The exponential convergence of kernel density estimator in L1 for ∅-mixing processes

Liangzhen Lei, Liming Wu

Pub. Inst. Stat. Univ. Paris
XXXXVIII, fasc. 1-2, 2004, 59 à 68

The exponential convergence of kernel density estimator in L^1 for ϕ-mixing processes

Liangzhen Lei* Liming WU†

October 2002

Abstract

In this paper we establish the exponential convergence of the nonparametric kernel density estimator f_n^* to the unknown density f in $L^1(\mathbb{R}^d, dx)$ for a $\mathbb{R}^d$-valued ϕ-mixing process verifying $\sum_k \phi_k < +\infty$ and yield an exponential inequality of Hoeffding's type.

Keywords : kernel density estimator, ϕ-mixingness, Hoeffding type's inequality, exponential convergence.

AMS Mathematical Subject Classification : 62G07, 60G10, 60F17.

1 Introduction

Let $\{X_i; i \geq 1\}$ be a sample taken from a ϕ-mixing process with values in $\mathbb{R}^d$, defined on probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with marginal distribution measure $d\mu = f(x)dx$, where the density $f \in L^1(\mathbb{R}^d)$ is unknown.

The empirical measure is $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$. Let K be a measurable function such that

(H1) $K \geq 0$, $\int_{\mathbb{R}^d} K dx = 1$,

and set $K_h(x) = \frac{1}{h^d} K(\frac{x}{h})$. The kernel density estimator of f is defined as usually as:

*Dept.Math., Wuhan University,430072-Hubei, China. E-mail address: zjhllz@163.com

†Laboratoire de Mathématiques Appliquées. CNRS-UMR 6620, Université Blaise Pascal, 63177 Aubière, France. E-mail address: Li-Ming.Wu@math.univ-bpclermont.fr and Dept. Math., Wuhan University, 430072-Hubei,China

$$f_n^*(x) = K_{h_n} * dL_n(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_n^d} K\left(\frac{x - X_i}{h_n}\right), \quad x \in \mathbb{R}^d \tag{1.1}$$

where $\{h_n, n \geq 1\}$ is a sequence of positive numbers (width of bands) satisfying

(H2) $h_n \to 0, \quad nh_n^d \to +\infty \quad as \quad n \to \infty.$

A natural measure of closeness of f_n^* to the unknown f is its L_1 distance below,

$$J_n = \int_{\mathbb{R}^d} |f_n^*(x) - f(x)| dx. \tag{1.2}$$

In the independent and identically distributed (i.i.d in short) case, L. Devroye, in a fundamental paper [5], proved

Theorem 1.1. *Let (X_i) be i.i.d. and K a nonnegative Borel measurable function on $\mathbb{R}^d$ with $\int k(x)dx = 1$ (i.e., (H1)), then the following conditions are equivalent:*

(i) $J_n \to 0$ in probability as $n \to \infty$ (weak consistency);

(ii) $J_n \to 0$ almost surely as $n \to \infty$ (strong consistency);

(iii) $J_n \to 0$ exponentially as $n \to \infty$, i.e., for any $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(J_n > \delta) < 0; \tag{1.3}$$

(iv) $\lim_n h_n = 0$ and $\lim_n nh_n^d = \infty$ (i.e., (H2)).

Recently Louani [11] (2000) prove the existence of limit in (1.3) and identifies that limit $-I(\delta)$ (i.e., a large deviation principle). And we (together with B. Xie) [10] establish a weak large deviation principle of f_n^* in L^1 (and it is known that the *good* large deviation principle fails). Gao [8] (2002) establishes the large and moderate deviation principle for f_n^* in L^∞ under mild condition on K, f and (h_n).

The study of kernel density estimators in the dependent cases were realized by many people from different points of view : see Peligrad [13] (1992), T.M. Adams and A.B. Nobel [1] (1998), Bosq, Merlevède and Peligrad [3] (1999) and the references therein. For instance Peligrad [13] established the uniform consistency of f_n^* (i.e., in L^∞ instead of L^1) under weaker condition on the ϕ-dependence coefficient (ϕ_k) than that used in this Note (nevertheless his conditions on $K, f, (h_n)$ are much stronger). In T.M. Adams and A.B. Nobel [1] (1998), a general procedure to construct ergodic processes for which the kernel density estimator *fails to be weakly consistent under (H1) and (H2)* is exhibited.

