Kernel Regression for $\tilde{\rho}$-mixing Observations

Mounir ARFI
College of Inverness
Crown Avenue
Inverness, Scotland
mounir.arfi@gmail.com

Abstract

We give the uniform almost sure convergence of the kernel estimate of the regression function over a sequence of compact sets which increases to $\mathbb{R}^d$ when n approaches the infinity and the observed process is $\tilde{\rho}$-mixing . The used estimator for the regression function is the kernel estimator proposed by Nadaraya, Watson (1964).

Kewwords: Kernel Estimate, Regression, $\tilde{\rho}$-mixing.
2000 Mathematics Subject Classifications: 62G05, 62G08.

1. Introduction

Let $(X_t, Y_t)_{t\in \mathbb{N}}$ be a strictly stationary process where (X_t, Y_t) takes value in $\mathbb{R}^d \times \mathbb{R}$ and is distributed as (X, Y). Suppose that a segment of data $(X_t, Y_t)_{t=1}^n$ has been observed.

We are interested in the study of the kernel estimate of the well known regression function defined by $r(x) = E(Y_t / X_t = x) \quad t \in \mathbb{N}$. A natural estimator of the function $r(.)$ is given by:

$$r_n(x) = \sum_{t=1}^{n} Y_t K(h_n^{-1}(x - X_t)) / \sum_{t=1}^{n} K(h_n^{-1}(x - X_t)) \quad \forall x \in E$$

where E stands for the subset $\{x \in R, f(x) > 0\}$, f being the density of the process (X_t). (h_n) is a positive sequence of real numbers such that $(h_n) \to 0$ and $nh_n \to \infty$ when $n \to \infty$ and K is a Parzen-Rosenblatt kernel type in the sense of a bounded integrable function satisfying $\int_{\mathbb{R}^d} K(x)dx = 1$ and $\lim_{\|x\|\to\infty} \|x\| K(x) = 0$, besides it will be assumed to be strictly positive and with bounded variation.

Such an approach has been subject to several investigations since many years. A number of distinguished papers is related to this topic. There are among others, Devroye (1981), Collomb (1984, 1985), Györfi *et al.* (1989), Härdle (1990), Bosq (1996) , Arfi (1996, 1997, 2003) and Walk (2006).

Watson (1964), for instance, considered the estimation of the conditional expectation as a predictor of Y and applied this method to some climatological time series data; Nadaraya (1964), established the same estimator independently.

The aim of the present paper is to study the almost sure convergence for $\widetilde{\rho}$-mixing random variable sequences over a sequence of compact sets which increases to $\mathbb{R}^d$.

2. Preliminaries and Assumptions

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and let $(X_t,\ t \in \mathbb{N})$ be a sequence of random variables and we write $\mathcal{F}_2 = \sigma(X_t,\ t \in S \subset \mathbb{N})$.

Given the σ-algebras $\mathcal{B}$ and $\mathcal{R}$ in $\mathcal{F}$, let

$$\rho(\mathcal{B},\ \mathcal{R}) = \sup[corr(X,Y), X \in L_2(\mathcal{B}), Y \in L_2(\mathcal{R})]$$

where $corr(X,Y) = (EXY - EXEY)/\sqrt{varXvarY}$.

Bradley (1990) introduced the following coefficients of dependence

$\widetilde{\rho}(k) = \sup\{\rho(\mathcal{F}_S, \mathcal{F}_T)\},\ \ k \ge 0$ where the supremum is taken over all finite subsets $S, T \subset \mathbb{N}$ such that $dist(S,T) \ge k.$

Obviously,

$$0 \le \widetilde{\rho}(k+1) \le \widetilde{\rho}(k) \le 1, \quad k \ge 0 \quad and \quad \widetilde{\rho}(0) = 1.$$

Definition 2.1. A random variable sequence $(X_t, t \ge 1)$ is said to be $\widetilde{\rho}$-mixing if there exists $k \in N$ such that $\widetilde{\rho}(k) < 1$.

