Recent results on weak dependence for causal sequences. Statistical applications to dynamical systems.

Clémentine PRIEUR

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

Summary. Many commonly used models for real-world phenomena do not satisfy classical mixing conditions. Modifications for these mixing conditions have therefore been introduced by number of authors. The coefficients under study in this paper measure the dependence between a real-valued random variable X and a $\sigma-$algebra $\mathcal{M}$ by considering the gap between the conditional distribution function of X given $\mathcal{M}$ and the distribution of X. Although weaker than usual mixing coefficients, they allow deriving limit theorems and statistical applications.

1 Introduction

There exists a wide literature on limit theorems under various classical mixing conditions such as strong mixing condition ($\alpha-$mixing), absolute regularity ($\beta-$mixing), or $\Phi-$mixing. For recent and complete results on the properties of these coefficients, we refer to the monographs by Doukhan [14], Rio [30] and Bradley [5]. However, many commonly used models do not satisfy these mixing conditions. For example, Andrews [1] proved that if $(\varepsilon_i)_{i\geq 1}$ is i.i.d. with marginal $\mathcal{B}\,(1/2)$, then the stationary solution $(X_i)_{i\geq 0}$ of the equation

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

is not $\alpha-$mixing. Many authors have therefore introduced modifications of these various mixing coefficients. 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 $\sigma-$algebra of $\mathcal{A}$. Let us first recall the definition of the usual mixing coefficients between $\mathcal{M}$ and $\sigma(X)$, introduced respectively by Rosenblatt [32], Volkonskii and Rozanov [33] and Ibragimov [19]:

$$\begin{aligned}
\alpha(\mathcal{M}, \sigma(X)) &= \frac{1}{2} \sup_{A\in\mathcal{B}(\mathbb{R})} \|\mathbb{P}_{X|\mathcal{M}}(A) - \mathbb{P}_X(A)\|_1 \\
\beta(\mathcal{M}, \sigma(X)) &= \| \sup_{A\in\mathcal{B}(\mathbb{R})} |\mathbb{P}_{X|\mathcal{M}}(A) - \mathbb{P}_X(A)|\, \|_1 \\
\phi(\mathcal{M}, \sigma(X)) &= \sup_{A\in\mathcal{B}(\mathbb{R})} \|\mathbb{P}_{X|\mathcal{M}}(A) - \mathbb{P}_X(A)\|_\infty \,.
\end{aligned}$$

We refer to the book of Doukhan [14] for the properties of these coefficients. However, to derive limit theorems for random processes modeling real-world phenomena which do not satisfy classical mixing conditions, it is useful to weaken the definitions above. Of course the goal is to catch more examples without losing too much of the nice properties of classical mixing processes. Following an idea of Rosenblatt [32], we can look at what happens when considering coarser sets than $\mathcal{M}$ or $\mathcal{B}(\mathbb{R})$. Changing $\mathcal{M}$ leads to coefficients which behave quite differently from the usual mixing coefficients. Let us cite for instance the work of Doukhan and Louhichi [16]. One different approach is to consider instead of $\mathcal{B}(\mathbb{R})$ the coarser set $\{]-\infty,t], t\in\mathbb{R}\}$. This has been done by Rio [30] and later by Peligrad [25] for the strong mixing coefficient. Dedecker and Prieur [11] have broadened and systematically developed this approach. The coefficients then obtained measure the difference between the conditional distribution $F_{X|\mathcal{M}}$ of $\mathbb{P}_{X|\mathcal{M}}$ and the distribution function F_X of $\mathbb{P}_X$. We define below the four dependence coefficients introduced in [11].

Definition 1 :

$\tau(\mathcal{M},X)=\int\|F_{X|\mathcal{M}}(t)-F_X(t)\|_1 dt$

$\alpha(\mathcal{M},X)=\sup_{t\in\mathbb{R}}\|F_{X|\mathcal{M}}(t)-F_X(t)\|_1$

$\beta(\mathcal{M},X)=\|\sup_{t\in\mathbb{R}}|F_{X|\mathcal{M}}(t)-F_X(t)|\,\|_1$

$\phi(\mathcal{M},X)=\sup_{t\in\mathbb{R}}\|F_{X|\mathcal{M}}(t)-F_X(t)\|_\infty$.

The coefficient $\alpha(\mathcal{M},X)$ was first introduced by Rio ([30], equation 1.10 c). It has then been used by Peligrad [25], while $\tau(\mathcal{M},X)$ was introduced in the current form by Dedecker and Prieur [10]. However, Rüschendorf first introduced the coefficient τ, in a "dual" form (see equation 10 in [34]). These coefficients are smaller than their corresponding mixing coefficients, but they are in many situations easier to compute. It is worth of interest to notice that the coefficients $\tau(\mathcal{M},X)$, $\alpha(\mathcal{M},X)$, $\beta(\mathcal{M},X)$ and $\phi(\mathcal{M},X)$, as other measures of dependence, can be defined as supremum over some family of functions. For this, we first need to introduce some definitions and notations.

Definition 2 : 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 $\|\mu\|=|\mu|(\mathbb{R})$.

We then get the result written in Lemma 1 below (compare to Theorem 4.4 in Bradley [5] for usual mixing coefficients).