Note however that in the ϕ-mixing case, how to extend those results of large and moderate deviations in [11], [8], [10] in the i.i.d. case is an interesting open question. This question is quite delicate because even for stationary Doeblin recurrent Markov

chains (for which ϕ_k decays exponentially to zero), the large deviation principle about partial sums fails in general (see Bryc and Dembo [4]).

In this Note we will carry out a first step towards the large deviations of f_n^*, i.e., to establish the exponential convergence of f_n^* to f in $L^1(\mathbb{R}^d, dx)$ for ϕ-mixing sequences verifying $\sum_k \phi_k < +\infty$, under **(H1)** and **(H2)**. Moreover we will yield an exponential inequality of Hoeffding type. Our main tool is the Hoeffding type inequality established recently by Rio [14](2000).

2 Main results

We briefly recall what is meant by the terminology of ϕ-mixing process. Given a sequence of random variables $(X_i)_{i\geq 1}$ with values in $\mathbb{R}^d$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$. For two sub-$\sigma$-algebras $\mathcal{A}, \mathcal{B}$ in $\mathcal{F}$, define

$$\phi(\mathcal{A}, \mathcal{B}) = \sup \left\{ \left| \mathbb{P}(V) - \frac{\mathbb{P}(U \bigcap V)}{\mathbb{P}(U)} \right| ;\ U \in \mathcal{A}, \mathbb{P}(U) \neq 0, V \in \mathcal{B} \right\}.$$

Define for every integer k,

$$\phi_k = \sup_{m \geq 1} \left\{ \phi(\sigma(X_1, \cdots, X_m), \sigma(X_{m+l}; l \geq k)) \right\}.$$

Theorem 2.1. *Let $(X_i)_{i\in\mathbb{N}^*}$ be a stationary sequence of $\mathbb{R}^d$-valued r.v. with marginal law $\mu(dx) = f(x)dx$. Assume that*

$$S_\phi := \sum_{k=1}^{\infty} \phi_k < +\infty. \tag{2.1}$$

Let K be a nonnegative measurable function on $\mathbb{R}^d$ with $\int K(x)dx = 1$ (i.e. (H1)) and (h_n) a sequence of positive numbers verifying (H2). Then $J_n \to 0$ exponentially as $n \to \infty$, i.e.,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(J_n > \delta) < 0, \qquad \forall \delta > 0.$$

Theorem 2.2. *In the context of Theorem 2.1, assume (2.1) and (H1) for K. Then for every $n \geq 1$ and all $r > 0$,*

$$\mathbb{P}(|J_n - \mathbb{E}J_n| > r/\sqrt{n}) \leq 2 \exp\left(-\frac{r^2}{2(1 + S_\phi)^2} \right). \tag{2.2}$$

Remarks: How to extend those results to random fields $(X_i)_{i\in\mathbb{Z}^d}$ is an interesting open question. The main point consists in extending the Rio's inequality in Lemma 3.1.

3 Some deviation inequalities for ϕ-mixing sequences

All of this Note is based on the following Hoeffding type inequality established recently by E. Rio [14] (2000).

Lemma 3.1. *([14]) Let $f : E^n \to \mathbb{R}$ satisfy*

$$|f(x) - f(y)| \leq L \tag{3.1}$$

for all $x, y \in E^n$ verifying $\#\{i;\ x_i \neq y_i\} = 1$. Then $\forall \lambda > 0$,

$$\begin{aligned} &\mathbb{E}\exp\left[\lambda\left(f(X_1, \cdots, X_n) - \mathbb{E}f(X_1, \cdots, X_n)\right)\right] \\ &\leq \exp\left(\frac{\lambda^2}{8} \cdot nL^2(1 + 2S_\phi)^2\right); \end{aligned} \tag{3.2}$$

and in particular $\forall t > 0$,

$$\mathbb{P}\left(f(X_1, \cdots, X_n) - \mathbb{E}f(X_1, \cdots, X_n) > t\right) \leq \exp\left(-\frac{2t^2}{nL^2(1 + 2S_\phi)^2}\right). \tag{3.3}$$

Indeed (3.3) is Rio [14], Corollaire 1 and (3.2) is an immediate consequence of [14], Théorème 1 together with the proof of his Corollaire 1.

