International Journal of Statistics and Probability; Vol. 5, No. 2; 2016
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

On the Convergence Rate for a Kernel Estimate of the Regression Function

Mounir ARFI[1]

[1] University of Sharjah, College of Sciences, Department of Mathematics, Sharjah, United Arab Emirates

Correspondence: Mounir ARFI, University of Sharjah, College of Sciences, Department of Mathematics, P.O. Box 27272 Sharjah, United Arab Emirates. E-mail: marfi@sharjah.ac.ae / mounir.arfi@gmail.com

Received: December 6, 2015 Accepted: December 28, 2015 Online Published: February 10, 2016

doi:10.5539/ijsp.v5n2p29 URL: http://dx.doi.org/10.5539/ijsp.v5n2p29

Abstract

We give the rate of the uniform convergence for 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 when the observed process is φ-mixing. The used estimator for the regression function is the kernel estimator proposed by Nadaraya, Watson (1964).

Keywords: Kernel estimate, φ-Mixing, regression

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 on values in $\mathbb{R}^d \mathrm{x} \mathbb{R}$ and 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 rate of convergence for a kernel estimate of the regression function, known as:

$$r(x) = E(Y_t|X_t = x) \qquad t \in \mathbb{N}.$$

A natural estimator of the function $r(.)$ is given by

$$r_n(x) = \frac{\sum_{t=1}^{n} Y_t K\left(\frac{x - X_t}{h_n}\right)}{\sum_{t=1}^{n} K\left(\frac{x - X_t}{h_n}\right)} \qquad \forall x \in E$$

Where E stands for the subset $\{x \in \mathbb{R}, f(x) > 0\}$, f being the density of the process (X_t) and (h_n) is a positive sequence of real numbers such that $h_n \to 0$ and $nh_n^d \to \infty$ when $n \to \infty$.

K is a Parzen-Rosenblatt kernel type in the sense of a bounded function satisfying

$$\int_{\mathbb{R}} K(x)dx = 1 \qquad \text{and} \qquad \lim_{||x|| \to \infty} ||x||K(x) = 0$$

Moreover, it is assumed to be strictly positive and with bounded variation.

The estimation of the regression function has been subject to several investigations, and many authors have been involved. Among others, Devroye (1981), Collomb (1984, 1985), Györfy *et al.* (1989), Härdle (1990), Bosq (1996), Arfi (1996), Arfi (1997) 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.

Gasser *et al.* (1984) introduced a kernel estimate for obtaining a nonparametric estimate of a regression function and its derivatives, Sara Van De Ger (1990) proposed an entropy approach to establish rates of convergence for estimators of a regression function and later on Hermann and Ziegler (2004) studied the rates of consistency for a nonparametric estimation of the mode in absence of smoothness assumptions.

Our work is devoted to the rate of the uniform convergence for a kernel estimate of the regression function over an increasing sequence of compact sets under a mixing condition.

2. Preliminaries and Assumptions

We assume that the process $(X_t)_{t\in\mathbb{N}}$ is stationary and φ-mixing that is

$$\varphi_n = \sup_{A\in\mathcal{M}_0^t}\ \sup_{B\in\mathcal{M}_{t+n}^\infty} \{|P(B/A) - P(A)|\} \longrightarrow 0, \quad n \to \infty$$

where $\mathcal{M}_0^t$ is the σ-field generated by $\{X_0, X_1, ..., X_t\}$ and $\mathcal{M}_{t+n}^\infty$ is the σ-field generated by $\{X_{t+n}, X_{t+1+n}, ...\}$

We will make use of the following assumptions

A1. $\exists\Gamma < \infty, \quad \forall x \in \mathbb{R}^d, \quad f(x) \leq \Gamma$

and

$\exists\gamma_n > 0, \quad \forall x \in C_n, \quad f(x) \geq \gamma_n$

where C_n is a sequence of compact sets such that $C_n = \{x : \|x\| \leq c_n\}$ and $c_n \to \infty$

A2. $\exists\beta \geq 2, \ \exists M < \infty$ such that $E(|Y|^\beta) \leq M$

A3. $\exists V < \infty, \quad \forall x \in \mathbb{R}^d, \quad E[(Y - r(x))^2|X = x] \leq V$

A4. The density f is twice differentiable and its second derivatives are bounded on $\mathbb{R}^d$

A5. The kernel K is Lipschitz of ratio L_k that is $|K(x) - K(y)| \leq L_k\|x - y\|^{\gamma_1}$

