Int. J. Open Problems Compt. Math., Vol 2, No. 1, March 2009

Nonparametric Prediction via Mode

MOUNIR ARFI
58 Corrour Road, Aviemore PH22 1SS, Inverness-Shire, Scotland.
Email: mounir.arfi@hotmail.co.uk

Abstract

It is shown that the (empirically determined) mode of the kernel estimate is uniformly convergent to the conditional mode function under the ergodic condition over a sequence of compact sets which increases to $\mathbf{R}^d$.

Key words: *Kernel density estimate; conditional mode; ergodicity, sequence of compact sets.*

1. Introduction

Let $\{(X_i, Y_i)\}_{i\in \mathrm{N}}$ be a stationary process where (X_i, Y_i) take values in $\mathrm{R}^d \times \mathrm{R}$ and distributed as (X, Y). Suppose that a segment of data $\{(X_i, Y_i)\}_{i=1}^n$ has been observed. We are interested in predicting Y from the data for a fixed value of X.

Such an approach has been investigated by several authors when the observed data are i.i.d. or when the process is mixing (see the surveys by Collomb [5] and Györfi *et al.* [7]).

However, we know that if the conditional distribution of Y given X has a dominant center peak and a smaller peak far from the center, then it is more reasonable to consider the conditional mode function.

The objective of this paper is to investigate the estimation of the conditional mode function, assuming that it is uniquely defined. Also, to establish the uniform almost sure convergence for the estimate of the conditional mode function, obtained from the conditional density under the ergodic hypothesis, which is more general than the i.i.d. case or even mixing situations over a sequence of compact sets which increases to R^d.

On the other hand, most of the results suppose that the data belong to a fixed compact set, this is rather cumbersome for the applications. In our paper we deal with sequences belonging to a sequence of compact sets which increases to R^d.

Such a subject has been studied by many authors, among others, Parzen [9] who studied the estimation of a probability density function and mode, Collomb& *al.* [6] considered the case of the conditional mode function, Arfi [2] used the mode function to investigate the prediction and Hermann & Ziegler [8] proposed rates of consistency for a nonparametric estimation of the mode in absence of smoothness assumptions.

The conditional mode is defined by means of the conditional density $f(y|x)$ of Y, given X, as follows: $\Theta(x) = \arg\max_{y \in R} f(y|x)$,

and the so-called *empirical mode predictor* is defined as the maximum of $f_n(y|x)$ over $y \in \mathrm{R}$, where $f_n(y|x)$ is the kernel estimate of $f(y|x)$ defined by:

$$f_n(y|x) = \frac{f_n(x,y)}{g_n(x)};$$

here $g_n(x) > 0$, is the kernel estimate of the density function of X, $g(x)$, and $f_n(x,y)$ is the kernel estimate of the joint density of the pair (X,Y), $f(x,y)$.

These kernel estimates are defined, respectively, as follows:

$$f_n(x,y) = \frac{1}{nh_n^{d+1}} \sum_{i=1}^{n} K_2\left(\frac{y - Y_i}{h_n}\right) K_1\left(\frac{x - X_i}{h_n}\right),$$

and

$$g_n(x) = \frac{1}{nh_n^d} \sum_{i=1}^{n} K_1\left(\frac{x - X_i}{h_n}\right);$$

here K_1 (K_2) are two Parzen-Rosenblatt kernels on R^d (R) with K_1 strictly positive and bounded variation, and K_2 compactly supported; h_n is a sequence of positive numbers such that: $h_n \to 0$ and $nh_n^{d+1} \to \infty$ when $n \to \infty$.

We show that the random function $\Theta_n(x) = \arg\max_{y \in R} f_n(y|x)$ converges uniformly over a sequence of compact sets C_n (which increases to R^d) to the mode function $\Theta(x)$.

2. Assumptions and Main Arguments

We denote by $\mathcal{F}_{i-1}$ and $\mathcal{G}_{i-1}$, the σ-fields generated by $\{(X_{i-j}, Y_{i-j})\ ;\ 1 \leq j < i\}$ and $\{X_{i-j}\ ;\ 1 \leq j < i\}$, respectively.

We assume the existence of the conditional densities $f_{X,Y}^{\mathcal{F}_{i-1}}(.,.)$ and $g_X^{\mathcal{G}_{i-1}}(.)$ of the variables (X, Y) and X with respect to $\mathcal{F}_{i-1}$ and $\mathcal{G}_{i-1}$.

It will be further assumed that $f(.,.)$ and $g(.) \in \mathrm{C}_0(\mathrm{R}^j)$, $j = d, d+1$ where $\mathrm{C}_0(\mathrm{R}^j)$ denotes the space of real-valued continuous functions on R^j tending to zero at infinity. The same assumption will be made for the conditional densities $f_{X,Y}^{\mathcal{F}_{i-1}}$ and $g_X^{\mathcal{G}_{i-1}}$.

