(P^i - \mu) \circ (Q_k(f) - Q_k(f)(0))\,\|_\infty\end{aligned} \qquad (7.12)$$

Here, we use a recent result of Collet *et al.* (2002). Denote by $\Lambda_{L_1,\dots,L_n}$ the set of functions f from $\mathbb{R}^n$ to $\mathbb{R}$ such that

$$|f(x_1, \dots, x_n) - f(y_1, \dots, y_n)| \leq L_1|x_1 - y_1| + \cdots + L_n|x_n - y_n| . \qquad (7.13)$$

Adapting Lasota-Yorke's approach to higher dimension Collet *et al.* prove (page 312 line 6) that there exist $K > 0$ and $0 \leq \sigma < 1$ such that, for any f in $\Lambda_{L_1,\dots,L_{k+1}}$,

$$\|dQ_k(f)\| \leq K \sum_{i=0}^{k} \sigma^i L_{i+1} . \qquad (7.14)$$

Applying (4.10), we infer from (7.12) and (7.14) that

$$\begin{aligned}\varphi(\sigma(Y_0), (Y_i, \dots, Y_{i+k})) &\leq C\rho^i \|Q_k(f) - Q_k(f)(0)\|_v \\ &\leq C\rho^i 2\|dQ_k(f)\| \leq C\rho^i 2K \sum_{j=0}^{k} \sigma^j .\end{aligned}$$

Moreover, according to (7.11), the same bound holds for $\varphi(\sigma(X_j, j \geq i+k), (X_0, \dots, X_k))$. For the Markov chain $(Y_i)_{i\geq 0}$ and the σ-algebras $\mathcal{M}_i = \sigma(Y_j, j \leq i)$ we obtain from (7.14) that

$$\varphi_\infty(i) \leq \Big(2\,CK \sum_{j\geq 0} \sigma^j\Big)\rho^i .$$

7.3. Bennett-type inequalities and Functional LIL

In this section, we recall some recent results for τ-dependent sequences obtained in Dedecker and Prieur (2003). The first Proposition extends Bennett's inequality for independent sequences to the case of τ-dependent sequences. For any positive integer q, we obtain an upper bound involving two terms: the first one is the classical Bennett's bound at level λ for a sum $\sum_n$ of independent variables ξ_i such that $\mathrm{Var}(\sum_n) = v_q$ and $\|\xi_i\|_\infty \leq qM$, and the second one is equal to $n\lambda^{-1}\tau_q(q+1)$.

Proposition 6. *Let $(X_i)_{i>0}$ be a sequence of real-valued random variables bounded by M, and $\mathcal{M}_i = \sigma(X_k, 1 \leq k \leq i)$. Let $S_k = \sum_{i=1}^k (X_i - \mathbb{E}(X_i))$ and $\overline{S}_n = \max_{1\leq k\leq n} |S_k|$. Let q be some positive integer, v_q some nonnegative number such that*

$$v_q \geq \|X_{q[n/q]+1} + \cdots + X_n\|_2^2 + \sum_{i=1}^{[n/q]} \|X_{(i-1)q+1} + \cdots + X_{iq}\|_2^2 .$$

and h the function defined by $h(x) = (1+x)\ln(1+x) - x$.

1. *For any positive λ, $\mathbb{P}(|S_n| \geq 3\lambda) \leq 4\exp\Big(-\frac{v_q}{(qM)^2} h\Big(\frac{\lambda q M}{v_q}\Big)\Big) + \frac{n}{\lambda}\tau_q(q+1)$.*
2. *For any $\lambda \geq Mq$,*

$$\mathbb{P}(\overline{S}_n \geq (\mathbb{1}_{q>1} + 3)\lambda) \leq 4\exp\Big(-\frac{v_q}{(qM)^2} h\Big(\frac{\lambda q M}{v_q}\Big)\Big) + \frac{n}{\lambda}\tau_q(q+1) .$$

Starting from the second inequality and using the coupling property of $\tau(\mathcal{M}, X)$ for real-valued random variables, we can prove a functional law of the iterated logarithm. We need some preliminary notations. Let $(X_i)_{i\in\mathbb{Z}}$ be a stationary sequence of real-valued random variables. let $Q = Q_{X_0}$ be defined as in Proposition 2 and let G be the inverse of $x \to \int_0^x Q(u)du$. Let $\mathcal{S}$ be the subset of $C([0,1])$ consisting of all absolutely continuous functions with respect to the Lebesgue measure such that $h(0) = 0$ and $\int_0^1 (h'(t))^2 dt \leq 1$.