3. Main Result

Theorem

Assuming that the assumptions A1 through A5 hold, we further assume that the function r is Lipschitz, bounded on $\mathbb{R}^d$ and that the bandwith sequence (h_n) satisfies with y_n:

$$n^\delta\gamma_n^{-1}h_n^{-d}y_n^{-\beta} \longrightarrow \infty \quad n \to \infty$$

$$\forall\epsilon_0 > 0 \quad \sum_n \left\{\frac{c_n^d y_n^{d/\gamma_1}(logn)^{d/\gamma_1}}{h_n^{d(1+d/\gamma_1)}}\right\} \exp\left(-\epsilon_0^2\frac{n^{1-2\delta}\gamma_n^2 h_n^d}{m_n y_n}\right) < \infty$$

Where $\delta \in]0, 1/2[$, m_n and y_n are two sequences such that:

$$1 \leq m_n \leq n/2 \ \text{ and } \ 1 \leq y_n \leq \sqrt{n/2}$$

If the kernel K is even with $\int z^2K(z)dz < \infty$ for $z = (z_1, ..., z_d)$ and if there exists a constant D such that $\gamma_n^{-1}y_n n^\delta h_n^d < D$ then if r is continuous, Lipschitz and bounded on $\mathbb{R}$ we have:

$$n^\delta \sup_{\|x\|\leq c_n} |r_n(x) - r(x)| = O(1) \quad a.s. \quad n \to \infty.$$

4. Preliminary Results

For practical reason, we make the following decomposition:

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

where $g_n(x) = \dfrac{1}{nh_n^d}\displaystyle\sum_{t=1}^n Y_tK\left(\frac{x - X_t}{h_n}\right)$

and $f_n(x) = \dfrac{1}{nh_n^d}\displaystyle\sum_{t=1}^n K\left(\frac{x - X_t}{h_n}\right)$

This leads to

$$\sup_{x\in C_n}|r_n(x) - r(x)| = \frac{1}{f(x)}\left\{\sup_{x\in C_n}|g_n(x) - r(x)f(x)| + \sup_{x\in C_n}|r_n(x)||f_n(x) - f(x)|\right\}$$

Then if

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

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

The following Lemma will be used in our proofs

Lemma (Collomb, 1984)

Let Z_t be a real centered and φ-mixing process such tha:

$$|Z_t| \leq d_n \qquad E(Z_t^2) \leq D_n \qquad E|Z_t| \leq \delta_n$$

then, $\forall\epsilon > 0$, $\forall n \in \mathbb{N}^*$ we have:

$$P\left\{|\sum_{t=1}^{n} Z_t| > \epsilon\right\} 2\exp\left\{-\alpha\epsilon + 3\sqrt{e}n\frac{\varphi_m}{m} + 6\alpha^2 n\left[D_n + 6\delta_n d_n\sum_{j=1}^{m}\varphi_j\right]\right\}$$

Lemma 1

Under the hypotheses of Theorem we have:

$$\gamma_n^{-1}n^{\delta}\sup_{x\in C_n}|g_n(x)-Eg_n(x)| \to 0 \quad a.s. \quad n\to\infty.$$

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

$$g_n^+(x) = \frac{1}{nh_n^d}\sum_{t=1}^{n} Y_t\mathbb{I}_{[|Y_t|>y_n]}K\left(\frac{x-X_t}{h_n}\right)$$

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

We start by showing that:

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

To this end, we write:

$$g_n^-(x) - Eg_n^-(x) = \sum_{t=1}^{n}\varphi_t \qquad \text{with}$$

$$\varphi_t = \frac{1}{nh_n^d}\left\{Y_t\mathbb{I}_{[|Y_t|\leq y_n]}K\left(\frac{x-X_t}{h_n}\right) - E\left[Y_t\mathbb{I}_{[|Y_t|\leq y_n]}K\left(\frac{x-X_t}{h_n}\right)\right]\right\}$$

therefore $E(\varphi_t) = 0$;

$|\varphi_t| \leq \dfrac{2K_1 y_n}{nh_n^d} = d_n$ where K_1 is an upperbound of K, which permits to write:

$$E|\varphi_t| \leq \frac{2\Gamma}{n}E\left|\frac{Y_t}{h_n^d}\mathbb{I}_{[|Y_t|\leq y_n]}K(\frac{x-X_t}{h_n})\right| \leq \frac{2\Gamma}{n}\int\frac{E(|Y_t|/X_t=u)}{h_n^d}K\left(\frac{x-u}{h_n}\right)du$$

Leading by Schwartz inequality and the assumption A3 to:

$$E|\varphi_t| \leq \frac{2\Gamma}{n}\int\frac{(r^2(u)+V)^{1/2}}{h_n^d}K\left(\frac{x-u}{h_n}\right)du \leq \tau_1 n^{-1}$$

where τ_1 is a positive constant.