Without loss of generality we may assume that $(X_t,\ t \ge 1)$ is such that $\widetilde{\rho}(1) < 1$ (see Bryc and Smolenski, 1993). In the study of $\widetilde{\rho}$-mixing sequences we refer to Bradley(1990, 1992) for the central limit theorem, Bryc and Smolenski (1993) for moment inequalities and almost sure convergence, Peligrad and Gut (1999) for almost sure results for a class of dependent random variables, Shixin (2004) for the almost sure convergence for $\widetilde{\rho}$-mixing random variable sequences obtaining new results which extend and improve some previous results , Sung (2010) for complete convergence for weighted sums.

We will make use of the following assumptions:

A1 The observed process $(X_t)_{t \in \mathbb{N}}$ is stationary and $\widetilde{\rho}$-mixing.

A2 $\exists \Gamma < \infty,\ \forall x \in \mathbb{R}^d,\ \ f(x) \le \Gamma$

and

$$\exists \gamma_n > 0;\ \forall x \in C_n, \quad f(x) \ge \gamma_n.$$

Where C_n is a sequence of compact sets such that $C_n = \{x : \| x \| \le c_n\}$

with $c_n \to \infty.$

A3 $\exists\ b > 2,\ \exists\ M < \infty\ \ E(|Y|^b) < M$.
A4 $\exists\ V < \infty,\ \forall\ \ x \in \mathrm{I}R^d,\ E\left[(Y - r(X))^2 \mid X = x\right] \leq V$.
A5 The density f is twice differentiable and its second derivatives are bounded on $\mathrm{I}R^d$.
A6 The kernel K is Lipschitz of ratio L_K that is $|K(x) - K(y)| \leq L_K \| x - y \|^k$.

2. Main Result

Theorem 2.1. *We suppose that the assumptions A1 to A6 hold. we further assume that the function r is Lipschitz, bounded on* I-.17em R^d *and that the bandwith sequence* (h_n) *satisfies with* y_n :

$$nh_n^d y_n^{-1} = \infty(Logn) \quad and \quad \gamma_n^{-1} h_n^{-d} y_n^{-b/2} \to 0, \quad n \to \infty$$

where y_n *is an unbounded and nondecreasing sequence chosen so that:*
$1 \leq y_n \leq n/2.$

If the kernel K *is even with* $\int z^2 K(z) dz < \infty$ *for* $z = (z_1, ..., z_d)$ *and if there exists a constant D such that* $\gamma_n^{-1} y_n h_n < D$. *Then we have:*

$$\sup_{\|x\| \leq c_n} |r_n(x) - r(x)| \to 0, \quad a.s. \quad n \to \infty.$$

3. Preliminary Results

We make the following decomposition:

$$r_n(x) - r(x) = \{f(x)\}^{-1} \{[g_n(x) - r(x) f(x)] - r_n(x)[f_n(x) - f(x)]\}$$

where $g_n(x) = (nh_n)^{-1} \sum_{t=1}^{n} Y_t K(h_n^{-1}(x - X_t))$ and

$$f_n(x) = (nh_n)^{-1} \sum_{t=1}^{n} K(h_n^{-1}(x - X_t)).$$

This leads to:

$$\sup_{x \in C_n} |r_n(x) - r(x)| \leq \{f(x)\}^{-1} \{\sup_{x \in C_n} |g_n(x) - r(x)| + \sup_{x \in C_n} |r_n(x)| |f_n(x) - f(x)|\}$$

Then if $\sup_{x \in C_n} |r_n(x)| \leq y_n \quad a.s.$ we obtain:

$$\sup_{x \in C_n} |r_n(x) - r(x)| \leq \gamma_n^{-1} \{\sup_{x \in C_n} |g_n(x) - r(x)| + y_n \sup_{x \in C_n} |f_n(x) - f(x)|\}$$

Lemma 3.1. *Under the assumptions of Theorem 2.1 we have:*

$$\gamma_n^{-1} \sup_{\|x\| \leq c_n} |g_n(x) - Eg_n(x)| \to 0, \ a.s. \quad n \to \infty.$$

Lemma 3.2. *Under the assumptions of Theorem 2.1, we have:*

$$\gamma_n^{-1} \sup_{x \in IR^d} | Eg_n(x) - r(x) f(x) | \to 0, \quad n \to \infty.$$

Lemma 3.3. *Under the assumptions of Theorem 2.1, we have:*

$$\lim_{n \to \infty} y_n \gamma_n^{-1} \sup_{||x|| \le c_n} | f_n(x) - Ef_n(x) | = 0 \quad a.s.$$

Lemma 3.4. *Under the assumptions of Theorem 2.1, we have:*

$$y_n \gamma_n^{-1} \sup_{x \in IR^d} | Ef_n(x) - f(x) | \to 0, \quad n \to \infty.$$

4. Proofs

Proof of Lemma 3.1.