Consider the Hamming distance on E^n:

$$d_H(x, y) := \#\{i;\ x_i \neq y_i\}.$$

Condition (3.1) is equivalent to

$$|f(x) - f(y)| \leq L d_H(x, y),\ \forall x, y \in E^n$$

i.e., the Lipchitzian coefficient of f w.r.t. the Hamming distance d_H is less than L. In such way we can translate Rio's inequality (3.2) into the following transportation inequality, by Bobkov-Goetze [2]:

Corollary 3.2. *Let μ_n be the law of $(X_1, \cdots, X_n)$. Then for any probability measure ν on E^n,*

$$W_1(\nu, \mu_n) \leq (1 + 2S_\phi)\sqrt{\frac{n}{2} \cdot h(\nu; \mu_n)}. \tag{3.4}$$

Here

$$W_1(\nu; \mu_n) := \inf \iint d_H(x, y) d\pi(x, y)$$

where the infimum is taken over all probability measures π on $E^n \times E^n$ with marginals ν and μ_n, is the Wasserstein distance between ν and μ_n; and

$$h(\nu, \mu_n) := \begin{cases} \int \log \frac{d\nu}{d\mu_n} d\nu & \text{if } \nu \ll \mu_n, \\ +\infty & \text{otherwise} \end{cases}$$

is the relative entropy (or Kullback information) of ν w.r.t. μ_n.

Notice that when $X_1, \cdots, X_n$ are independent, $S_\phi = 0$ and inequality (3.4) is a consequence of the Pinsky inequality (i.e., (3.4) with $n = 1$ and $S_\phi = 0$) together with the tensorization technique. Inequality (3.4) was proved at first by Marton [12] (1996) for Doeblin recurrent Markov chains, and next extended by Samson [15] (2000) for general ϕ-mixing sequences. But the condition in [15] is $\sum_k \sqrt{\phi_k} < +\infty$, stronger than the condition here.

See Ledoux [9] for a systematic treatment (and references) and application of such a transportation inequality to concentration of measure, and H. Djellout, A. Guillin and the second author [7] (2002) for some further extensions of Rio's result above.

4 Proofs of the main results

4.1 Proof of Theorem 2.2

Let

$$g(x_1, \cdots, x_n) := \int_{\mathbb{R}^d} \left| \frac{1}{n} \sum_{j=1}^{n} K_h(u - x_i) - f(u) \right| du, \ \forall x = (x_1, \cdots, x_n) \in E^n.$$

Then $J_n = g(X_1, \cdots, X_n)$. Fix $i = 1, \cdots, n$. For all $x, y \in E^n$ such that $x_j = y_j$ for all j except $j = i$, we have

$$|g(x) - g(y)| \leq \frac{1}{n} \int_{\mathbb{R}^d} |K_h(u - x_i) - K_h(u - y_i)| \, du \leq \frac{2}{n}.$$

In other words g verifies (3.1) with $L = 2/n$. Thus applying (3.3) to g and $-g$, we get (2.2).

4.2 Proof of Theorem 2.1

For the convenience of the reader we recall two well known lemmas in Analysis:

Lemma 4.1. *(L_1 version of Bochner's theorem) Let K be a nonnegative Borel function on $\mathbb{R}^d$ with $\int K(x)dx = 1$. Then $\lim_{h \to 0+} \int |K_h * f(x) - f(x)|dx = 0$, where $K_h(x) = h^{-d} K(x/h)$.*

See L.Devroye [5] (Lemma 1, p.897).

Lemma 4.2. *(Lebesgue density theorem)If f is a density on $\mathbb{R}^d$ and B is a compact set of $\mathbb{R}^d$ with $\lambda(B) > 0$ where λ is the Lebesgue measure, then*

$$\lim_{h \to 0} \lambda^{-1}(hB) \int_{x+hB} f(y)dy = f(x), \quad \text{for almost all} \quad x.$$

See L.Devroye [5] (Lemma 2, p.898).
We now go to the

Proof of Theorem 2.1. . Its proof will be divided into three steps, where the first two steps are close to that of [5].