Now, same arguments give:

$$E(\varphi_t)^2 \leq \frac{2\Gamma}{n^2}\int\frac{(r^2(u)+V)}{h_n^{2d}}K\left(\frac{x-u}{h_n}\right)du \leq \upsilon n^{-2}h_n^{-d}$$

where υ is a positive constant.

We apply the Collomb inequality with $\alpha = 1/(4m_n d_n)$ and we obtain for all $\epsilon_n > 0$:

$$P\left(\left|g_n^-(x) - Eg_n^-(x)\right| > \epsilon_n\right) \leq C_1\exp\left(-C_2\epsilon_n^2\frac{nh_n^d}{m_n y_n}\right)$$

where C_1 and C_2 are two positive constants.

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

And we make the following decomposition:

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

and we have

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

The kernel K being Lipschitz we obtain

$$|g_n^-(x) - g_n^-(x_{jn})| \le L_K \frac{y_n}{h_n^{d+\gamma_1}} \|x - x_{jn}\|^{\gamma_1}$$

$$|g_n^-(x) - g_n^-(x_{jn})| \le L_K \frac{y_n}{h_n^{d+\gamma_1}} c_n^{\gamma_1} \mu_n^{-\gamma_1}$$

$$|g_n^-(x) - g_n^-(x_{jn})| \le \frac{1}{logn}$$

If we choose

$$\mu_n = L_K^{1/\gamma_1} \frac{y_n^{1/\gamma_1} c_n (logn)^{1/\gamma_1}}{h_n^{d/\gamma_1 + 1}} \longrightarrow \infty.$$

Thus we obtain:

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

Therefore, if we apply μ_n^d times the Lemma of Collomb, we obtain

$$P\left(\sup_{x \in C_n} |g_n^-(x) - Eg_n^-(x)| > 2\epsilon_n\right) \le C_1 \mu_n^d \exp\left(-C_2 \epsilon_n^2 \frac{nh_n^d}{m_n y_n}\right)$$

Now if we choose $\epsilon_n = n^{-\delta} \gamma_n \epsilon_0$ for a certain $\epsilon_0 > 0$, we obtain accordingly with the hypotheses of the Theorem:

$$P\left(\gamma_n^{-1} n^\delta \sup_{\|x\| \le c_n} |g_n^-(x) - Eg_n^-(x)| > 2\epsilon_0\right) \le C_1 L_K^{d/\gamma_1} \frac{y_n^{d/\gamma_1} c_n^d (logn)^{d/\gamma_1}}{h_n^{d(1+d/\gamma_1)}} \exp\left(-C_2 \epsilon_0^2 \frac{n^{1-2\delta} \gamma_n^2 h_n^d}{m_n y_n}\right)$$

The hypotheses of the Theorem permit to conclude that:

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

It remains to show that:

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

For practical reason, we write:

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

Where,

$$E_n = \frac{n^\delta \gamma_n^{-1}}{nh_n^d} \sup_{\|x\| \le c_n} \left| \sum_{t=1}^{n} Y_t \mathbb{I}_{(|Y_t| > y_n)} K\left(\frac{x - X_t}{h_n}\right)\right|$$

And we have

$$(E_n \neq 0) \subset \{\exists\, t_0 \in [1, 2, 3, ..., n] \text{ such that } |Y_{t_0}| > y_n\}$$

the above leads to

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

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

$$\sum_n P(E_n \neq 0) \leq \sum_n P(|Y_t| > y_n) \leq \sum_n n y_n^{-\beta} E|Y|^{\beta}$$

$$\sum_n P(E_n \neq 0) \leq c_4 \sum_n n y_n^{-\beta} < \infty$$

where c_4 is a positive constant.

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

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

Moreover,

$$F_n = \frac{n^{\delta}}{\gamma_n n h_n^d} \sup_{\|x\| \leq c_n} \left| \sum_{t=1}^{n} E\left[Y_t \mathbb{I}_{[|Y_t| > y_n]} K\left(\frac{x - X_t}{h_n} \right) \right] \right|$$

$$F_n \leq \frac{n^{\delta}}{\gamma_n h_n^d} K_1 E[|Y| \mathbb{I}_{|Y| > y_n}]$$

$$F_n \leq \frac{n^{\delta}}{\gamma_n h_n^d} K_1 (E(Y^2))^{1/2} (P[|Y| > y_n])^{1/2} \leq c_5 n^{\delta} \gamma_n^{-1} h_n^{-d} y_n^{-\beta/2} \to 0, \quad n \to \infty$$

where c_5 is being a positive constant.