Under the previous conditions, the Theorem in Beck [4] implies the following condition named the (T) condition:

$$T_{1,n} = \sup_{(x,y)\in \mathrm{R}^d\times\mathrm{R}} |n^{-1}\sum_{i=1}^{n} f_{X,Y}^{\mathcal{F}_{i-1}}(x,y) - f(x,y)| \xrightarrow{a.s.} 0,\ n \to \infty$$

$$T_{2,n} = \sup_{x\in\mathrm{R}^d} |n^{-1}\sum_{i=1}^{n} g_X^{\mathcal{G}_{i-1}}(x) - g(x)| \xrightarrow{a.s.} 0,\ n \to \infty,$$

for ergodic processes satisfying some further mild regularity conditions (Györfi *et al.* [7]).

In the sequel, we suppose that the (T) condition holds and we suppose that $T_{1,n} = o(n^{-a})$ and $T_{2,n} = o(n^{-a-1})$.

Moreover, the conditional densities $f_{X,Y}^{\mathcal{F}_{i-1}}(.)$ and $g_X^{\mathcal{G}_{i-1}}(.)$ are assumed to be Lipschitz, in the sense that:

$$|f_{X,Y}^{\mathcal{F}_{i-1}}(x,y) - f_{X,Y}^{\mathcal{F}_{i-1}}(x', y')| \leq ||(x,y) - (x', y')||_{\mathrm{R}^d\times\mathrm{R}},$$

$$|g_X^{\mathcal{G}_{i-1}}(x) - g_X^{\mathcal{G}_{i-1}}(x')| \leq ||x - x'||_{\mathrm{R}^d}.$$

We will also make use of the following assumptions:

A1. The process $(X_i, Y_i)_{i\in\mathrm{N}}$ is strictly stationary and ergodic

A2. The joint distribution $P_{(X,Y)}$ of the pair (X, Y) is absolutely continuous with regard to the Lebesgue measure on $\mathrm{R}^d \times \mathrm{R}$.

A3. There exists $a > 0$, such that $g(x) \geq n^{-a}$, $n \geq 1$, for all $x \in C_n$, where $C_n = \{x : ||x|| \leq c_n\}$ with $c_n \longrightarrow \infty$, $n \to \infty$.

A4. The kernels K_j, $j = 1, 2$ are Lipschitz of order $\gamma_1 > 0$, in the sense that: $\exists L_K < \infty \quad |K_j(u) - K_j(v)| \leq L_K|u - v|^{\gamma_1} \quad j = 1, 2.$

A5. K_j, $j = 1, 2$ are bounded and integrate to one.

A6. The mode function $\Theta(.)$ satisfies the following condition on a sequence of compact sets C_n:

$$\forall \epsilon_n > 0, \quad \exists \beta_n > 0, (\forall \zeta \quad C_n \to \mathrm{R}^d)$$

$$\text{if} \qquad \sup_{x \in C_n} |\Theta(x) - \zeta(x)| \geq \epsilon_n, \qquad \text{then} \qquad \sup_{x \in C_n} |f(\Theta(x)|x) - f(\zeta(x)|x)| \geq \beta_n.$$

A7. There exists $\xi > 2$ and $M < \infty$ such that $E|Y|^\xi < M$.

3. Main Result

Our main result is stated in the theorem below

Theorem

We suppose that the assumptions A1 to A7 hold. We further assume that the sequence h_n satisfies:

$$\lim_{n \to \infty} \frac{n h_n^{2(d+1)}}{\mathrm{Log} n} = \infty, \qquad n^{a+1} h_n^k \longrightarrow 0 \quad \text{for } a > 0 \ \text{ and } k \geq 3 \tag{1}$$

and

$$\forall \epsilon_n > 0, \qquad \sum_n \frac{n^{d(a+1)/\gamma_1} c_n^d (\mathrm{Log} n)^{1/\gamma_1}}{h_n^{d(d+1+\gamma_1)/\gamma_1}} h_n^{-\frac{1}{\mu} - \eta} \exp\{-\epsilon_n^2 n h_n^{2(d+1)}\} < \infty$$

for $\eta > 1 + (d+2)/\gamma_1 + (d+1)/\gamma_1^2$ *, with* $a > 0$ *and* μ *a positive constant. If the kernel* K_1 *is even with* $\int z^k K_1(z) dz < \infty$ *for* $k \geq 1$*, then*

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

Remarks

1) As sequences h_n and c_n we can choose $h_n = O(n^{-b})$ with $b < 1/2(d+1)$ and $c_n = O((Logn)^{1/\gamma_1})$.