Step 1 By Lemma 4.1, it suffices to show that $\int |f_n^*(x) - K_h * f(x)|dx \to 0$ exponentially as $n \to \infty$. Note that

$$f_n^*(x) = K_h * L_n = h^{-d} \int K(\frac{x-y}{h}) dL_n(y).$$

Given $\varepsilon > 0$, we can find finite positive constants M, L, m, $a_1, \cdots, a_m$ and disjoint finite rectangles $A_1, \cdots, A_m$ in $\mathbb{R}^d$ of form $\prod_{i=1}^d [x_i, x_i + a_i)$ such that the function

$$K^{(\varepsilon)}(x) = \sum_{j=1}^m a_j I_{A_j}(x)$$

satisfies: $K^{(\varepsilon)} \leq M, K^{(\varepsilon)} = 0$ outside $[-L, L]^d$, and $\int |K(x) - K^{(\varepsilon)}|dx < \varepsilon$. Define

$$f_n^{(\varepsilon),*} := K_h^{(\varepsilon)} * L_n.$$

Then

$$\begin{aligned}
\int |f_n^*(x) - K_h * f(x)|dx \leq & \int |f_n^*(x) - f_n^{(\varepsilon),*}(x)|dx \\
& + \int |f_n^{(\varepsilon),*}(x) - K_h^{(\varepsilon)} * f(x)|dx + \int |K_h^{(\varepsilon)} * f - K_h * f(x)|dx \\
\leq & \int h^{-d} \int |K^{\varepsilon}(\frac{x-y}{h}) - K(\frac{x-y}{h})|L_n(dy)dx \\
& + \int |f_n^{(\varepsilon),*}(x) - K_h^{(\varepsilon)} * f(x)|dx \\
& + \int h^{-d} \int |K^{(\varepsilon)}(\frac{x-y}{h}) - K(\frac{x-y}{h})|f(y)dydx \\
\leq & 2\varepsilon + \int |f_n^{(\varepsilon),*}(x) - K_h^{(\varepsilon)} * f(x)|dx.
\end{aligned} \tag{4.1}$$

Noting that $d\mu = fdx$, then

$$\begin{aligned}
\int |f_n^{(\varepsilon),*}(x) - K_h^{(\varepsilon)} * f(x)|dx \leq & \sum_{j=1}^m |a_j| \int |h^{-d} \int_{x+hA_j} f(y)dy - h^{-d} \int_{x+hA_j} L_n(dy)|dx \\
\leq & Mh^{-d} \sum_{j=1}^m \int |\mu(x + hA_j) - L_n(x + hA_j)|dx.
\end{aligned}$$

Consequently for Theorem 2.1, it is enough to prove that for any finite rectangle $A := \prod_{i=1}^d [x_i, x_i + a_i)$ of $\mathbb{R}^d$,

$$h^{-d} \int |L_n(x + hA) - \mu(x + hA)|dx \to 0 \quad \text{exponentially as} \quad n \to \infty. \tag{4.2}$$

Step 2. Fix such a rectangle $A := \prod_{i=1}^{d}[x_i, x_i + a_i)$, and let $\varepsilon > 0$ be arbitrary. Consider the partition of $\mathbb{R}^d$ into sets B that are d-fold products of intervals of the form $[\frac{(i-1)h}{N}, \frac{ih}{N})$, where i is an integer, and N is a fixed integer to be chosen later. Call the partition Ψ.

Let N be such that $\min_i a_i \geq \frac{2}{N}$, $A^* = \Pi_{i=1}^{d}[x_i + \frac{1}{N}, x_i + a_i - \frac{1}{N})$. Define

$$C_x = x + hA - \bigcup_{B\in\Psi, B\subseteq x+hA} B \subseteq x + h(A\backslash A^*).$$

Clearly,

$$\begin{aligned} &\int |\mu(x + hA) - L_n(x + A)|dx \\ &\leq \int \sum_{B\in\Psi, B\subseteq x+hA} |\mu(B) - L_n(B)|dx + \int \{\mu(C_x) + L_n(C_x)\}dx \end{aligned} \tag{4.3}$$

Using the fact that for any set C, and any probability measure ν on $\mathbb{R}^d$,

$$\int \nu(x + hC)dx = \lambda(hC),$$

where λ is the Lebesgue measure, the last term in (4.3) is bounded from above by

$$\begin{aligned} 2\lambda(h(A\backslash A^*)) &= 2h^d\lambda(A\backslash A^*) = 2h^d\left(\prod_{i=1}^{d} a_i - \prod_{i=1}^{d}\left(a_i - \frac{2}{N}\right)\right) \\ &= 2h^d\lambda(A)\left(1 - \prod_{i=1}^{d}\left(1 - \frac{2}{Na_i}\right)\right) \\ &\leq \varepsilon h^d \end{aligned}$$

once if N verifies

$$\min_i a_i \geq \frac{2}{N}, \text{ and } 2\lambda(A)\left(1 - \prod_{i=1}^{d}\left(1 - \frac{2}{Na_i}\right)\right) \leq \varepsilon.$$

Fix such N which is independent of n.