Lemma 1 *[11] 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 $\|dh\|\leq 1$. 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. $\phi(\mathcal{M},X)=\sup\{\|\mathbb{E}(f(X)|\mathcal{M})-\mathbb{E}(f(X))\|_\infty, f\in BV_1\}$.

This paper is a survey of recent results on dependence in the causal frame. For the non-causal frame, we refer e.g. to Doukhan and al or Ango Nze and Doukhan [15, 2]. In Section 2 we state useful tools in limit theory for dependent sequences. In Section 3 we give a way to compare the different coefficients of dependence. Section 4 is devoted to statistical applications. In Section 5, we give exponential inequalites. We then see in Section 6 how to extend some of the coefficients of Definition 1 to the multidimensionnal case. To conclude, we give in Section 7 some results for particular classes of dynamical systems on $[0,1]$.

2 Main tools for statistical applications

2.1 Covariance inequalities

For statistical applications, it was made clear by Viennet [35] that covariance inequalities in the style of Delyon [13] are more efficient than the usual covariance inequalities. Viennet's result applies to linear estimators and provides optimal results for the mean integrated square error.

In Proposition 1 below, we give two covariance inequalities. Inequality (1) is a weak version of that of Delyon [13] in which appear two random variables $b_1(\sigma(Y),\sigma(X))$ and $b_2(\sigma(X),\sigma(Y))$ each having mean $\beta(\sigma(Y),\sigma(X))$. Inequality (2) is a weak version of that of Peligrad [24], where the dependence coefficients are $\phi(\sigma(Y),\sigma(X))$ and $\phi(\sigma(X),\sigma(Y))$.

Proposition 1 *[9] Let X and Y be two real-valued random variables on the probability space $(\Omega,\mathcal{A},\mathbb{P})$. Let $F_{X|Y}=\{t\to P_{X|Y}(]-\infty,t])\}$ be a distribution function of X given Y and let F_X be the distribution function of X. Define the random variable $b(\sigma(Y),X)=\sup_{x\in\mathbb{R}}|F_{X|Y}(x)-F_X(x)|$. For any conjugate exponents p and q, we have the inequalities*

$$|\mathrm{Cov}(Y,X)| \le 2\{\mathbb{E}(|X|^p b(\sigma(X),Y))\}^{\frac{1}{p}}\{\mathbb{E}(|Y|^q b(\sigma(Y),X))\}^{\frac{1}{q}} \tag{1}$$

$$\le 2\phi(\sigma(X),Y)^{\frac{1}{p}}\phi(\sigma(Y),X)^{\frac{1}{q}}\|X\|_p\|Y\|_q\,. \tag{2}$$

In order to derive the MISE of the unknown marginal density of a stationary sequence (see Section 4.2), we need Corollary 1 below.

Corollary 1 *[9] Let f_1,f_2,g_1,g_2 be four increasing functions, and let $f=f_1-f_2$ and $g=g_1-g_2$. For any random variable Z, let $\varDelta_p(Z)=\inf_{a\in\mathbb{R}}\|Z-a\|_p$ and $\varDelta_{p,\sigma(X),Y}(Z)=\inf_{a\in\mathbb{R}}(\mathbb{E}(|Z-a|^p b(\sigma(X),Y)))^{1/p}$. For any conjugate exponents p and q, we have the inequalities*

$$|\mathrm{Cov}(g(Y),f(X))| \le 2\big\{\varDelta_{p,\sigma(X),Y}(f_1(X))+\varDelta_{p,\sigma(X),Y}(f_2(X))\big\} \big\{\varDelta_{q,\sigma(Y),X}(g_1(Y))+\varDelta_{q,\sigma(Y),X}(g_2(Y))\big\}\,,$$

$$|\mathrm{Cov}(g(Y),f(X))| \le 2\phi(\sigma(X),Y)^{\frac{1}{p}}\phi(\sigma(Y),X)^{\frac{1}{q}}\ \big\{\varDelta_p(f_1(X))\ +\ \varDelta_p(f_2(X))\big\} \big\{\varDelta_q(g_1(Y))+\varDelta_q(g_2(Y))\big\}\,.$$

In particular, if μ is a signed measure with total variation $\|\mu\|$ and $f(x)=\mu(]-\infty,x])$, we have

$$|\mathrm{Cov}(Y,f(X))| \le \|\mu\|\mathbb{E}(|Y|b(\sigma(Y),X)) \le \phi(\sigma(Y),X)\|\mu\|\,\|Y\|_1\,. \tag{3}$$

Two different proofs of inequality (3) above can be found in [11] and [9].

2.2 Coupling

Coupling is another popular and useful method to obtain limit theorems for sequences of dependent random variables. The coupling result stated in this section just concerns real valued sequences. Thanks to conditional quantile transformation (see Major [21]), we get an explicit formula for the coupled variable. A more general result in higher dimension has been stated by Rüschendorf [34] (see Section 6.2).