2) In the ergodic case, there is no general theoretical result to determine the precise rate of convergence. The convergence can be arbitrarily fast.

4. Preliminary Results

$$\sup_{x \in C_n} \sup_{y \in \mathrm{R}} |f_n(y|x) - f(y|x)| \leq$$

$$\frac{1}{\inf\limits_{x\in C_n} g(x)} \times \left\{ \sup_{x\in C_n}\sup_{y\in \mathrm{R}} |f_n(x,y)-f(x,y)| + \sup_{x\in C_n}\sup_{y\in R}|f_n(y|x)||g_n(x)-g(x)| \right\} \leq$$

$$n^a \left\{ \sup_{x\in C_n}\sup_{y\in \mathrm{R}} |f_n(x,y)-f(x,y)| + \sup_{x\in C_n}\sup_{y\in R}|f_n(y|x)||g_n(x)-g(x)| \right\}$$

with

$$\sup_{y\in R}|f_n(y|x)| \leq \frac{\widetilde{K}}{h_n\overline{K}_1} \qquad \text{then} \qquad n^{-1}\sup_{y\in R}|f_n(y|x)| \leq \frac{\widetilde{K}}{nh_n\overline{K}_1} < M_1 < \infty$$

where $M_1$is a positive constant and $\widetilde{K} = \max\left\{ \sup\limits_{x\in R^d} K_1(x), \sup\limits_{y\in R} K_2(y), 1 \right\}$ $\overline{K}_1$ is an upperbound of K_1 and we can write

$$\sup_{x\in C_n}\sup_{y\in \mathrm{R}} |f_n(y|x)-f(y|x)| \leq n^a \sup_{x\in C_n}\sup_{y\in \mathrm{R}} |f_n(x,y)-f(x,y)| + M_1 n^{a+1} \sup_{x\in C_n} |g_n(x)-g(x)|$$

Definition

A process $(X_i)_{i\in \mathrm{N}}$ is called a martingale difference, if it is real valued and satisfies:

$$E(X_i|\mathcal{A}_{i-1}) = 0 \quad \forall i \in \mathrm{N}^*,$$

where $\mathcal{A}_{i-1}$ denotes the σ-field generated by the past of the process (X_i).

Lemma 1 (Azuma [3])

If $(X_i)_{i\in \mathrm{N}}$ is a martingale difference with $|X_i| \leq B$ a.s., then for all $\epsilon > 0$

$$P\left\{ |\sum_{i=1}^{n} X_i| > \epsilon \right\} \leq 2\exp\left\{ -\frac{\epsilon^2}{2nB^2} \right\}.$$

If $(X_i)_{i\in \mathrm{N}}$ is real-valued with $|X_i| \leq B$ a.s., then for all integers $m > 0$ such that $i - m > 0$ and all $\epsilon > 0$,

$$P\left\{ |\sum_{i=1}^{n} [X_i - E(X_i|\mathcal{F}_{i-m})]| > \epsilon \right\} < 2m\exp\left\{ -\frac{\epsilon^2}{2nm^2B^2} \right\}.$$

Lemma 2
Under assumptions A1 to A5, we have:

$$n^{a+1}\sup_{x\in C_n}|g_n(x)-g(x)| \xrightarrow{a.s.} 0, \quad n\to\infty.$$

Proof:
Consider the following decomposition:

$$g_n(x)-g(x)=\sum_{i=1}^{n} Z_i(x)+V_n(x)$$

with

$$Z_i(x)=\frac{1}{nh_n^d}\left\{K_1\left(\frac{x-X_i}{h_n}\right)-E^{\mathcal{G}_{i-1}}K_1\left(\frac{x-X_i}{h_n}\right)\right\},$$

$$\text{and} \qquad V_n(x)=\frac{1}{nh_n^d}\sum_{i=1}^{n}E^{\mathcal{G}_{i-1}}K_1\left(\frac{x-X_i}{h_n}\right)-g(x),$$

where, $E^{\mathcal{G}_{i-1}}(.)=E(.|\mathcal{G}_{i-1})$ and $\mathcal{G}_{i-1}=\sigma(X_{i-j};\ 1\leq j<i)$.

For fixed x, Z_i is a martingale difference with $|Z_i(x)|\leq \frac{B}{nh_n^d}$ *a.s.*, where B is a positive constant. Then, by applying Lemma 1, we obtain

$$P\left\{n^{a+1}|\sum_{i=1}^{n}Z_i(x)|>\epsilon\right\}=P\left\{|\sum_{i=1}^{n}Z_i(x)|>\epsilon_n\right\}\leq 2\exp\left\{-\frac{\epsilon_n^2}{2B^2}nh_n^{2d}\right\}; \qquad (2)$$

$\forall\epsilon>0$ *and* $\epsilon_n=\epsilon n^{-a-1}$ .The choice of h_n in the Theorem allows us to conclude that:

$$n^{a+1}|\sum_{i=1}^{n}Z_i(x)|\longrightarrow 0, a.s. \quad \text{when} \quad n\to\infty.$$

Next, we show that: $n^{a+1}\sup_{x\in C_n}|\sum_{i=1}^{n}Z_i(x)| \xrightarrow{a.s.} 0,\ n\to\infty$.