Theorem 1. *Let $(X_i)_{i\in\mathbb{Z}}$ be a stationary sequence of zero-mean square integrable random variables, and $\mathcal{M}_i = \sigma(X_j, j \leq i)$. Let $S_n = X_1 + \cdots + X_n$ and define the partial sum process $S_n(t) = S_{[nt]} + (nt - [nt])X_{[nt]+1}$. If*

$$\sum_{k=1}^{\infty} \int_0^{\tau_\infty(k)} Q \circ G(u)\, du < \infty \tag{7.15}$$

then $\mathrm{Var}(S_n)$ converges to $\sigma^2 = \sum_{k\in\mathbb{Z}} \mathrm{Cov}(X_0, X_k)$. If furthermore $\sigma > 0$ then the process $\{\sigma^{-1}(2n\ln\ln n)^{-1/2} S_n(t) \; : \; t \in [0,1]\}$ is almost surely relatively compact in $C([0,1])$ with limit set $\mathcal{S}$.

Remark 7. Proposition 6 and Theorem 1 remain valid when replacing the definition of $\tau(\mathcal{M}, X)$ given in (7.1) by the weaker one

$$\tau(\mathcal{M}, X) = \sup_{f\in\Lambda_1(\mathcal{X})} \Big\| \sup\Big\{ \Big| \int g\circ f(x)\mathbb{P}_{X|\mathcal{M}}(dx) - \int g\circ f(x)\mathbb{P}_X(dx)\Big|, g\in\Lambda_1(\mathbb{R})\Big\}\Big\|_1 . \quad (7.16)$$

The coefficient $\tau_\infty(i)$ obtained from (7.16) instead of (7.1) is comparable to the usual strong mixing coefficient $\alpha'_\infty(i) = \alpha(\mathcal{M}_0, \sigma(X_k, k \geq i))$. In particular, keeping the same notations as in Theorem 1, we have that

$$\int_0^{\tau_\infty(k)} Q\circ G(u)\,du \leq 2\int_0^{\alpha'_\infty(i)} Q^2(u)\,du ,$$

so that condition (7.15) is weaker than Rio's condition (1995) for the functional LIL.

7.4. A concentration inequality for Lipschitz functions.

Recall that if $(\mathcal{X}, d)$ is a Polish space, we put the distance d_1 on the product space $\mathcal{X}^n$: $d_1(x, y) = d(x_1, y_1) + \cdots + d(x_n, y_n)$. The space $\Lambda_1(\mathcal{X}^n)$ is the space of 1-Lipschitz functions from $\mathcal{X}^n$ to $\mathbb{R}$ with respect to d_1.

The following inequality is a straightforward consequence of Theorem 1 in Rio (2000b).

Theorem 2. *Let $(X_1, \ldots, X_n)$ be a sequence of random variables with values in a Polish space $(\mathcal{X}, d)$ and $\mathcal{M}_i = \sigma(X_1, \ldots, X_i)$. Let $\Delta_i = \inf\{2\|d(X_i, x)\|_\infty, x\in \mathcal{X}\}$ and define*

$$B_n = \Delta_n \quad \text{and for } 1\leq i < n, \quad B_i = \Delta_i + 2\varphi(\mathcal{M}_i, (X_{i+1}, \ldots, X_n)) .$$

For any f in $\Lambda_1(\mathcal{X}^n)$, we have that

$$\mathbb{P}(f(X_1, \ldots, X_n) - \mathbb{E}(f(X_1, \ldots, X_n)) \geq x) \leq \exp\Big(\frac{-2x^2}{B_1^2 + \cdots + B_n^2}\Big) .$$

This theorem applies to the examples given in Section 7.2. Recall that the set $\Lambda_{L_1,\ldots,L_n}$ has been defined in (7.13).