Lemma 2 *[34, 10] Let $(\Omega,\mathcal{A},\mathbb{P})$ be a probability space, X an integrable real-valued random variable, and $\mathcal{M}$ a σ-algebra of $\mathcal{A}$. Assume that there exists a random variable δ uniformly distributed over $[0,1]$, independent of the σ-algebra generated by X and $\mathcal{M}$. Then there exists a random variable X^*, measurable with respect to $\mathcal{M}\vee\sigma(X)\vee\sigma(\delta)$, independent of $\mathcal{M}$ and distributed as X, such that*

$$\|X - X^*\|_1 = \tau(\mathcal{M}, X)\,. \tag{4}$$

Let $U = F_{X|\mathcal{M}}(X-0) + \delta\left(F_{X|\mathcal{M}}(X) - F_{X|\mathcal{M}}(X-0)\right)$, *where* $F_{X|\mathcal{M}}(t-0) = \sup_{s<t} F_{X|\mathcal{M}}(s)$. *The random variable* U *is uniformly distributed over* $[0,1]$ *and independent of* $\mathcal{M}$. *An explicit solution for the coupling is then given by* $X^* = F_X^{-1}(U)$.

3 Comparison of coefficients

The four coefficients introduced in Definition 1 can be compared to each other. It is the purpose of Proposition 2 below. Comparing coefficients to each other is useful for example when deriving upper bounds for $\tau(\mathcal{M}, X)$, $\alpha(\mathcal{M}, X)$, $\beta(\mathcal{M}, X)$ and $\phi(\mathcal{M}, X)$.

Proposition 2 *[11] 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* $\alpha(\mathcal{M}, X) \leq \beta(\mathcal{M}, X) \leq \phi(\mathcal{M}, X)$.
2. *Let* Q_X *be the generalized inverse of the tail function* $t \to \mathbb{P}(|X| > t)$*: if* $u \in]0,1[$, $Q_X(u) = \inf\{t \in \mathbb{R} : \mathbb{P}(|X| > t) \leq u\}$. *We have the inequality*

$$\tau(\mathcal{M}, X) \leq 2 \int_0^{\alpha(\mathcal{M}, X)} Q_X(u)du\,.$$

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*

$$\beta(\mathcal{M}, X) \leq \frac{2\tau(\mathcal{M}, X)}{g^{-1}(\tau(\mathcal{M}, X))}\,. \tag{5}$$

In particular, if F *is Hölder, that is there exist* $C > 0$ *and* $\alpha \in]0,1]$ *such that for all* (x, y) $|F(x) - F(y)| \leq C|x-y|^\alpha$, *then*

$$\beta(\mathcal{M}, X) \leq 2C^{1/(\alpha+1)} \left(\tau(\mathcal{M}, X)\right)^{\alpha/(\alpha+1)}\,.$$

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

$$\beta(\mathcal{M}, X) \leq 2\sqrt{K\tau(\mathcal{M}, X)}\,. \tag{6}$$

4 Application to mean integrated square error

We deal in this section with the problem of estimating the unknown marginal density f from the observations $(X_1, \ldots, X_n)$ of a stationary sequence $(X_i)_{i\geq 0}$. There exist many works on density estimation under various mixing conditions. The results of Proposition 3 are comparable to results of Mokkadem (see Theorem 1.2 in the book of Rio [30], and also [22]). The ones of Proposition 4 are to be compared to Theorem 1.3 (a) in [30]. These results of Mokkadem and Rio are obtained for the strong mixing coefficients $\alpha\left(\sigma(X_0), \sigma(X_i)\right)$. Viennet [35] proved, under the minimal assumption on the $\beta-$mixing coefficients

$$\sum_{k>0} \beta(\sigma(X_0), \sigma(X_k)) < +\infty\,, \tag{7}$$

that the mean integrated square error (MISE) is of the same order than in the i.i.d. case. Let us now state the results we obtain when working with our weaker coefficient β defined in Definition 1. We first define the coefficients $\tau(i)$, $\alpha(i)$, $\beta(i)$, and $\phi(i)$ related to a sequence of real-valued random variables.

Definition 3 : 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

$$\tau(i) = \sup_{k\geq 0} \tau(\mathcal{M}_k, X_{i+k})\,. \tag{8}$$

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

According to Definition 3 above and to the stationarity of $(X_i)_{i\geq 0}$, we let $\beta(i) = \beta(\sigma(X_0), X_i)$. Both Propositions 3 and 4 give upper bounds for the variance of estimators of the marginal density f.

Proposition 3 *[11] 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*

$$Y_{k,n} = h^{-1}K(h^{-1}(x - X_k)) \quad \textit{and} \quad f_n(x) = \frac{1}{n}\sum_{k=1}^{n} Y_{k,n}\,. \tag{9}$$

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 *[11] 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*

$$Y_{j,n} = \frac{1}{n}\sum_{k=1}^{n} \varphi_j(X_k) \quad \textit{and} \quad f_n = \sum_{j=1}^{m} Y_{j,n}\varphi_j\,. \tag{10}$$

The following inequality holds

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

These upper bounds allow deriving some quite sharp rates of convergence for some estimators of the unknown density f. We can write indeed $f_n(x) - f(x) = (f_n(x) - \mathbb{E}f_n(x)) + (\mathbb{E}f_n(x) - f(x))$. The bias term $\mathcal{BIAS}_n(x) = \mathbb{E}f_n(x) - f(x)$ does not depend on the dependence properties of the stationary sequence $(X_i)_{i\geq 0}$, but only on the regularity of the marginal density f. We refer to Prieur [28] page 63 for a detailed study of this term.