We cover C_n by μ_n spheres in the shape of $\{x:||x-x_{nj}||\leq c_n\mu_n^{-1}\}$ for $1\leq j\leq\mu_n^d$, $c_n\longrightarrow\infty$ and μ_n chosen such that $\mu_n\longrightarrow\infty$ to be defined later and we make the following decomposition.

$$\begin{aligned}\left|\sum_{i=1}^{n}Z_i(x)\right| \leq\ & \frac{1}{nh_n^d}\left|\sum_{i=1}^{n}\left[K_1\left(\frac{x-X_i}{h_n}\right)-K_1\left(\frac{x_{nj}-X_i}{h_n}\right)\right]\right|+\\ & \frac{1}{nh_n^d}\left|\sum_{i=1}^{n}E^{\mathcal{G}_{i-1}}\left[K_1\left(\frac{x-X_i}{h_n}\right)-K_1\left(\frac{x_{nj}-X_i}{h_n}\right)\right]\right|+\\ & \frac{1}{nh_n^d}\left|\sum_{i=1}^{n}\left[K_1\left(\frac{x_{nj}-X_i}{h_n}\right)-E^{\mathcal{G}_{i-1}}K_1\left(\frac{x_{nj}-X_i}{h_n}\right)\right]\right|.\end{aligned}$$

We have:

$$\frac{n^{a+1}}{nh_n^d}\left|\sum_{i=1}^{n}\left[K_1\left(\frac{x-X_i}{h_n}\right)-K_1\left(\frac{x_{nj}-X_i}{h_n}\right)\right]\right| \leq \frac{n^{a+1}L_K}{h_n^{d+\gamma_1}}||x-x_{nj}||^{\gamma_1} \leq$$

$$\frac{L_K n^{a+1}}{h_n^{d+\gamma_1}}c_n^{\gamma_1}\mu_n^{-\gamma_1} = \frac{1}{Logn}$$

where μ_n is chosen such that

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

Then:

$$n^{a+1}\sup_{x\in C_n}\left|\sum_{i=1}^{n}Z_i(x)\right| \leq$$

$$\sup_{1\leq j\leq \mu_n^d}\frac{n^{a+1}}{nh_n^d}\left|\sum_{i=1}^{n}\left[K_1\left(\frac{x_{nj}-X_i}{h_n}\right)-E^{\mathcal{G}_{i-1}}K_1\left(\frac{x_{nj}-X_i}{h_n}\right)\right]\right| + \frac{2}{Logn}.$$

For all $n \geq n_1(\epsilon)$ and for all $\epsilon > 0$

$$P\left(n^{a+1}\sup_{x\in C_n}\left|\sum_{i=1}^{n}Z_i(x)\right| > 2\epsilon\right) = P\left(\sup_{x\in C_n}\left|\sum_{i=1}^{n}Z_i(x)\right| > 2\epsilon_n\right) \leq$$

$$\sum_{j=1}^{\mu_n^d}P\left(\frac{1}{nh_n^d}\left|\sum_{i=1}^{n}\left[K_1\left(\frac{x_{nj}-X_i}{h_n}\right)-E^{\mathcal{G}_{i-1}}K_1\left(\frac{x_{nj}-X_i}{h_n}\right)\right]\right| > \epsilon_n\right).$$

Applying Azuma's Lemma μ_n^d times we obtain:

$$P\left(\sup_{x\in C_n}\left|\sum_{i=1}^{n}Z_i(x)\right| > 2\epsilon_n\right) \leq 2\mu_n^d \exp\left(-\frac{\epsilon_n^2 n h_n^{2d}}{8\overline{K}_1^2}\right) \leq$$

$$2h_n^{-d(d/\gamma_1+1)}L_K^{d/\gamma_1}c_n^d n^{d(1+a)/\gamma_1}(Logn)^{d/\gamma_1}\exp\left(-\frac{\epsilon_n^2 n h_n^{2d}}{8\overline{K}_1^2}\right)$$

where $\overline{K}_1$ is an upperbound of K_1.

The hypotheses of the Theorem and Borel-Cantelli lemma permit to conclude.