Lemma 2

Under the assumptions of the Theorem, we have:

$$n^{\delta} \gamma_n^{-1} \sup_{x \in \mathbb{R}^d} |E g_n(x) - r(x) f(x)| \to 0, \quad n \to \infty.$$

Proof:

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

$$E g_n(x) - r(x) f(x) = \frac{1}{h_n^d} \int_{\mathbb{R}^d} r(u) K\left(\frac{x - u}{h_n} \right) f(u) du - r(x) f(x)$$

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

$$E g_n(x) - r(x) f(x) = \int_{\mathbb{R}^d} [r(x - z h_n) - r(x)] K(z) f(x - z h_n) dz + r(x) \int K(z) [f(x - z h_n) - f(x)] dz$$

Assuming that the function $r(.)$ is Lipschitz of ratio 1 and order 1 provides

$$\left| \int_{\mathbb{R}^d} [r(x - z h_n) - r(x)] K(z) f(x - z h_n) dz \right| \leq h_n \Gamma \int |Z| K(z) dz$$

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

$$n^{\delta} \gamma_n^{-1} \sup_{x \in \mathbb{R}^d} |E g_n(x) - r(x) f(x)| \to 0, \quad n \to \infty.$$

Lemma 3

Under the assumptions of the Theorem, we have:

$$\lim_{n \to \infty} \frac{y_n n^{\delta}}{\gamma_n} \sup_{\|x\| \leq c_n} |f_n(x) - E f_n(x)| = 0 \ \ a.s.$$

Proof:

This is a particular case of Lemma 1 when $Y_t = 1$ and $\epsilon = \epsilon_0 \gamma_n n^{-\delta} y_n^{-1}$ for a certain $\epsilon_0 > 0$.

Lemma 4

Under the assumptions of the Theorem, we have:

$$\lim_{n\to\infty} \frac{y_n n^{\delta}}{\gamma_n} \sup_{x\in\mathbb{R}^d} |Ef_n(x) - f(x)| = 0$$

Proof:

We write

$$Ef_n(x) - f(x) = \frac{1}{h_n^d} \int [f(u) - f(x)] K\left(\frac{u-x}{h_n}\right) du$$

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

5. Proof of the Theorem

The lemma 1, Lemma 2, Lemma 3, and Lemma 4 permit to conclude.

Acknowledgments

The author would like to thank the referees for their comments and criticisms about this paper.

References

Arfi, M. (1996). *Sur la Regression non parametrique d' un processus stationnaire mélangeant ou ergodique*. Doctorate Thesis of University Paris 6.

Arfi, M. (1997). On General Estimate of the Regression Function. *Statistics & Decisions, 15*, 191-200. http://dx.doi.org/10.1524/strm.1997.15.2.191

Bosq, D. (1996). Nonparametric statistics for stochastic processes. *Lecture Notes in Statistics, 110*, Springer Verlag. http://dx.doi.org/10.1007/978-1-4684-0489-0

Collomb, (1984). Proprietes de Convergence Presque Sure du Predicteur a Noyau. *Z. Wahrsch. Verw., Gebiete., 66*, 441-460. http://dx.doi.org/10.1007/BF00533708

Collomb, G. (1985) Nonparametric regression; an up to date bibliography. *Statistics, 16*, 309-324. http://dx.doi.org/10.1080/02331888508801860

Devroye, L. (1981). On the Almost Eveywhere Convergence of Nonparametric Regression Function Estimates. *The Annals of Statistics, 9*, 1310-1319. http://dx.doi.org/10.1214/aos/1176345647

Gasser, T., & Hans-Georg Müller. (1984). Estimating Regression Functions and Their Derivatives by the Kernel Method. *Scand. J. Statist., 11*, 171-185.

Györfy, L., Härdle, W., Sarda, P., & Vieu, P. (1989). Nonparametric Curve Estimation from Time Series. *Lecture Notes in Statistics, 60*, Springer Verlag.

Härdle, W. (1990). Applied Nonparametric Regression. Cambridge Univ. Press.

Hermann, E., & Ziegler, K. (2004). Rates of Consistency for a Nonparametric Estimation of the Mode in Absence of Smoothness Assumptions. *Statistics & Probability Letters, 68*, 359-368. http://dx.doi.org/10.1016/j.spl.2004.04.005

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Application, 9*, 141-142. http://dx.doi.org/10.1137/1109020

Sara Van De Ger. (1990). Estimating a Regression Function. *The Annals of Statistics, 18*(2), 907-924. http://dx.doi.org/10.1214/aos/1176347632

Walk, H. (2006). Almost sure Convergence Properties of Nadaray-Watson Regression Estimates. *International Series in Operations Research & Management Science, 46*, Modeling Uncertainty, 201-223.

Watson, G. S. (1964) Smooth regression analysis. *Sankhya, A 26*, 359-372.