Examples 1 and 2: causal functions of stationary sequences and iterated random functions. Keeping the same notations as in Examples 1 and 2 of Section 7.2, let $\delta'_i = \|\mathbb{E}(|X_i - X_i^*|\,|\mathcal{M}_0)\|_\infty$ and define

$$M_n = L_n\Delta_0 \quad \text{and for } 1\leq i < n, \quad M_i = L_i\Delta_0 + 2(L_{i+1}\delta'_1 + \cdots + L_n\delta'_{n-i}) .$$

For f any function f belonging to $\Lambda_{L_1,\ldots,L_n}$, we have

$$\mathbb{P}(f(X_1, \ldots, X_n) - \mathbb{E}(f(X_1, \ldots, X_n)) \geq x) \leq \exp\Big(\frac{-2x^2}{M_1^2 + \cdots + M_n^2}\Big) . \quad (7.17)$$

Example 3: Markov kernels. Let $(X_n)_{n\in\mathbb{N}}$ be a stationary Markov chain with values in $\mathcal{X}$, with marginal distribution μ and transition kernel P satisfying Condition H of Example 3, Section 4.3. For any function f belonging to $\Lambda_{L_1,\dots,L_n}$, the bound (7.17) holds with

$$M_n = L_n\Delta_0 \quad \text{and for } 1 \le i < n, \quad M_i = \Delta_0(L_i + 2L_{i+1}\kappa + \cdots + 2L_n\kappa^{n-i})\,.$$

Example 4: Expanding maps. Let T be an expending map from [0, 1] to [0, 1] satisfying the assumptions of Section 7.2. Let $X_i = T^i$ and Y_i be the associated Markov chain (cf. Section 7.2). Starting from (7.14) and (4.10), we infer that the bound (7.17) holds for $f(Y_1, \dots, Y_n)$ with

$$M_n = L_n\Delta_0 \quad \text{and for } 1 \le i < n, \quad M_i = \Delta_0 L_i + 4CK\rho(L_{i+1} + \cdots + L_n\sigma^{n-i-1})\,.$$

Since $(X_1, \dots, X_n)$ has the same distribution as $(Y_n, \dots, Y_1)$, we obtain the bound (7.17) for $f(X_1, \dots, X_n)$ with

$$\begin{aligned} M_n &= L_1\Delta_0 \quad \text{and for } 1 \le i < n, \\ M_i &= \Delta_0 L_{n-i+1} + 4CK\rho(L_{n-i} + \cdots + L_1\sigma^{n-i+1})\,. \end{aligned}$$

Remark 8. Assume that (7.17) holds for $M_i = \delta_0 L_i + \delta_i L_{i+1} + \cdots + \delta_{n-i} L_n$ (which is the case in the four examples studied above) and let $C_n = \delta_0 + \cdots + \delta_{n-1}$. Applying Cauchy-Schwarz's inequality, we obtain the bound $M_i^2 \le C_n \sum_{j=i}^n \delta_{j-i} L_i^2$, and consequently $\sum_{i=1}^n M_i^2 \le C_n^2 \sum_{i=1}^n L_i^2$. Hence, (7.17) yield the upper bound

$$\mathbb{P}(f(X_1, \dots, X_n) - \mathbb{E}(f(X_1, \dots, X_n)) \ge x) \le \exp\Big(\frac{-2x^2}{C_n^2(L_1^2 + \cdots + L_n^2)}\Big)\,. \tag{7.18}$$

For expanding maps (Example 4 above) (7.18) has been proved by Collet *et al* (2002).

7.5. A Berry-Esseen inequality

The following Berry-Esseen bound is due to Rio (1996), Theorem 1.

Theorem 3. *Let $(X_i)_{i\in\mathbb{Z}}$ be a stationary sequence of real-valued bounded and centered random variables and $\mathcal{M}_i = \sigma(X_j, j \le i)$. Let $S_n = X_1 + \cdots + X_n$ and $\sigma_n = \|S_n\|_2$. If $\limsup_{n\to\infty} \sigma_n = \infty$ and*

$$\sum_{n>0} n\varphi_3(n) < \infty\,, \tag{7.19}$$

then σ_n^2 converges to $\sigma^2 = \sum_{k\in\mathbb{Z}} \mathrm{Cov}(X_0, X_k)$. Moreover $\sigma > 0$ and

$$\sup_{x\in\mathbb{R}}\Big|\mathbb{P}(S_n \le x\sigma_n) - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^x \exp(-x^2/2)dx\Big| \le \frac{C}{\sqrt{n}}\,,$$

where C depends only on $\|X_0\|_\infty$, $(\varphi_3(k))_{k\ge 0}$ and σ.