Now, we show that: $n^{a+1}\sup_{x\in C_n}|V_n(x)| \xrightarrow{a.s.} 0, \;\; n\to\infty.$ Write

$$V_n(x) = \frac{1}{nh_n^d}\int_{\mathrm{R}^d} K_1\left(\frac{u-x}{h_n}\right)\sum_{i=1}^{n} g_X^{\mathcal{G}_{i-1}}(u)du - g(x),$$

and set $z = (u-x)/h_n$ to obtain:

$$\sup_{x\in C_n}|V_n(x)| \leq \sup_{x\in C_n}\left|\int_{\mathrm{R}^d} K_1(z)n^{-1}\sum_{i=1}^{n}\left[g_X^{\mathcal{G}_{i-1}}(zh_n+x) - g_X^{\mathcal{G}_{i-1}}(x)\right]dz\right|$$

$$+\sup_{x\in C_n}\left|\int_{\mathrm{R}^d} K_1(z)\left[n^{-1}\sum_{i=1}^{n} g_X^{\mathcal{G}_{i-1}}(x) - g(x)\right]dz\right|.$$

By the assumption that the conditional densities satisfy the Lipschitz condition, we obtain

$$n^{a+1}\sup_{x\in C_n}|V_n(x)| \leq n^{a+1}h_n^k\int_{\mathrm{R}^d} z^k K_1(z)dz +$$

$$n^{a+1}\sup_{x\in C_n}\left|n^{-1}\sum_{i=1}^{n} g_X^{\mathcal{G}_{i-1}}(x) - g(x)\right|\int_{\mathrm{R}^d} K_1(z)dz.$$

The condition (T), the choice $T_{2,n} = o(n^{-a-1})$ and the assumption about the kernel K_1 permit us to conclude that:

$$n^{a+1}\sup_{x\in C_n}|V_n(x)| \xrightarrow{a.s.} 0, \;\; n\to\infty.$$

Lemma 3

Under the assumptions of the Theorem, we have:

$$n^a\sup_{x\in C_n}\sup_{y\in\mathrm{R}}|f_n(x,y) - f(x,y)| \xrightarrow{a.s.} 0, \;\; n\to\infty.$$

Proof:

$$f_n(x,y) - f(x,y) = \sum_{i=1}^{n} Z_i(x,y) + T_n(x,y),$$

where

$$T_n(x,y) = \frac{1}{nh_n^{d+1}} \sum_{i=1}^{n} E^{\mathcal{F}_{i-1}} \left\{ K_2\left(\frac{y-Y_i}{h_n}\right) K_1\left(\frac{x-X_i}{h_n}\right) \right\} - f(x,y),$$

and

$$Z_i(x,y) = \frac{1}{nh_n^{d+1}} K_2\left(\frac{y-Y_i}{h_n}\right) K_1\left(\frac{x-X_i}{h_n}\right) -$$

$$\frac{1}{nh_n^{d+1}} E^{\mathcal{F}_{i-1}} \left[K_2\left(\frac{y-Y_i}{h_n}\right) K_1\left(\frac{x-X_i}{h_n}\right) \right]$$

Z_i is a martingale difference with $|Z_i| \le \frac{2\widetilde{K}^2}{nh_n^{d+1}}$, where

$$\widetilde{K} = \max\left\{ \sup_{x\in\mathrm{R}^d} K_1(x), \sup_{y\in\mathrm{R}} K_2(y),\ 1 \right\}.$$

Then, apply Lemma 1 to obtain:

$$\forall \epsilon > 0, \qquad P\left\{ |\sum_{i=1}^{n} Z_i| > n^{-a}\epsilon \right\} = P\left\{ |\sum_{i=1}^{n} Z_i| > \epsilon_n \right\} \le$$

$$2\exp\left\{ -C_1 \epsilon_n^2 n h_n^{2(d+1)} \right\}, \tag{3}$$

where C_1 is a positive constant.
Condition (1) in the Theorem permits us to conclude:

$$\sum_n P\left\{ n^a |\sum_{i=1}^{n} Z_i| > \epsilon \right\} < \infty.$$

Next, we show that: $n^a \sup_{x\in C_n} \sup_{y\in\mathrm{R}} |\sum_{i=1}^n Z_i(x,y)| \xrightarrow{a.s.} 0,\ n\to\infty.$

We cover C_n by μ_n^d spheres: $\{x\ :\ ||x-x_{nj}|| \le c_n\mu_n^{-1}\}$, $\quad 1\le j\le \mu_n^d$, where $c_n \longrightarrow \infty$, and μ_n is chosen so that $\mu_n \to \infty$, to be defined precisely later.