Because of the possible large values for the variables Y_t , we use a truncation technique which consists in decomposing g_n in g_n^+ and g_n^- where

$g_n(x) = (nh_n)^{-1} \sum_{t=1}^n Y_{t(|Y_t| > y_n)} K(h_n^{-1}(x - X_t))$ and

$g_n^-(x) = g_n(x) - g_n^+(x),$ where y_n is the unbounded sequence defined in the Theorem 2.1.

We start by showing that:

$$\gamma_n^{-1} \sup_{||x|| \le c_n} | g_n^-(x) - Eg_n^-(x) | \to 0, \; a.s. \; n \to \infty.$$

We write $g_n^-(x) - Eg_n^-(x) = \sum_{t=1}^n \psi_t$ with

$$\psi_t = (nh_n^d)^{-1} \{ Y_{t(|Y_t| \le y_n)} K(h_n^{-1}(x - X_t)) - E[Y_{t(|Y_t| \le y_n)} K(h_n^{-1}(x - X_t))] \}.$$

Then $E(\psi_t) = 0;$ $\quad |\psi_t| \le (nh_n^d)^{-1} 2K_1 y_n = d_n$ where K_1 is an upperbound of K, which permits to write

$$E|\psi_t| \le 2n^{-1} E[|Y_t| h_n^{-d}{}_{(|Y_t| \le y_n)} K(h_n^{-1}(x - X_t))] \le$$

$$2n^{-1} E[|Y_t| h_n^{-d} K(h_n^{-1}(x - X_t))].$$

$$E|\psi_t| \le 2\Gamma n^{-1} \int E[|Y_t| / X_t = u] h_n^{-d} K(h_n^{-1}(x - u)) du.$$

Leading by Schwartz inequality and the assumption A4 to :

$$E|\psi_t| \le 2\Gamma n^{-1} \int (r^2(u) + V)^{1/2} h_n^{-d} K(h_n^{-1}(x - u)) du \le \tau n^{-1} = \delta_n.$$

Where τ is a positive constant. Same arguments give

$$E\psi_t^2 \le 2\Gamma n^{-2}\int (r^2(u)+V)h_n^{-2d}K(h_n^{-1}(x-u))du \le \upsilon n^{-2}h_n^{-d} = D_n$$

where υ is a positive constant.

Now, let us write

$$\sum_{n=1}^{\infty}P(\gamma_n^{-1}\,|\,g_n^-(x)-Eg_n^-(x)\,|>\varepsilon)=\sum_{n=1}^{\infty}P(\gamma_n^{-1}\,|\sum_{t=1}^{n}\psi_t\,|>\varepsilon)$$

and we write

$W_{nt}=\psi_{t(|\psi_t|\le n^\alpha)}$ and $Z_{nt}=\psi_{t(|\psi_t|>n^\alpha)}$ for $\alpha>1$ and $1\le t\le n$.

Then,

$$|\sum_{t=1}^{n}\psi_t\,|\le|\sum_{t=1}^{n}(W_{nt}-EW_{nt})\,|+|\sum_{t=1}^{n}Z_{nt}\,|+|\sum_{t=1}^{n}EW_{nt}\,|\,(3.1)$$

We need to show the followings

$$\sum_{n=1}^{\infty}P(\gamma_n^{-1}\,|\sum_{t=1}^{n}(W_{nt}-EW_{nt})\,|>\varepsilon n^\alpha/3)<\infty(3.2)$$

$$\sum_{n=1}^{\infty}P(\gamma_n^{-1}\,|\sum_{t=1}^{n}Z_{nt}\,|>\varepsilon n^\alpha/3)<\infty(3.3)$$

$$\gamma_n^{-1}\,|\sum_{t=1}^{n}EW_{nt}\,|/n^\alpha\to 0,\ \ n\to\infty.(3.4)$$

We start by showing (3.2).