For kernel density estimators, optimal results can be obtained for the MISE under the condition

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

as far as we assume that the kernel K is BV and Lebesgue integrable. This condition is weaker than condition (7). For projection estimators, it does not work in the general case. We have to assume moreover that the basis is well localized, because the variance inequality of Proposition 4 is less precise than that of Viennet. It is due to the fact that the covariance inequality (3) of Corollary 1 is not symetric contrary to Delyon's covariance inequality [13] used in Viennet [35]. We refer to [11] for more detailed results as far as for the proofs.

5 Exponential inequalities

In this section, we state Hoeffding-type inequalities (Propositions 5 and 6) for partial sums under $\phi-$mixing conditions. The coefficients $\phi(k)$ are defined as in Definition 3.

Proposition 5 *[11] Let $(X_i)_{i\geq 0}$ be a sequence of centered and square integrable 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

$$\|S_n(h) - \mathbb{E}(S_n(h))\|_p \leq \Big(2p \sum_{i=1}^{n} b_{i,n}\Big)^{1/2} \leq \|dh\| \Big(2p \sum_{k=0}^{n-1} (n-k)\phi(k)\Big)^{1/2} . \qquad (11)$$

We also have that

$$\mathbb{P}(|S_n(h) - \mathbb{E}(S_n(h))| > x) \leq e^{1/e} \exp\left(\frac{-x^2}{4e\|dh\|^2 \sum_{k=0}^{n-1}(n-k)\phi(k)}\right) . \qquad (12)$$

Proposition 6 *Let $(X_i)_{i\geq 0}$ be a sequence of centered and square integrable random variables and $\mathcal{M}_i = \sigma(X_j, 1 \leq j \leq i)$. For any BV function h, define $S_n(h) = \sum_{i=1}^{n} h(X_i)$. For any $p \geq 2$ we have the inequality*

$$\mathbb{P}(|S_n(h) - \mathbb{E}(S_n(h))| > x) \leq 2\exp\left(\frac{-x^2}{2\|dh\|^2 \sum_{i=1}^{n}\big(1 + 2\sum_{k=1}^{n-i+1} \phi(k)\big)^2}\right) . \qquad (13)$$

Inequalities (12) and (13) are of the same type as soon as $\sum_{k>0} \phi(k)$ is finite. Proposition 5 applies to obtain an empirical central limit theorem for classes of BV functions. We refer to Dedecker and Prieur [11] for statement and proof of the empirical central limit theorem as far as for applications.

6 The higher dimension case

6.1 Extension of the coefficients

To get more precise inequalities and limit theorems, it is often necessary to consider the dependence between a past $\sigma-$algebra and several points in the future of the sequence. Therefore we focus in this section on the problems arising when extending coefficients based on the conditional distribution function to dimension greater than 2. This is a rather complicated problem. One way to proceed is to start with the functional definition (see Definition 2) of the coefficients. Even by doing so, it remains difficult to extend $\alpha(\mathcal{M}, X)$, $\beta(\mathcal{M}, X)$ and $\phi(\mathcal{M}, X)$ because the notion of bounded variation is delicate as soon as we are in dimension greater or equal to 2. But dealing with τ is more efficient. We get the following immediate extension, whose “dual” form has been introduced by Rüschendorf [34].

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 [17], 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

$$\tau(\mathcal{M}, X) = \Big\|\sup\Big\{\Big|\int f(x)\mathbb{P}_{X|\mathcal{M}}(dx) - \int f(x)\mathbb{P}_X(dx)\Big|, f \in \Lambda_1(\mathcal{X})\Big\}\Big\|_1 . \qquad (14)$$

If $d(0, X)$ is bounded, we can also define the uniform version of τ, which was first introduced by Rio [29]:

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

Let us just mention here that Rio's definition of φ [29] is slightly different from the definition above. He defines the class $\Lambda_1(\mathcal{X})$ as the set of $1-$Lipschitz functions from $\mathcal{X}$ to $[0,1]$.

Thanks to such definitions in spaces of higher dimension, we now 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.

In the following, if $(\mathcal{X}, d)$ is some Polish space, we put on $\mathcal{X}^k$ the distance

$$d_1(x, y) = d(x_1, y_1) + \cdots + d(x_k, y_k). \tag{15}$$

We then define

$$\tau_k(i) = \max_{1\leq l\leq k} \frac{1}{l} \sup\{\tau(\mathcal{M}_p, (X_{j_1}, \ldots, X_{j_l})), p + i \leq j_1 < \cdots < j_l\} \tag{16}$$

$$\text{and } \tau_\infty(i) = \sup_{k>0} \tau_k(i).$$

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

6.2 Coupling

A coupling result for the one-dimensional case and for the coefficient τ has already been stated in Lemma 2 of Section 2.2. In order to extend this result, let us first give some definitions.