Consider the following decomposition:

$$\sum_{i=1}^{n} Z_i(x,y) = \sum_{i=1}^{n} [\Upsilon_i(x,y) - \Upsilon_i(x_{nj},y)] -$$

$$\sum_{i=1}^{n} E^{\mathcal{F}_{i-1}} [\Upsilon_i(x,y) - \Upsilon_i(x_{nj},y)] \ + \ \sum_{i=1}^{n} [\Upsilon_i(x_{nj},y) - E^{\mathcal{F}_{i-1}} \Upsilon_i(x_{nj},y)],$$

where $\Upsilon_i(.,y) = \frac{1}{nh_n^{d+1}} K_2\left(\frac{y-Y_i}{h_n}\right) K_1\left(\frac{.-X_i}{h_n}\right)$.

By the fact that the kernel K_1 is Lipschitz, we obtain:

$$n^a \sup_{x\in C_n} \sup_{y\in \mathrm{R}} |\sum_{i=1}^{n}[\Upsilon_i(x,y)-\Upsilon_i(x_{nj},y)] \leq \frac{L_K\widetilde{K}n^a}{h_n^{d+1+\gamma_1}}||x-x_{nj}||^{\gamma_1} \leq \frac{L_K\widetilde{K}n^a}{h_n^{d+1+\gamma_1}}c_n^{\gamma_1}\mu_n^{-\gamma_1}$$

$$= \frac{1}{\mathrm{Log}n},$$

where μ_n is chosen so that: $\mu_n = \frac{L_K^{1/\gamma_1}\widetilde{K}^{1/\gamma_1}c_n n^{a/\gamma_1}(\log n)^{1/\gamma_1}}{h_n^{(d+1+\gamma_1)/\gamma_1}} \to \infty$. Thus,

$$n^a \sup_{x\in C_n} \sup_{y\in \mathrm{R}} |\sum_{i=1}^{n} Z_i(x,y)| \leq$$

$$n^a \sup_{1\leq j\leq \mu_n^d} \sup_{y\in \mathrm{R}} |\sum_{i=1}^{n}[\Upsilon_i(x_{nj},y) - E^{\mathcal{F}_{i-1}}\Upsilon_i(x_{nj},y)]| + \frac{2}{\mathrm{Log}n},$$

and then, for all $n \geq n_1(\epsilon)$ and all $\epsilon > 0$, if we put $\epsilon_n = n^{-a}\epsilon$ we have:

$$P\left\{\sup_{x\in C_n} \sup_{y\in \mathrm{R}} |\sum_{i=1}^{n} Z_i(x,y)| > 2\epsilon_n\right\} \leq$$

$$\sum_{j=1}^{\mu_n^d} P\left\{\sup_{y\in \mathrm{R}} |\sum_{i=1}^{n}[\Upsilon_i(x_{nj},y) - E^{\mathcal{F}_{i-1}}\Upsilon_i(x_{nj},y)]| > \epsilon_n\right\}. \qquad (4)$$

For fixed j, set:

$$\sum_{i=1}^{n}[\Upsilon_i(x_{nj},y) - E^{\mathcal{F}_{i-1}}\Upsilon_i(x_{nj},y)] = \Delta_n(x_{nj},y) \qquad \text{if} \qquad |y| \leq v_n$$

$$\sum_{i=1}^{n}[\Upsilon_i(x_{nj},y) - E^{\mathcal{F}_{i-1}}\Upsilon_i(x_{nj},y)] = \delta_n(x_{nj},y) \qquad \text{if} \qquad |y| > v_n$$

where v_n is defined by $v_n = h_n^{-\frac{1}{\mu}}$ with μ being a positive constant.

Then we have

$$\sup_{y\in \mathrm{R}}|\sum_{i=1}^{n}[\Upsilon_i(x_{nj},y)-E^{\mathcal{F}_{i-1}}\Upsilon_i(x_{nj},y)]|\leq \sup_{|y|\leq v_n}|\Delta_n(x_{nj},y)|+\sup_{|y|>v_n}|\delta_n(x_{nj},y)|.$$

Cover $[-v_n,v_n]$ by l_n spheres B_s with centers t_s and radii less than or equal to h_n^{η}, where $l_n\leq v_nh_n^{-\eta}$ and η is a fixed number. Then by arguments similar to those in the proof of Lemma 2, we obtain:

$$\sup_{|y|\leq v_n}|\widetilde{\Delta_n}(x_{nj},y)|\leq \lambda_0 h_n^{\gamma_1(\eta-1)-(d+1)} \quad a.s.,$$

where $\widetilde{\Delta_n}(x_{nj},y)=\Delta_n(x_{nj},y)-\Delta_n(x_{nj},t_s)$ and λ_0 is a positive constant.