Remark 9. In fact, in Rio's theorem, the condition is $\sum_{n>0} n\varphi'_3(n) < \infty$, where

$$\varphi'_3(i) = \sup_{p+i\leq j_1<j_2<j_3} \sup\{\|\mathbb{E}(f(X_{j_1}, X_{j_2}, X_{j_3})|\mathcal{M}_p) - \mathbb{E}(f(X_{j_1}, X_{j_2}, X_{j_3}))\|_\infty, f \in \Lambda'_1(\mathbb{R}^3)\}$$

and $\Lambda'_1(\mathbb{R}^3)$ is the set of f from $\mathbb{R}^3$ to $[0, 1]$ such that $|f(x)-f(y)| \leq \max_{1\leq i\leq 3} |x_i - y_i|$. Under the assumptions of Theorem 3, we have $\varphi'_3(i)/3 \leq \varphi_3(i) \leq (1 \vee 2\|X_0\|_\infty)\varphi'_3(i)$, so that (7.19) is equivalent to Rio's condition.

Acknowledgements. At the beginning of this paper, there is a discussion with Magda Peligrad on covariance inequalities. The first version of Proposition 1 has been established with Magda Peligrad after a day of work. The original proof was much longer and not as clear as the one presented here. Sana Louhichi pointed out the reference to Yu (1993), which helped us greatly in the writing of the present proof. The fact that the definition of τ can be extended to Polish spaces without losing the coupling property became clear after a fruitful discussion with Emmanuel Rio. The idea that one can compute φ_∞ for iterates of expanding maps first occured to us after a conference given by Bernard Schmitt in Versailles. Finally, we thank a referre for having pointed out many gaps in the first version of this paper.

References

1. Andrews, D.W.K.: Nonstrong mixing autoregressive processes. J. Appl. Probab. **21**, 930–934 (1984)
2. Barbour, A.D., Gerrard, R.M., Reinert, G.: Iterates of expanding maps. Probab. Theory Relat. Fields **116**, 151–180 (2000)
3. Barron, A., Birgé, L., Massart, P.: Risk bounds for model selection via penalization. Probab. Theory Relat. Fields **113**, 301–413 (1999)
4. Berbee, H.C.P.: Random walks with stationary increments and renewal theory. Math. Cent. Tracts. Amsterdam, 1979
5. Billingsley, P.: (1968) Convergence of probability measures. Wiley, New-York
6. Borovkova, S., Burton, R., Dehling, H.: Limit theorems for functionals of mixing processes with application to U-statistics and dimension estimation. Trans. Amer. Math. Soc. **353**, 4261–4318 (2001)
7. Bradley, R.C.: Approximations theorems for strongly mixing random variables. Michigan Math. J. **30**, 69–81 (1983)
8. Bradley, R.C.: Introduction to Strong Mixing Conditions, Volume 1. Technical Report, Department of Mathematics, I. U. Bloomington, 2002
9. Broise, A.: . Transformations dilatantes de l'intervalle et théorèmes limites. Études spectrales d'opérateurs de transfert et applications. Astérisque **238**, 1–109 (1996)
10. Castaing, C., Raynaud de Fitte, P., Valadier, M.: Young Measures on Topological Spaces. With Applications in Control Theory and Probability Theory. Kluwer Academic Publishers, Dordrecht, 2004
11. Collet, P., Martinez, S., Schmitt, B.: Exponential inequalities for dynamical measures of expanding maps of the interval. Probab. Theory. Relat. Fields **123**, 301–322 (2002)
12. Deddens, J., Peligrad, M., Yang, T.: (1987) On strong consistency of kernel estimators under dependence assumptions. Mathematical statistics and probability theory Vol. B 33–41, (Bad Tatzmannsdorf, 1986) Reidel, Dordrecht
13. Dedecker, J., Doukhan, P.: . A new covariance inequality and applications. Stochastic Process. Appl. **106**, 63–80 (2003)