Definition 4 :
In the following,

- $(\mathcal{X}, d)$ is a Polish space,
- $(\Omega, \mathcal{A}, \mathbb{P})$ is a given probability space,
- $c : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ is a cost function satisfying

$$c(x, y) = \sup_{u\in Lip_{\mathcal{X}}^{(c)}} |u(x) - u(y)|\,, \tag{17}$$

  where $Lip_{\mathcal{X}}^{(c)}$ is the class of continuous bounded functions u on $\mathcal{X}$ such that $|u(x) - u(y)| \leq c(x, y)\cdot$

We assume that $\int c(x, x_0)\mathbb{P}_X(dx)$ is finite for some (and therefore any) x_0 in $\mathcal{X}$.

We then introduce the following generalization of the coefficient τ defined by (14)

$$\tau_c(\mathcal{M}, X) = \left\|\sup\left\{\left|\int f(x)\mathbb{P}_{X|\mathcal{M}}(dx) - \int f(x)\mathbb{P}_X(dx)\right|, f \in Lip_{\mathcal{X}}^{(c)}\right\}\right\|_1. \tag{18}$$

Let us notice that if $Lip_{\mathcal{X}}^{(c)}$ is a separating class, this coefficient measures the dependence between $\mathcal{M}$ and X ($\tau_c(\mathcal{M}, X) = 0$ if and only if X is independent of $\mathcal{M}$).

We are now in position to state a general coupling result.

Lemma 3 *[34, 12] Let X be a random variable with values in $(\mathcal{X}, d)$. Assume that there exists a random variable δ uniformly distributed over $[0,1]$, independent of the σ-algebra generated by X and $\mathcal{M}$. Then there exists a random variable X^*, measurable with respect to $\mathcal{M}\vee\sigma(X)\vee\sigma(\delta)$, independent of $\mathcal{M}$ and distributed as X, such that*

$$\tau_c(\mathcal{M}, X) = \mathbb{E}(c(X, X^*))\,. \tag{19}$$

Remark 1.
Lemma 3 above can be extended to the case where $\mathcal{X}$ is not a Polish space (see [12]).
If $c(x,y) = \mathbb{I}_{x\neq y}$, (19) was first written by Berbee [4]. In the case where c is a distance for which $\mathcal{X}$ is a Polish space, the result has been proved by Rüschendorf [34] in the particular case where Ω is Polish. The proof of Lemma 3 can be found in [34, 12]. It mainly relies on a parametrized version of the Kantorovich-Rubinštein Theorem (see Proposition 4 in [34] and, for the duality, Theorem 2.1 in [12]).

6.3 Exponential inequalities and statistical results

Working with the coefficients (16), we obtain further results for $\tau-$dependent sequences. We sum up these results below.

Proposition 7 *[10] 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 $\lambda \geq Mq$, $\mathbb{P}(\overline{S}_n \geq (\mathbf{1}_{q>1}+3)\lambda) \leq 4\exp\Big(-\frac{v_q}{(qM)^2}h\Big(\frac{\lambda q M}{v_q}\Big)\Big) + \frac{n}{\lambda}\,\tau_q(q+1)\,.$*

Proposition 7 above extends Bennett's inequality for independent sequences. Starting from the second inequality of Proposition 7 and from the coupling property of Lemma 2, we can prove a functional law of the iterated logarithm (Theorem 1 below). 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 *[10] Let $(X_i)_{i\in\mathbb{Z}}$ be a stationary sequence of real-valued 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*

$$\sum_{k=1}^{\infty} \int_0^{\tau_\infty(k)} Q \circ G(u)\, du < \infty \tag{20}$$

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 $\mathcal{S}$.

To conclude with the statistical applications, we would like to mention two results: the first one is a concentration inequality which is a straightforward consequence of Theorem 1 in Rio [31], and the second one is a Berry-Esseen inequality due to Rio [29].

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 defined by (15).

Theorem 2 *[11] 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

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

Theorem 3 *[29] Let $(X_i)_{i\in\mathbb{Z}}$ be a stationary sequence of real-valued bounded and centered random variables and $\mathcal{M}_i=\sigma(X_j, j\leq i)$. Let $S_n=X_1+\cdots+X_n$ and $\sigma_n=\|S_n\|_2$. If $\limsup_{n\to\infty}\sigma_n=\infty$ and*

$$\sum_{n>0} n\varphi_3(n)<\infty\,, \tag{21}$$

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

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

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

7 Application to dynamical systems on $[0,1]$

The mixing conditions studied in this paper are similar to the usual mixing conditions. However, they are weaker and therefore applicable to substantially broader classes of processes. In this section, we are interested in classes of dynamical systems on $[0,1]$. We refer to [10, 11] for other classes of examples.

