

Rates of strong uniform consistency for local least squares kernel regression estimators

David Blondin *

L.S.T.A., Université Paris VI, 175 rue du Chevaleret, 75013 Paris, France

Abstract

We establish exact rates of strong uniform consistency for the multivariate Nadaraya-Watson kernel estimator of the regression function and its derivatives. As a special case, we treat the local linear estimator of the regression and the local polynomial smoothers of derivatives of the regression in the more convenient univariate setting. Our methods of proofs are based upon modern empirical process theory in the spirit of the results of Einmahl and Mason [4] and Deheuvels and Mason [2] relative to uniform deviations of nonparametric kernel estimators.

Key words: nonparametric regression, derivative estimation, kernel estimation, local linear least squares kernel estimator, local polynomial fitting, strong uniform consistency, rate of convergence, uniform limit law of the logarithm.

1. Introduction

Let $(\mathbf{X}, \mathbf{Y}) = (\mathbf{X}_1, \mathbf{Y}_1), (\mathbf{X}_2, \mathbf{Y}_2), \dots$, be independent and identically distributed random vectors in $\mathbb{R}^p \times \mathbb{R}^d$. In this paper, we are concerned with the estimation of the conditional expectation of a functional ψ of \mathbf{Y} given \mathbf{X} . The following notation and assumptions will be in force. Let \mathbf{x} (respectively \mathbf{y}) denote a vector in \mathbb{R}^p (resp. \mathbb{R}^d). We assume that the distribution function of (\mathbf{X}, \mathbf{Y}) has a Lebesgue density $f_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) = g(\mathbf{x}, \mathbf{y})$ on $\mathbb{R}^p \times \mathbb{R}^d$. We denote by $f_{\mathbf{X}}(\mathbf{x}) = f(\mathbf{x})$ the marginal density of \mathbf{X} assumed to exist for each $\mathbf{x} \in \mathbb{R}^p$. Let $\psi : \mathbb{R}^d \rightarrow \mathbb{R}^q$ a Borel measurable function, bounded on each compact set of \mathbb{R}^d . We consider the regression function of $\psi(\mathbf{Y})$ given $\mathbf{X} = \mathbf{x}$, defined by

$$m_\psi(\mathbf{x}) = \mathbb{E}[\psi(\mathbf{Y})|\mathbf{X} = \mathbf{x}] = \frac{1}{f(\mathbf{x})} \int_{\mathbb{R}^d} \psi(\mathbf{y})g(\mathbf{x}, \mathbf{y})d\mathbf{y} := \frac{r_\psi(\mathbf{x})}{f(\mathbf{x})}, \quad (1.1)$$

whenever this regression function is meaningful (see, e.g., (F.1–3) below). Throughout this paper, we will investigate the strong consistency of kernel-type estimators of the regression $m_\psi(\mathbf{x})$ and its derivatives.

* Fax: +33-1-44-27-33-42

Email address: blondin@ccr.jussieu.fr (David Blondin).

Denote by $\mathbf{I} = \prod_{i=1}^p [a_i, b_i]$ and $\mathbf{J} = \prod_{i=1}^p [a'_i, b'_i] \supset \mathbf{I}$ two fixed closed hyper-rectangles of \mathbb{R}^p such that

$$-\infty < a'_i < a_i < b_i < b'_i < \infty, \quad \text{for } i = 1, \dots, p.$$

We will work in the multivariate framework, where $p \geq 1$, $q \geq 1$, $d \geq 1$ are arbitrary integers, and impose the following set of assumptions upon the distribution of (\mathbf{X}, \mathbf{Y}) .

(F.1) $g(\cdot, \cdot)$ is continuous on $\mathbf{J} \times \mathbb{R}^d$;

(F.2) $f(\cdot)$ is continuous and strictly positive on \mathbf{J} ;

(F.3) $\mathbf{Y} \mathbb{1}_{\{\mathbf{X} \in \mathbf{J}\}}$ is bounded on \mathbb{R}^d .