Furthermore,

$$\omega_n=P\left\{\max_{s=1,\dots,l_n}|\Delta_n(x_{nj},t_s)|>\epsilon_n/2\right\}\leq\sum_{s=1}^{l_n}P\left\{|\Delta_n(x_{nj},t_s)|>\epsilon_n/2\right\}\leq$$

$$l_n\sup_{|y|\leq v_n}P\left\{|\Delta_n(x_{nj},y)|>\epsilon_n/2\right\}.$$

Then inequality (3) implies: $\omega_n\leq 2v_nh_n^{-\eta}\exp\{-C_1\epsilon_n^2nh_n^{2(d+1)}\}$.

Applying Lemma 1, μ_n^d times, we obtain:

$$P\left\{\sup_{x\in C_n}\sup_{|y|\leq v_n}|\sum_{i=1}^{n}Z_i(x,y)|>\epsilon_n\right\}\leq$$

$$\frac{n^{ad/\gamma_1}c_n^dL_K^{d/\gamma_1}\widetilde{K}^{d/\gamma_1}(\mathrm{Log}n)^{d/\gamma_1}}{h_n^{d(d+1+\gamma_1)/\gamma_1}}h_n^{-\eta-\frac{1}{\mu}}\exp\{-C_1\epsilon_n^2nh_n^{2(d+1)}\}.$$

The assumptions of the Theorem permit us to conclude that:

$$n^a\sup_{x\in C_n}\sup_{|y|\leq v_n}|\sum_{i=1}^{n}Z_i(x,y)|\xrightarrow{a.s.}0.$$

It remains to show that: $n^a\sup_{|y|>v_n}|\delta_n(x_{nj},y)|\xrightarrow{a.s.}0$. We have

$$\sup_{|y|>v_n}|\delta_n(x_{nj},y)|\leq\sup_{|y|>v_n}|\sum_{i=1}^{n}\Upsilon_i(x_{nj},y)|+\sup_{|y|>v_n}|\sum_{i=1}^{n}E^{\mathcal{F}_{i-1}}\Upsilon_i(x_{nj},y)|,$$

and by the compactness of the support of K_2,

$$K_2\left(\frac{y-Y}{h_n}\right)\leq\widetilde{K}\mathrm{I}_{[|Y|>v_n/2]}.$$

Therefore

$$\sup_{|y|>v_n} |\sum_{i=1}^{n} \Upsilon_i(x_{nj}, y)| \leq \frac{1}{nh_n^{d+1}} \widetilde{K}^2 \sum_{i=1}^{n} \mathrm{I}_{[|Y_i|>v_n/2]} \tag{5}$$

with

$$P(|Y| > v_n/2) \leq (2v_n^{-1})^{\xi}(E|Y|^{\xi}) \tag{6}$$

for a certain $\xi > 0$ such that $\xi > \mu\gamma_1(\eta - 1)$.

For all $\epsilon > 0$, we have

$$P\left\{\sup_{|y|>v_n} |\sum_{i=1}^{n} \Upsilon_i(x_{nj}, y)| > \epsilon_n\right\} \leq \epsilon_n^{-1} E\left[\sup_{|y|>v_n} |\sum_{i=1}^{n} \Upsilon_i(x_{nj}, y)|\right].$$

Then, using (5) and (6) we obtain:

$$P\left\{\sup_{|y|>v_n} |\sum_{i=1}^{n} \Upsilon_i(x_{nj}, y)| > \epsilon_n\right\} \leq \epsilon_n^{-1}\widetilde{K}^2 h_n^{-d-1}(2v_n^{-1})^{\xi}(E|Y|^{\xi}) = \epsilon_n^{-1}\widetilde{K}^2 h_n^{-d-1+\frac{\xi}{\mu}} 2^{\xi}(E|Y|^{\xi}).$$

Inequality (4) implies:

$$P\left\{\sup_{x\in C_n} \sup_{|y|>v_n} |\sum_{i=1}^{n} Z_i(x, y)| > 2\epsilon_n\right\} \leq A\mu_n^d h_n^{-d-1+\frac{\xi}{\mu}}(E|Y|^{\xi}),$$

where A is a positive constant.

The choice of ξ and the assumptions of the Theorem permit us to conclude that:

$$n^a \sup_{x\in C_n} \sup_{y\in \mathrm{R}} |\sum_{i=1}^{n} Z_i(x, y)| \xrightarrow{a.s.} 0$$

To complete the proof of Lemma 3, we need to show that:

$$n^a \sup_{x\in C_n} \sup_{y\in \mathrm{R}} |T_n(x, y)| \xrightarrow{a.s.} 0, \quad n \to \infty.$$