The Markov inequality and Chebyshev's inequality lead to:

$$\sum_{n=1}^{\infty}P(\gamma_n^{-1}\,|\sum_{t=1}^{n}(W_{nt}-EW_{nt})\,|>\varepsilon n^\alpha/3)\le$$

$$c_1\sum_{n=1}^{\infty}\sum_{t=1}^{n}\gamma_n^{-1}E\,|\,W_{nt}\,|^b\,/n^{\alpha b}\le c_2\sum_{n=1}^{\infty}\gamma_n^{-1}h_n^{-d}n^{-1-\alpha b}<\infty.$$

with the choice $\gamma_n=n^{-a}$ for $a>0$ and $h_n=n^{-\tau}$ for $0<\tau<1/2$ where c_1 and c_2 are two positive constants and b such that $b>(a+\tau d)/\alpha$.

The Borel Cantelli lemma permits to conclude for (3.2).

Now, we show (3.3).

Note that $(|\sum_{t=1}^{n}Z_{nt}\,|>\varepsilon n^\alpha/3)\subset\bigcup_{t=1}^{n}(|\,\psi_t\,|>n^\alpha)$ hence,

$$\sum_{n=1}^{\infty}P(\gamma_n^{-1}\,|\sum_{t=1}^{n}Z_{nt}\,|>\varepsilon n^\alpha/3)\le\sum_{n=1}^{\infty}nP(\gamma_n^{-1}\,|\,\psi_t\,|>n^\alpha)$$

$$\le\sum_{n=1}^{\infty}n\gamma_n^{-1}E\,|\,\psi_t\,|^b\,/n^{\alpha b}\le c_3\sum_{n=1}^{\infty}n^{-1-\alpha b}\gamma_n^{-1}h_n^{-d}<\infty$$

with the choice $\gamma_n = n^{-a}$ for $a > 0$ and $h_n = n^{-\tau}$ for $0 < \tau < 1/2$; b such that $b > (a + \tau d)/\alpha$ and where c_3 is a positive constant.

Lastly we show that (3.4) holds.

We can write:

$$\gamma_n^{-1} n^{-\alpha} \,|\sum_{t=1}^{n} EW_{nt}| \le \gamma_n^{-1} n^{-\alpha} \sum_{t=1}^{n} |EZ_{nt}| = \gamma_n^{-1} n^{-\alpha} \sum_{t=1}^{n} E|\psi_t|_{(|\psi_t| > n^{\alpha})}$$

$$= n^{a-\alpha} E|\psi_t|_{(|\psi_t| > n^{\alpha})} \to 0, \quad n \to \infty \text{ with } \alpha > a > 0.$$

Next, we cover C_n by μ_n^d spheres in the shape of $\{x : \| x - x_{jn} \| \le c_n \mu_n^{-1}\}$ with $1 \le j \le \mu_n^d$ and we make the folowing decomposition:

$$|g_n^-(x) - Eg_n^-(x)| \le |g_n^-(x) - g_n^-(x_{jn})| + |g_n^-(x_{jn}) - Eg_n^-(x_{jn})|$$
$$+ |Eg_n^-(x_{jn}) - Eg_n^-(x)|$$

then we have

$$|g_n^-(x) - g_n^-(x_{jn})| \le$$
$$(nh_n^d)^{-1} y_n \sum_{t=1}^{n} |K(h_n^{-1}(x - X_t)) - K(h_n^{-1}(x_{jn} - X_t))|.$$

The kernel K being Lipschitz we obtain

$$|g_n^-(x) - g_n^-(x_{jn})| \le L_K y_n h_n^{-d-k} \| x - x_{jn} \|^k \le$$
$$L_K y_n h_n^{-d-k} c_n^k \mu_n^{-k} = 1/Logn$$

if we choose $\mu_n = L_K^{1/k} y_n^{1/k} h_n^{-(d/k+1)} c_n (Logn)^{1/k} \to \infty.$