Remark 1.1 *The above assumption (F.3) is introduced for technical reasons related to our forthcoming proofs. At times, we will replace this assumption by*

(F.4) *for some $s > 2$, $\sup_{\mathbf{x} \in \mathbf{J}} \mathbb{E}[\psi(\mathbf{Y})^s | \mathbf{X} = \mathbf{x}] < \infty$.*

We will see in the sequel that our methods do not allow us to establish a compact law of the iterated logarithm when (F.4) does not hold. Details on this problem are given in Einmahl and Mason [4] and Deheuvels and Mason [2].

Here and elsewhere $k \geq 1$ denotes a fixed order of differentiation in the following sense. Let $\zeta = (\zeta_1, \dots, \zeta_q)$ be an arbitrary measurable function with $\zeta_j : \mathbb{R}^p \rightarrow \mathbb{R}$, $j = 1, \dots, q$. In this section, we consider estimation of functionals of ζ defined at $\mathbf{x} = (x_1, \dots, x_p) \in \mathbf{J}$. For each p -uple of non-negative integers $k_1 \geq 0, \dots, k_p \geq 0$, $\mathbf{k} = (k_1, \dots, k_p)$, we introduce the differential operator $\mathbf{D}^{(\mathbf{k})}$ of order $|\mathbf{k}| = k = k_1 + \dots + k_p$, defined by

$$\mathbf{D}^{(\mathbf{k})}\zeta(\mathbf{x}) = \zeta^{(\mathbf{k})}(\mathbf{x}) = \left(\mathbf{D}^{(\mathbf{k})}\zeta_1(\mathbf{x}), \dots, \mathbf{D}^{(\mathbf{k})}\zeta_q(\mathbf{x}) \right),$$

where

$$\mathbf{D}^{(\mathbf{k})}\zeta_j(\mathbf{x}) = \left(\frac{\partial}{\partial x_1} \right)^{k_1} \dots \left(\frac{\partial}{\partial x_p} \right)^{k_p} \zeta_j(\mathbf{x}), \quad j = 1, \dots, q.$$

At first, we treat the Nadaraya-Watson kernel estimator (see [15] and [20]) and its partial derivatives when the predictor variables are \mathbb{R}^p -valued. For this purpose we introduce a general kernel function $\mathbf{K} : \mathbb{R}^p \rightarrow \mathbb{R}$, fulfilling the conditions

(K.1) (i) \mathbf{K} is bounded;

(ii) $\mathbf{K}(x_1, \dots, x_p)$ is a right-continuous function of x_1, \dots, x_p ;

(iii) $\mathbf{K}(\mathbf{u}) = \Phi(P(\mathbf{u}))$, where P is a polynomial, and Φ is a real valued function of bounded variation;

(K.2) \mathbf{K} is compactly supported;

(K.3) $\int_{\mathbb{R}^p} \mathbf{K}(\mathbf{u}) d\mathbf{u} = 1$;

(K.4) \mathbf{K} is k -times differentiable, with partial derivatives satisfying (K.1).

Remark 1.2 1) *When $k \geq 2$, the hypotheses (K.2) and (K.4) imply \mathbf{K} to be of bounded variation on \mathbb{R}^p (see, e.g., Hobson [12]).*

2) *Consider the classes of functions*

$$\mathcal{K} := \{ \mathbf{K}((\mathbf{x} - \cdot)/h) : \mathbf{x} \in \mathbb{R}^p, h \in \mathbb{R}^+ \} \quad \text{and} \quad \mathcal{K}_{\mathbf{k}} := \{ \mathbf{K}^{(\mathbf{k})}((\mathbf{x} - \cdot)/h) : \mathbf{x} \in \mathbb{R}^p, h \in \mathbb{R}^+ \}.$$