For any finite constant $R > 0$, letting $S_{OR} := \{x \in \mathbb{R}^d;\ |x| \leq R\}$, we can bound the first term in (4.3) from above by

$$\sum_{B\in\Psi, B\cap S_{OR}\neq\phi} |L_n(B) - \mu(B)| \int_{B\subseteq x+hA} dx + \int_{B\subseteq x+hA} dx\{L_n(S_{OR}^c) - \mu(S_{OR}^c) + 2\mu(S_{OR}^c)\}.$$

Here $(\cdot)^c$ denotes the complement of a set. Clearly, $h^{-d}\int_{B\subseteq x+hA} dx \leq \lambda(A)$, and $\mu(S_{OR}^c) < \varepsilon$ by sufficiently large R.

By Lemma 3.1,

$$\mathbb{P}\{L_n(S_{OR}^c) - \mu(S_{OR}^c) > \delta\} \le \exp\left(-\frac{2n\delta^2}{(1+2S_\phi)^2}\right), \ \forall \delta > 0.$$

Consequently for (4.2) it remains to establish

$$\sum_{B\in\Psi, B\cap S_{OR}\neq\phi} |L_n(B) - \mu(B)| \to 0, \text{ exponentially.} \tag{4.4}$$

Step 3. Our proof of the key estimate (4.4) is very different from that in [5] and it is the main new point here. Set

$$\tilde{\Psi} = \{B; B \in \Psi, B \cap S_{OR} \neq \emptyset\}, \ C := \left(\bigcup_{B\in\tilde{\Psi}} B\right)^c$$

$$\mathcal{B}(\tilde{\Psi}) = \sigma\{B; B \in \tilde{\Psi}\}.$$

Regarding L_n and μ as probability measures on $\mathcal{B}(\tilde{\Psi})$, and denoting the total variation of $L_n - \mu$ on $\mathcal{B}(\tilde{\Psi})$ by $\|L_n - \mu\|_{\mathcal{B}(\tilde{\Psi})}$, we have

$$\sum_{B\in\Psi, B\cap S_{OR}\neq\emptyset} |L_n(B) - \mu(B)| \le \|L_n - \mu\|_{\mathcal{B}(\tilde{\Psi})} = 2 \max_{\tilde{B}\in\mathcal{B}(\tilde{\Psi})} |L_n(\tilde{B}) - \mu(\tilde{B})|.$$

Therefore,

$$\begin{aligned}\mathbb{P}\left(\sum_{B\in\Psi, B\cap S_O R\neq\emptyset} |L_n(B) - \mu(B)| > \varepsilon\right) &\le \mathbb{P}\left(\max_{\tilde{B}\in\mathcal{B}(\tilde{\Psi})} |L_n(\tilde{B}) - \mu(\tilde{B})| > \frac{\varepsilon}{2}\right) \\ &\le \sum_{\tilde{B}\in\mathcal{B}(\tilde{\Psi})} \mathbb{P}\left(|L_n(\tilde{B}) - \mu(\tilde{B})| > \frac{\varepsilon}{2}\right)\end{aligned}$$

At first by Lemma 3.1, for each $\tilde{B} \in \mathcal{B}(\tilde{\Psi})$,

$$\begin{aligned}\mathbb{P}\left(|L_n(\tilde{B}) - \mu(\tilde{B})| > \frac{\varepsilon}{2}\right) &\le \mathbb{P}\left(L_n(\tilde{B}) > \mu(\tilde{B}) + \frac{\varepsilon}{2}\right) + \mathbb{P}\left(L_n(\tilde{B}) < \mu(\tilde{B}) - \frac{\varepsilon}{2}\right) \\ &\le 2\exp\left(-\frac{n\varepsilon^2}{2(1+2S_\phi)^2}\right).\end{aligned}$$