The statistical study of dynamical systems is really important because even very simple and determinist dynamical systems may behave in an unpredictable way. Indeed, if we consider two close initial conditions, one can obtain after some finite time two orbits with quite different behaviours. In the litterature on dynamical systems, mixing in the ergodic-theoric sense (MES) is different from mixing in the sense of Rosenblatt [32]. A dynamical system (T^n,μ) is said to be 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. As MES 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. There exist number of works on the statistical properties of dynamical systems. Of course, we do not pretend to give a complete account of the literature. One way to derive covariance inequalities for dynamical systems is to study the spectral properties of the associated Perron-Frobenius operator in some well chosen Banach space (see for example [18, 20, 23]). More recently Pène obtained rates of convergence in the central limit theorem for two-dimensional dispersive billiards [26] and also rate of convergence in multidimensional limit theorems for the Prokhorov metric [27]. Chazottes, Collet, Martinez and Schmitt gave exponential inequalities for expanding maps of the interval [8], Devroye inequalities [6] and statistical applications [7] for a class of non-uniformly hyperbolic dynamical systems. A first approach to study dynamical systems using tools of the theory of weak dependence, in the sense of Doukhan and Louhichi [16], can be found in Prieur [28]. Some more recent works of Dedecker and Prieur [11] prove that furhter results can be obtained by working with the coefficient Φ defined in Definition 1 of Section 1 and with the Markov chain associated to the dynamical system (see Section 7.1 for a precise definition of this Markov chain).

We now introduce the model (Section 7.1), and then apply some of the results of Sections 4 and 6 (Sections 7.2, 7.3 below) to this model.

7.1 Introduction of the model

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 $\mathbb{L}^1(I,\mu)$,

$$|\mathrm{Cov}(h(X_0), k(X_n))| \leq a_n \|k(X_n)\|_1 (\|h\|_{1,\lambda} + \|dh\|)\,, \tag{22}$$

for some nonincreasing sequence a_n tending to zero as n tends to infinity. Note that if (22) holds, then

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

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

$$|\mathrm{Cov}(h(X_0), k(X_n))| \leq 2a_n \|k(X_n)\|_1 \|dh\|\,. \tag{23}$$

Inequality (23) 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 $\phi(\sigma(X_n), X_0) \leq 2a_n$. The answer is positive, due to the following Lemma.

Lemma 4 *[11] 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*

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

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

$$\text{If (22) holds, then } \phi(\sigma(X_n), X_0) \leq 2a_n. \tag{24}$$

In many cases, Inequality (22) follows from the spectral properties of the Markov operator associated to T. In these cases, due to the underlying Markovian structure, (24) holds with $\mathcal{M}_n = \sigma(X_i, i \geq n)$ 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

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

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

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

The converse inequality follows immediately from Lemma 1 of Section 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 [20] and Hofbauer and Keller [18]). 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

$$\mathcal{L}^n(h) = \lambda(h) f_\mu + \Psi^n(h) \quad \text{with} \quad \|\Psi^n(h)\|_v \le K\rho^n \|h\|_v. \tag{25}$$

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

$$I_* = \{f_\mu \ne 0\} \text{ is an interval, and there exists } \gamma > 0 \text{ such that } f_\mu > \gamma^{-1} \text{ on } I_*. \tag{26}$$

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

$$P(h)(x) = \frac{\mathcal{L}(f_\mu h)(x)}{f_\mu(x)}. \tag{27}$$

It is easy to check (see for instance Barbour *et al.* [3]) that $(X_0, X_1, \ldots, X_n)$ has the same distribution as $(Y_n, Y_{n-1}, \ldots, Y_0)$ where $(Y_i)_{i\ge 0}$ is a stationary Markov chain with invariant distribution μ and transition kernel P. Since $\|fg\|_\infty \le \|fg\|_v \le 2\|f\|_v \|g\|_v$, we infer that, taking $C = 2K\gamma(\|df_\mu\| + 1)$,

$$P^n(h) = \mu(h) + g_n \quad \text{with} \quad \|g_n\|_\infty \le C\rho^n \|h\|_v. \tag{28}$$

This estimate implies (22) 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

$$\phi(\sigma(X_n, \ldots, X_{m+n}), X_0) = \phi(\sigma(Y_0, \ldots Y_m), Y_{n+m}) = \phi(\sigma(Y_m), Y_{n+m}) = \phi(\sigma(X_n), X_0) .$$

This being true for any integer m, it holds for $\mathcal{M}_n = \sigma(X_i, i \ge n)$. We conclude that if (25) and (26) hold then there exists $C > 0$ and $0 \le \rho < 1$ such that

$$\phi(\sigma(X_i, i \ge n), X_0) \le 2C\rho^n . \tag{29}$$

Application: Expanding maps. Let $([a_i, a_{i+1}[)_{1\le i\le N}$ be a finite partition of $[0, 1[$. We make the same assumptions on T as in Collet *et al* [8].

1. For each $1 \le j \le N$, 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 $n_0 \ge 1$ such that $T^{-n}(U) \cap V \ne \emptyset$ for all $n \ge n_0$.

If T satisfies 1. 2. and 3. then (25) holds. If furthermore (26) holds (see Morita [23] for sufficient conditions), then (29) holds.

7.2 MISE

Let us now specify the results concerning the MISE in the particular case of dynamical systems described in Section 7.1 above. We know via Proposition 2 that $\beta(\mathcal{M}, X) \leq \phi(\mathcal{M}, X)$. Hence, Propositions 3 and 4 apply to dynamical systems satisfying (22) with $2\sum_{i=1}^{n-1} a_k$ instead of $\sum_{i=1}^{n-1} \beta(k)$. For kernel estimators such a result can also be deduced from a variance estimate (see Prieur [28]).