Thus we obtain

$$\sup_{x \in C_n} |g_n^-(x) - Eg_n^-(x)| \le \sup_{1 \le j \le \mu_n^d} |g_n^-(x_{jn}) - Eg_n^-(x_{jn})| + (2/Logn)$$

so that for all $n \ge n_1(\varepsilon_n),\ \forall \varepsilon_n > 0$ and we have

$$P(\sup_{x \in C_n} |\textstyle\sum_{t=1}^{n} \psi_t| > 2\varepsilon_n) \le \sum_{j=1}^{\mu_n} P(|g_n^-(x_{jn}) - Eg_n^-(x_{jn})| > \varepsilon_n).$$

Now using similar decomposition as in (3.1) μ_n times; the use of ($\mu_n^d \quad n^{\alpha}$) instead of n^{α} permit to conclude that :

$$\gamma_n^{-1} \sup_{\|x\| \le c_n} |g_n^-(x) - Eg_n^-(x)| \to 0, \qquad a.s., \qquad n \to \infty.$$

It remains to show that:

$$\gamma_n^{-1} \sup_{\|x\| \le c_n} |g_n^+(x) - Eg_n^+(x)| \to 0, \qquad a.s., \qquad n \to \infty.$$

To this aim, we write:

$$\gamma_n^{-1} \sup_{\|x\|\le c_n} | g_n^+(x) - Eg_n^+(x) | \le E_n + F_n$$

where,

$$E_n = (\gamma_n n h_n^d)^{-1} \sup_{\|x\|\le c_n} | \sum_{t=1}^n Y_{t(|Y_t|> y_n)} K(h_n^{-1}(x - X_t)) | .$$

We have $(E_n \ne o) \subset \{\exists t_0 \in [1,2,...,n]$ such that $|Y_{t_0}|> y_n\}$ and we can write

$$(E_n \ne 0) \subset \bigcup_{t=1}^n \{|Y_t|> y_n\}$$

$$P(E_n \ne 0) \le \sum_{t=1}^n P(|Y_t|> y_n) = nP(|Y|> y_n)$$

$$\sum_n P(E_n \ne 0) \le \sum_n nP(|Y|> y_n) \le \sum_n n y_n^{-b} E|Y|^b$$

$$\sum_n P(E_n \ne 0) \le c_4 \sum_n n y_n^{-b} < \infty$$

where c_4 is a positive constant.

Then $E_n \to 0, \quad a.s., \quad n \to \infty$ and $\sup_{1\le t\le n} \sup |Y_t| \le y_n \quad a.s.$

The kernel K being strictly positive, we conclude that $|r_n(x)| \le y_n \quad a.s.$

Moreover,

$$F_n = (\gamma_n n h_n^d)^{-1} \sup_{\|x\|\le c_n} | \sum_{t=1}^n E[Y_{t(|Y_t|> y_n)} K(h_n^{-1}(x - X_t))] |$$

$$F_n = \gamma_n^{-1} h_n^{-d} K_1 E[|Y|_{(|Y_t|> y_n)}]$$

$$\le \gamma_n^{-1} h_n^{-d} K_1 (E(Y^2))^{1/2} (P[|Y|> y_n])^{1/2}$$

$$\le c_5 \gamma_n^{-1} h_n^{-d} y_n^{-b/2} \to 0, \quad n \to \infty$$

with c_5 being a positive constant.

Proof of Lemma 3.2.

$$Eg_n(x) - r(x)f(x) =$$

$$(nh_n^d)^{-1} E\left\{\sum_{t=1}^n Y_t K(h_n^{-1}(x - X_t))\right\} - r(x)f(x)$$

$$Eg_n(x) - r(x)f(x) = h_n^{-d} \int_{\mathrm{IR}^d} r(u) K(h_n^{-1}(x-u)) f(u) du - r(x)f(x)$$

we write $z = (x-u)/h_n$ and we obtain:

$$Eg_n(x) - r(x)f(x) = \int_{\mathrm{IR}^d} [r(x - zh_n) - r(x)] K(z) f(x - zh_n) dz +$$

$$r(x) \int K(z) [f(x - zh_n) - f(x) \quad] dz.$$

If we assume that the function $r(.)$ is Lipschitz of ratio 1 and order 1 we get

$$|\int_{IR^d}[r(x-zh_n)-r(x)]K(z)f(x-zh_n)dz| \leq h_n\Gamma\int|z|K(z)dz.$$

Now a Taylor expansion , the Bochner lemma and the fact that the function r is bounded permit to conclude that:

$$\gamma_n^{-1}\sup_{x\in IR^d}|Eg_n(x)-r(x)f(x)| \to 0, \quad n\to\infty.$$

Proof of Lemma 3.3.