Secondly, the number of elements $\#\tilde{\Psi}$ in $\tilde{\Psi}$ is not greater than $\left(\frac{2RN}{h} + 2\right)^d = o(n)$ by (H2), $\tilde{\mathcal{B}}(\phi)$ has $2^{\#\tilde{\Psi}} = 2^{o(n)}$ elements for n large enough. Consequently

$$\mathbb{P}\left(\sum_{B\in\Psi, B\cap S_O R\neq\emptyset} |L_n(B) - \mu(B)| > \varepsilon\right) \le 2^{o(n)} 2\exp\left(-\frac{n\varepsilon^2}{2(1+2S_\phi)^2}\right)$$

where the desired (4.4) follows. This completes the proof of Theorem 2.1.

□

5 Concluding remarks

From the proof of Theorem 2.1 above, we see clearly that for proving the exponential convergence of f_n^* to the unknown density f in L^1, it is enough to show the key relation (4.4) together with the exponential convergence of $L_n(S_{OR}^c)$ to $\mu(S_{OR}^c)$. Thus by following the proof of (4.4), we see that Theorem 2.1 will remain valid once if we can prove the following exponential deviation inequality

$$\mathbb{P}(|L_n(A) - \mu(A)| > \delta) \leq C_1(\delta)e^{-C_2(\delta)n}, \ \forall A \subset \mathcal{B}(\mathbb{R}^d) \tag{5.1}$$

for some constants $C_1(\delta), C_2(\delta)$ depending only on δ (independent of n, A) and for any $\delta > 0$.

In this Note we have applied Rio's inequality which is stronger than (5.1). The reader certainly guess that (5.1) holds in a much wider situation than the uniform mixing case treated in this paper.

References

[1] T.M. Adams and A.B. Nobel (1998), On density estimation from ergodic processes, *Ann. Probab. 26, No.2, 794-804.*

[2] S.G. Bobkov and F. Götze, Exponential integrability and transportation cost related to logarithmic Sobolev inequalities, *J. Funct. Anal.* **163**,*1-28*(1999)

[3] D. Bosq, F. Merlevède, M. Peligrad, Asymptotic normality for density kernel estimators in discrete and continuous time. *J. Multivariate Anal.* **68**, *No. 1, 78-95* (1999)

[4] W. Bryc and A. Dembo(1996), Large deviations and strong mixing, *Ann. Inst. H. Poincaré (Série Proba. et Stat.), 32, 549-569.*

[5] L. Devroye (1983), The equivalence of weak, strong and complete convergence in L_1 for kernel density estimates, *Ann. Stat. 1983, Vol.11, No.3., 896-904*

[6] L. Devroye (1988), The kernel estimate is relatively stable. *Prob. Theo. Rel. Field. 1988, Vol.77, p537-550*

[7] H. Djellout, A. Guillin and L. Wu (2002), Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Preprint*, 2002. *To appear in Ann. Probab.*

[8] F.Q. Gao (2001), Moderate deviations and large deviations for the kernel density estimators. Preprint.

[9] M. Ledoux,(1999), Concentration of measure and logarithmic sobolev inequalities, *S 'eminaire de Probab. XXXIII,LNM* **1709**,*p120-216*

[10] L. Lei, L. Wu and B. Xie (2002), Large deviations and deviation Inequality for Kernel Density Estimator in $L_1(R^d)$-distance, In *Development of Modern Statistics and Related Topics (Series in Biostatistics Vol.1), p89-97* (2003).

[11] D. Louani(2000), Large deviations for the L_1-distance in kernel density estimation. *Jour. Stat. Plan. Infer. p177-182*

[12] K.Marton(1996), A measure concentration inequality for contracting Markov chains. *Geom.Funct.Anal.* **7**. *556-571*

[13] M. Peligrad, Properties of uniform consistency of the kernel estimators of density and of regression functions under dependence assumption. *Stoch. and Stoch. Reports, Vol. 40, 147-168* (1992)

[14] E. Rio, Inégalités de Hoeffding pour les fonctions lipchitziennes de suites dépendantes, *CRAS, Paris, t.330, Série I, 905-908* (2000)

[15] P.M. Samson (2000), Concentration of measure inequalites for Markov chains and ϕ-mixing processes, *Ann. Prob. 2000, Vol 28, No.1., 416-461*