By (K.1)(i), there exist a $\kappa < \infty$ such that $\|\mathbf{K}\|_{\infty} = \kappa$, so that the function \mathbf{K} is bounded in absolute value by κ . For $\epsilon > 0$, set $N(\epsilon, \mathcal{K}) = \sup_Q N(\kappa\epsilon, \mathcal{K}, d_Q)$, where the supremum is taken over all probability

measures Q on $(\mathbb{R}^p, \mathcal{B})$, d_Q is the $L_2(Q)$ -metric and $N(\kappa\epsilon, \mathcal{K}, d_Q)$ denotes the minimal number of balls $\{f : d_Q(f, g) < \epsilon\}$ of d_Q -radius needed to cover \mathcal{K} . The above class of functions \mathcal{K} is known to have a polynomial covering number under the entropy condition that is, for some $C, \nu > 0$,

$$N(\epsilon, \mathcal{K}) \leq C\epsilon^{-\nu}, \quad 0 < \epsilon < 1. \quad (1.2)$$

3) On the other hand, **(K.1)** (resp. **(K.4)**) ensures that the class of functions \mathcal{K} (resp. $\mathcal{K}_{\mathbf{k}}$) has a measurable envelope function and a uniform polynomial covering number (see, e.g., Lemma 22, p. 797, in Nolan and Pollard [16] and Chapter 4 in Dudley [3]). The sharpness of the polynomial covering number assumption (1.2) is discussed in Mason [13] and [14] (Remarks 1 and 2, p. 1396).

4) The continuity condition **(K.1)(ii)** entails that \mathcal{K} is a pointwise measurable class, in the sense that it admits a dense countable subclass with respect to the pointwise convergence topology. This measurability assumption is instrumental with respect to the symmetrization procedure used later on in our proofs. The measurability condition is easy to check in all the examples of interest (see, e.g., [4]).

Let h_n , $n = 1, 2, \dots$, denotes a non-random (bandwidth) sequence of positive constants satisfying assumptions:

$$(H.1) \quad h_n \searrow 0 \text{ and } nh_n \nearrow \infty, \text{ as } n \rightarrow \infty;$$

$$(H.2) \quad nh_n^{2k+p} / \log(h_n^{-p}) \rightarrow \infty, \text{ as } n \rightarrow \infty;$$

$$(H.3) \quad \log(h_n^{-p}) / \log \log n \rightarrow \infty, \text{ as } n \rightarrow \infty.$$

Remark 1.3 1) The conditions $h_n \rightarrow 0$ and $nh_n^{2k+p} / \log n \rightarrow \infty$, as $n \rightarrow \infty$, which are slightly weaker than **(H.1–2)**, turn out to be, in general, necessary and sufficient to ensure the strong consistency of the multivariate Nadaraya-Watson estimator of the partial derivatives of order k of the regression function (consult Collomb [1]). These two conditions are also enough strong to obtain weak limit laws (see, e.g., Deheuvels and Mason [2]).

2) The assumptions **(H.1–3)** are tailored with respect to almost sure convergence of our estimators.

3) The assumptions **(H.2–3)** are sharp, in the sense that if either one is not fulfilled, then the conclusions of our forthcoming Theorems fail to hold in general. For the data-driven bandwidth case, refer to [2] and, more recently, to Einmahl and Mason [5].

We introduce the kernel estimators of $f(\mathbf{x})$, $r_\psi(\mathbf{x})$ and $m_\psi(\mathbf{x})$, defined by

$$\hat{f}_n(\mathbf{x}) = \frac{1}{nh_n^p} \sum_{i=1}^n \mathbf{K}\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right), \quad \hat{r}_{\psi;n}(\mathbf{x}) = \frac{1}{nh_n^p} \sum_{i=1}^n \psi(\mathbf{Y}_i) \mathbf{K}\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right),$$

$$\hat{m}_{\psi;n}(\mathbf{x}) = \frac{\hat{r}_{\psi;n}(\mathbf{x})}{\hat{f}_n(\mathbf{x})} \quad \text{when } \hat{f}_n(\mathbf{x}) \neq 0.$$