14. Dedecker, J., Prieur, C.: (2003) Coupling for τ-dependent sequences and applications. Preprint LSP 2003-02. Université Paul Sabatier, Toulouse. Accepted for publication in J. Theoret. Probab.
15. Dedecker, J., Prieur, C.: Couplage pour la distance minimale. *C. R. Acad. Sci. Paris, Ser I* **338**, 805–808 (2004)
16. Dedecker, J., Rio, E.: On the functional central limit theorem for stationary processes. Ann. Inst. H. Poincaré Probab. Statist. **36**, 1–34 (2000)
17. Dehling, H.: A note on a theorem of Berkes and Philipp. Z. Wahrsch. Verw. Gebiete **62**, 39–42 (1983)
18. Delyon, B.: (1990) Limit theorems for mixing processes. Tech. Report 546 IRISA, Rennes I
19. DeVore, R.A., Lorentz, G.G.: Constructive approximation. Springer, Berlin, Heidelberg, New York, 1993
20. Diaconis, P., Freedman, D.: Iterated random functions. SIAM Rev. **41**, 45–76 (1999)
21. Dobrushin, R.L.: Prescribing a system of random variables by conditional distributions. Theory Probab. Appl. **15**, 458–486 (1970)
22. Doukhan, P.: Mixing: properties and examples. Lecture Notes in Statist. **85**, Springer-Verlag, 1994
23. Doukhan, P., Louhichi, S.: A new weak dependence condition and applications to moment inequalities. Stochastic Process. Appl. **84**, 313–342 (1999)
24. Dudley, R.M.: Real analysis and probability. Wadworsth Inc., Belmont, California, 1989
25. Hofbauer, F., Keller, G.: Ergodic properties of invariant measures for piecewise monotonic transformations. Math. Z. **180**, 119–140 (1982)
26. Ibragimov, I.A.: Some limit theorems for stationary processes. Theory Probab. Appl. **7**, 349–382 (1962)
27. Lasota, A., Yorke, J.A.: On the existence of invariant measures for piecewise monotonic transformations. Trans. Amer. Math. Soc. **186**, 481–488 (1974)
28. Ledoux, M., Talagrand, M.: Probability in Banach spaces. Springer, New-York, 1991
29. Major, P.: On the invariance principle for sums of identically distributed random variables. J. Multivariate Anal. **8**, 487–517 (1978)
30. Merlevède, F., Peligrad, M.: (2002) On the coupling of dependent random variables and applications. Empirical process techniques for dependent data. 171–193 Birkhäuser
31. Morita, T.: Local limit theorem and distribution of periodic orbits of Lasota-Yorke transformations with infinite Markov partition. J. Math. Soc. Japan **46**, 309–343 (1994)
32. Peligrad, M.: Some remarks on coupling of dependent random variables. Stat. Prob. Letters **60**, 201–209 (2002)
33. Philipp, W., Stout, W.: Almost sure invariance principles for partial sums of weakly dependent random variables. Mem. Amer. Math. Soc **161**, (1975)
34. Prieur, C.: Density Estimation For One-Dimensional Dynamical Systems. ESAIM, Probab. & Statist. www.emath.fr/ps **5**, 51–76 (2001)
35. Rachev, S.T., Rüschendorf, L.: Mass Transportation Problems, Volume 1: Theory. Springer. Berlin, Heidelberg, New-York, 1998
36. Rio, E.: The functional law of the iterated logarithm for stationary strongly mixing sequences. Ann. Probab. **23**, 1188–1203 (1995)
37. Rio, E.: Sur le théorème de Berry-Esseen pour les suites faiblement dépendantes. Probab. Theory Relat. Fields **104**, 255–282 (1996)
38. Rio, E.: Théorie asymptotique des processus aléatoires faiblement dépendants. Collection Mathématiques & Applications **31**, Springer, Berlin, (2000a)
39. Rio, E.: Inégalités de Hoeffding pour les fonctions lipschitziennes de suites dépendantes. C. R. Acad. Sci. Paris Série I **330**, 905–908 (2000b)

40. Rosenblatt, M.: . A central limit theorem and a strong mixing condition. Proc. Nat. Acad. Sci. U. S. A. **42**, 43–47 (1956)
41. Rozanov, Y.A., Volkonskii, V.A.: Some limit theorems for random functions I. Theory Probab. Appl. **4**, 178–197 (1959)
42. van der Vaart, A.W., Wellner, J.A.: Weak convergence and empirical processes. Springer, Berlin, 1996
43. Viennet, G.: Inequalities for absolutely regular sequences: application to density estimation. Probab. Theory Relat. Fields **107**, 467–492 (1997)
44. Yu, H.: A Glivenko-Cantelli lemma and weak convergence for empirical processes of associated sequences. Probab. Theory Relat. Fields **95**, 357–370 (1993)