To this end:

$$T_n(x, y) = \frac{1}{nh_n^{d+1}} \sum_{i=1}^{n} E^{\mathcal{F}_{i-1}}\left\{K_2\left(\frac{y - Y_i}{h_n}\right) K_1\left(\frac{x - X_i}{h_n}\right)\right\} - f(x, y),$$

with

$$E^{\mathcal{F}_{i-1}}\left\{K_2\left(\frac{y - Y_i}{h_n}\right) K_1\left(\frac{x - X_i}{h_n}\right)\right\} =$$

$$\int\int_{\mathrm{R}^d\times\mathrm{R}} K_2\left(\frac{y-v}{h_n}\right) K_1\left(\frac{x-u}{h_n}\right) f_{X,Y}^{\mathcal{F}_{i-1}}(u,v)dudv.$$

Properties of the Bochner's integral permit to write

$$T_n(x,y) =$$

$$\frac{1}{h_n^{d+1}}\int\int_{\mathrm{R}^d\times\mathrm{R}} K_2\left(\frac{y-v}{h_n}\right) K_1\left(\frac{x-u}{h_n}\right) n^{-1}\sum_{i=1}^{n} f_{X,Y}^{\mathcal{F}_{i-1}}(u,v)dudv - f(x,y).$$

Then if we set $z_1 = (x-u)/h_n$, $z_2 = (y-v)/h_n$, we obtain

$$T_n(x,y) = \int\int_{\mathrm{R}^d\times\mathrm{R}} K_2(z_2)K_1(z_1)n^{-1}\sum_{i=1}^{n} f_{X,Y}^{\mathcal{F}_{i-1}}(x-z_1h_n, y-z_2h_n)dz_1dz_2 - f(x,y).$$

Condition (T) and the fact that the conditional densities $f_{X,Y}^{\mathcal{F}_{i-1}}$ are Lipschitz and similar arguments to those used before yield:

$$n^a \sup_{x\in C_n}\sup_{y\in\mathrm{R}} |T_n(x,y)| \xrightarrow{a.s.} 0, \quad n\to\infty$$

5. Proof of the Main Result

By the definitions of $\Theta_n(x)$ and $\Theta(x)$, we have

$$\begin{aligned}|f(\Theta_n(x)|x) - f(\Theta(x)|x)| &\leq |f_n(\Theta_n(x)|x) - f(\Theta_n(x)|x)| + |f_n(\Theta_n(x)|x) - f(\Theta(x)|x)|\\ &\leq \sup_{y\in\mathrm{R}} |f_n(y|x) - f(y|x)| + |\sup_{y\in\mathrm{R}} f_n(y|x) - \sup_{y\in\mathrm{R}} f(y|x)|\\ &\leq 2\sup_{y\in\mathrm{R}} |f_n(y|x) - f(y|x)|.\end{aligned}$$

Assumption A6 implies that for all $\epsilon_n > 0$ there exists $\beta_n > 0$ such that:

$$P\left(\sup_{x\in C_n} |\Theta_n(x) - \Theta(x)| \geq \epsilon_n\right) \leq P\left(\sup_{x\in C_n}\sup_{y\in\mathrm{R}} |f_n(y|x) - f(y|x)| \geq \beta_n\right),$$

which completes the proof of the Theorem.

The Open Problem

The rate of convergence remains up to now very hard to control because it could be arbitrarily fast, one can consider this study in the case when the process is ergodic on each compact set separately and find a function to conclude for the whole space.

Acknowledgements

The author is grateful to the referees for their comments and citicisms

References

[1] Arfi, M. *Sur la régression non paramétrique d'un processus stationnaire mélangeant ou ergodique.* Thèse de Doctorat de l'Université Paris VI, 1996

[2] Arfi, M. Nonparametric prediction from ergodic samples. *J. Nonparametric Statistics.* **9,** 1998, 23 - 37

[3] Azuma, K. Weighted sums of certain dependent random variables. *Tôhoku Math. J.*, **19**, 1967, 357-367.

[4] Beck, A. On the strong law of large numbers. *Ergodic theory*, F. B. Wright Ed Academic Press, 1963, 21-53

[5] Collomb, G. Nonparametric regression: an up to date bibliography. *Statistics*, **16**, 1985, 309-324

[6] Collomb, G., Hardle, G., Hassani, S. A note on prediction via estimation of the conditional mode function. *J. Statist. Plann. Inference.* **15,** 1987, 227 - 236

[7] Györfi, L.; Härdle, W.; Sarda, P.; Vieu, P. Nonparametric curve estimation from time series. *Lecture Notes in Statistics*, **60**, 1989, Springer-Verlag

[8] Hermann, E., Ziegler, K. Rates of consistency for a nonparametric estimation of the mode in absence of smothness assumptions. *Statistics&Probability Letters,* **68,** 2004, 359 - 368

[9] Parzen, E. On the estimation of a probability density function and mode. *Ann. Statist.,* **33,** 1962, 1065 - 1076