7.3 Exponential inequalities

This section is devoted to exponential inequalities which can be obtained for dynamical systems. We first need to give the dependence properties of the Markov chain associated to our dynamical system. Let T be a map from $[0,1]$ to $[0,1]$ satisfying Conditions 1. 2. and 3. of Section 7.1. Assume moreover that the density f_μ of the invariant probability μ satisfies (26). Let $X_i = T^i$ and define P as in (27). We know from Section 7.1 that on $([0,1], \mu)$, the sequence $(X_n, \ldots, X_0)$ has the same distribution as $(Y_0, \ldots, Y_n)$ where $(Y_i)_{i\geq 0}$ is the stationary Markov chain with Markov Kernel P. Consequently

$$\varphi(\sigma(X_j, j \geq i+k), (X_0, \ldots, X_k)) = \varphi(\sigma(Y_0), (Y_i, \ldots, Y_{i+k}))\,. \tag{30}$$

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

$$\int_0^1 Q_k(f)(x) g(x) f_\mu(x) d\lambda(x) = \int_0^1 f(T^k(x), T^{k-1}(x), \ldots, x) g(T^k(x)) f_\mu(x) d\lambda(x)\,.$$

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

$$\begin{aligned} \varphi(\sigma(Y_0), (Y_i, \ldots, Y_{i+k})) &= \sup\nolimits_{f \in \Lambda_1(\mathbb{R}^{k+1})} \|P^i \circ Q_k(f) - \mu(Q_k(f))\|_\infty \\ &= \sup\nolimits_{f \in \Lambda_1(\mathbb{R}^{k+1})} \|(P^i - \mu) \circ (Q_k(f) - Q_k(f)(0))\|_\infty \end{aligned} \tag{31}$$

Here, we use a recent result of Collet *et al.* [8]. Denote by $\Lambda_{L_1, \ldots, L_n}$ the set of functions f from $\mathbb{R}^n$ to $\mathbb{R}$ such that

$$|f(x_1, \ldots, x_n) - f(y_1, \ldots, y_n)| \leq L_1|x_1 - y_1| + \cdots + L_n|x_n - y_n|\,. \tag{32}$$

Adapting Lasota-Yorke's approach to higher dimension Collet *et al.* prove (page 312 line 6 [8]) that there exist $K > 0$ and $0 \leq \sigma < 1$ such that, for any f in $\Lambda_{L_1, \ldots, L_{k+1}}$,

$$\|dQ_k(f)\| \leq K \sum_{i=0}^{k} \sigma^i L_{i+1}\,. \tag{33}$$

Applying (28), we infer from (31) and (33) that

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

Moreover, according to (30), the same bound holds for $\varphi(\sigma(X_j, j \geq i+k), (X_0, \ldots, X_k))$. 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 (33) that

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

Several exponential inequalities for dynamical systems are derived from dependence properties of the associated Markov chain. Exponential inequalities of Propositions 5 and 6 in Section 4 can be applied to the sequence $(X_i - \mathbb{E}X_i)_{i\in\mathbb{N}}$. In this section, we would like to focus on the concentration inequality for Lipschitz functions stated in Theorem 2. Starting from (33) and (28) we get that, for any function f belonging to $\Lambda_{L_1,\ldots,L_n}$,

$$\mathbb{P}(f(Y_1,\ldots,Y_n) - \mathbb{E}(f(Y_1,\ldots,Y_n)) \geq x) \leq \exp\Big(\frac{-2x^2}{M_1^2+\cdots+M_n^2}\Big)\,. \tag{34}$$

with

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

Since $(X_1,\ldots,X_n)$ has the same distribution as $(Y_n,\ldots,Y_1)$, the bound (34) holds for $f(X_1,\ldots,X_n)$ with

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

Remark 8. Assume that (34) 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 $M_i^2 \leq C_n \sum_{j=i}^n \delta_{j-i}L_i^2$, and consequently $\sum_{i=1}^n M_i^2 \leq C_n^2 \sum_{i=1}^n L_i^2$. Hence, (34) yields the upper bound

$$\mathbb{P}(f(X_1,\ldots,X_n) - \mathbb{E}(f(X_1,\ldots,X_n)) \geq x) \leq \exp\Big(\frac{-2x^2}{C_n^2(L_1^2+\cdots+L_n^2)}\Big)\,. \tag{35}$$

For expanding maps, (35) has been proved by Collet *et al* [8].

Aknowledgment. I would like to thank Jérôme Dedecker for suggestions concerning the organization of the paper.