This is a particular case of Lemma 3.1 when $Y_t = 1$ and $\varepsilon = \varepsilon_0\gamma_n y_n^{-1}$ for a certain $\varepsilon_0 > 0.$

Proof of Lemma 3.4.

We write:

$$Ef_n(x)-f(x) = h_n^{-d}\int_{IR^d}[f(u)-f(x)]K(h_n^{-1}(u-x))du$$

A Taylor expansion, the hypotheses of Theorem 2.1 and the Bochner lemma permit to conclude.

Proof Theorem 2.1.

Lemmas 3.1, 3.2, 3.3 and 3.4 permit to conclude that:

$$\sup_{x\in C_n}|r_n(x)-r(x)| \to 0, \quad a.s. \quad n\to\infty.$$

Acknowledgements

The author would like to thank the referees for their relevant remarks and criticisms.

References

1. Arfi, M. (1996). Sur la Régression Nonparametrique d'un Processus Stationnaire Mélangeant ou Ergodique. Thèse de Doctorat de l 'Université Paris 6.
2. Arfi, M. (1997). On General Estimate of the Regression Function. Statisitcs & Decisions, 15, 191-200.
3. Arfi, M. (2003). Convergence Rate for a Kernel Estimate of the Regression Function. Jour. of Mathematical Sciences, 14, 11-17.
4. Bosq, D. (1996). Nonparametric statistics for stochastic processes. Lecture Notes in Statistics, 110, Springer-Verlag.
5. Bradley, R. C. (1990). Equivalent Mixing Conditions of Random Fields. Technical Report No 336. Center for Stochastic Processes, University of North Carolina, Chapel Hill.
6. Bradley, R. C. (1992). On the Spectral Density and Asymptotic Normality of Weakly Dependent Random Fields. J. Theoret. Probab., 5, 355-374.

7. Bryc, W., Smolenski, W. (1993). Moment Conditions for Almost Sure Convergence of Weakly Correlated Random Variables. Proc. Amer. Math. Soc., 119, 629-635.
8. Collomb, G. (1984). Proprietés de Convergence Presque Sure du Predicteur à Noyau. Z. Wahrsch. Verw., Gebiete., 66, 441-460.
9. Collomb, G. (1985). Nonparametric Regression: An Up to Date Bibliography. Statistics, 16, 309-324.
10. Devroye, L. (1981). On the Almost Everywhere Convergence of Nonparametric Regression Function Estimates. The Annals of Statistics, 9(6), 1310-1319.
11. Györfi, L., Härdle, W., Sarda, P., Vieu, P. (1989). Nonparametric Curve Estimation from Time Series. Lecture Notes in Statistics, 60, Springer-Verlag.
12. Härdle, W. (1990). Applied Nonparametric Regression. Cambridge Univ. Press
13. Peligrad, M., Gut, A. (1999). Almost Sure Results for a Class of Dependent Random Variables. J. Theoret. Probab., 12, 87-104.
14. Nadaraya, E. A. (1964). On Estimating Regression. Theory of Probability and its Application, 9, 141-142.
15. Shixin, G. (2004). Almost Sure Convergence for $\widetilde{\rho}$-Mixing Random Variable Sequences. Statistics & Probability Letters, 67, 289-298.
16. Sung, S. H. (2010). Complete Convergence for Weighted Sums of $\widetilde{\rho}$-Mixing Random Variables. Hindawi Publishing Corporation, Discrete Dynamics in Nature and Society, Vol. 2010.
17. Walk, H. (2006). Almost sure convergence properties of Nadaraya-Watson regression estimates. International Series in Operations Research & Management Science, Vol. 46, Modeling Uncertainty, Part 4, 201-223.
18. Watson, G.S. (1964). Smooth Regression Analysis. Sankhya, A 26, 359-372.