References

1. Andrews, D. W. K.: Nonstrong mixing autoregressive processes. J. Appl. Probab., **21**, 930-934 (1984).
2. Ango Nze, P., Doukhan, P.: Weak Dependence: Models and Applications. In: Empirical process, Techniques for dependent data. Birkhäuser (2002).
3. Barbour, A.D., Gerrard, R.M., Reinert, G.: Iterates of expanding maps. Probab. Theory Relat. Fields, **116**, 151-180 (2000).
4. Berbee, H.C.P.: Random walks with stationary increments and renewal theory. Math. Cent. Tracts. Amsterdam (1979).
5. Bradley, R.C.: Introduction to Strong Mixing Conditions, Volume 1. Technical Report, Department of Mathematics, I. U. Bloomington (2002).
6. Chazottes, J.R., Collet, P., Schmitt B.: Devroye Inequality for a Class of Non-Uniformly Hyperbolic Dynamical Systems. Preprint (2004). Available on `http://arxiv.org/abs/math/0412166`.
7. Chazottes, J.R., Collet, P., Schmitt B.: Statistical Consequences of Devroye Inequality for Processes. Applications to a Class of Non-Uniformly Hyperbolic Dynamical Systems. Preprint (2005). Available on `http://arxiv.org/abs/math/0412167`.
8. 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).
9. Dedecker, J.: Inégalités de covariance. C. R. Acad. Sci. Paris, Ser I, **339**, 503-506 (2004).
10. Dedecker, J., Prieur, C.: Coupling for τ-dependent sequences and applications. J. Theoret. Probab., **17 no 4**, 861-885 (2004).
11. Dedecker, J., Prieur, C.: New dependence coefficients. Examples and applications to statistics. To appear in Probab. Theory Relat. Fields (2005).

12. Dedecker, J., Prieur, C., Raynaud De Fitte, P.: Parametrized Kantorovitch-Rubinštein Theorem and coupling for the minimal distance (2005).
13. Delyon, B.: Limit theorems for mixing processes. Tech. Report 546 IRISA, Rennes I (1990).
14. Doukhan, P.: Mixing: properties and examples. Lecture Notes in Statist., **85**, Springer-Verlag (1994).
15. Doukhan, P.: Models inequalities and limit theorems for stationary sequences. In: Doukhan, P. *et alii* (ed) Theory and applications of long range dependence, Birkhäuser (2002).
16. Doukhan, P., Louhichi, S.: A new weak dependence condition and applications to moment inequalities. Stochastic Process. Appl., **84**, 313-342 (1999).
17. Dudley, R.M.: Real analysis and probability. Wadworsth Inc., Belmont, California (1989).
18. Hofbauer, F., Keller, G.: Ergodic properties of invariant measures for piecewise monotonic transformations. Math. Z., **180**, 119-140 (1982).
19. Ibragimov, I.A.: Some limit theorems for stationary processes. Theory Probab. Appl., **7**, 349-382 (1962).
20. Lasota, A., Yorke, J.A.: On the existence of invariant measures for piecewise monotonic transformations. Trans. Amer. Math. Soc., **186**, 481-488 (1974).
21. Major, P.: On the invariance principle for sums of identically distributed random variables. J. Multivariate Anal., **8**, 487-517 (1978).
22. Mokkadem, A.: Study of risks of kernel estimators. Teor. Veroyatnost. i Primenen, **35**, no. 3, 531-538 (1990); translation in Theory Probab. Appl., **35** (1990), no. 3, 478-486 (1991).
23. 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).
24. Peligrad, M.: A note on two measures of dependence and mixing sequences. Adv. Appl. Probab., **15**, 461-464 (1983).
25. Peligrad, M.: Some remarks on coupling of dependent random variables. Stat. Prob. Letters, **60**, 201-209 (2002).
26. Pène, F.: Rates of convergence in the CLT for two-dimensional dispersive billiards. Comm. Math. Phys. **225**, no. 1, 91-119 (2002).
27. Pène, F.: Multiple decorrelation and rate of convergence in multidimensional limit theorems for the Prokhorov metric. Ann. Probab. **32**, no. 3B, 2477-2525 (2004).
28. Prieur, C.: Density Estimation For One-Dimensional Dynamical Systems. ESAIM, Probab. & Statist., `WWW.emath.fr/ps`, **5**, 51-76 (2001).
29. Rio, E.: Sur le théorème de Berry-Esseen pour les suites faiblement dépendantes. Probab. Theory Relat. Fields, **104**, 255-282 (1996).
30. Rio, E.: Théorie asymptotique des processus aléatoires faiblement dépendants. Collection Mathématiques & Applications **31**. Springer, Berlin (2000).
31. Rio, E.: Inégalités de Hoeffding pour les fonctions lipschitziennes de suites dépendantes. C. R. Acad. Sci. Paris Série I, **330**, 905-908 (2000).
32. Rosenblatt, M.: A central limit theorem and a strong mixing condition. Proc. Nat. Acad. Sci. U. S. A., **42**, 43-47 (1956).
33. Rozanov, Y.A., Volkonskii, V.A.: Some limit theorems for random functions I. Theory Probab. Appl., **4**, 178-197 (1959).
34. Rüschendorf, L.: The Wasserstein Distance and Approximation Theorems. Z. Wahrscheinlichkeitstheorie verw. Gebiete, **70**, 117-129 (1985).
35. Viennet, G.: Inequalities for absolutely regular sequences: application to density estimation. Probab. Theory Relat. Fields, **107**, 467-492 (1997).

Mathematics Subject Classifications (2000): 62G07, 60J10, 60E15, 37C30.
Key words: dependence coefficients, mixing, covariance inequalities, coupling, dynamical systems, ergodic theory, density estimation, exponential inequalities, concentration inequalities, Berry-Esseen inequality.