Our estimators of the functional $\mathbf{D}^{(\mathbf{k})}$ of $f(\mathbf{x})$ and $r_\psi(\mathbf{x})$ are then defined as follows

$$\hat{f}_n^{(\mathbf{k})}(\mathbf{x}) = \mathbf{D}^{(\mathbf{k})}(\hat{f}_n(\mathbf{x})) = \frac{1}{nh_n^{k+p}} \sum_{i=1}^n \mathbf{K}^{(\mathbf{k})}\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right),$$

$$\hat{r}_{\psi;n}^{(\mathbf{k})}(\mathbf{x}) = \mathbf{D}^{(\mathbf{k})}(\hat{r}_{\psi;n}(\mathbf{x})) = \frac{1}{nh_n^{k+p}} \sum_{i=1}^n \psi(\mathbf{Y}_i) \mathbf{K}^{(\mathbf{k})}\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right).$$

Unless otherwise specified, we will limit most of our exposition to the case where $\mathbf{k} = (k_1, \dots, k_p)$ is such that $k_j = 1$ and $k_l = 0$ for $l \neq j$ and denote by $\mathbf{k}_j = (0, \dots, 1, \dots, 0)$ the corresponding p -uple. It will become obvious later on that our methods allow to treat likewise the case of an arbitrary \mathbf{k} .

$$\hat{m}_{\psi;n}^{(\mathbf{k}_j)}(\mathbf{x}) = \mathbf{D}^{(\mathbf{k}_j)}(\hat{m}_{\psi;n}(\mathbf{x})) = \frac{\hat{r}_{\psi;n}^{(\mathbf{k}_j)}(\mathbf{x})}{\hat{f}_n(\mathbf{x})} - \frac{\hat{r}_{\psi;n}(\mathbf{x})\hat{f}_n^{(\mathbf{k}_j)}(\mathbf{x})}{\hat{f}_n^2(\mathbf{x})}.$$

We note that, when $k_j \geq 2$ for some j , $\hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x}) = \mathbf{D}^{(\mathbf{k})}(\hat{m}_{\psi;n}(\mathbf{x}))$ may be obtained likewise through the usual Leibniz expansion of derivatives of products. One of the motivation to this study comes from the fact that the estimation of the proper partial derivatives of m_ψ is needed to implement plug-in bandwidth selection strategies. For details, consult Section 2.4 in Deheuvels and Mason (2004).

Set

$$f_n^{(\mathbf{k})}(\mathbf{x}) = \mathbb{E}[\hat{f}_n^{(\mathbf{k})}(\mathbf{x})] \quad \text{and} \quad r_{\psi;n}^{(\mathbf{k})}(\mathbf{x}) = \mathbb{E}[\hat{r}_{\psi;n}^{(\mathbf{k})}(\mathbf{x})]. \quad (1.3)$$

For the study of the rate of convergence of our estimator of the regression function and its differentials, it will be convenient to center $\hat{m}_n(\mathbf{x})$ and $\hat{m}_{\psi;n}^{(\mathbf{k}_j)}(\mathbf{x})$, respectively, by

$$\tilde{\mathbb{E}}[\hat{m}_{\psi;n}(\mathbf{x})] := \frac{r_{\psi;n}(\mathbf{x})}{f_n(\mathbf{x})}, \quad \tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{(\mathbf{k}_j)}(\mathbf{x})] := \frac{r_{\psi;n}^{(\mathbf{k}_j)}(\mathbf{x})}{f_n(\mathbf{x})} - \frac{r_{\psi;n}(\mathbf{x})f_n^{(\mathbf{k}_j)}(\mathbf{x})}{f_n^2(\mathbf{x})}. \quad (1.4)$$

A similar setup applies for operators $\mathbf{D}^{(\mathbf{k})}$ with $|\mathbf{k}| = k \geq 2$.

The main purpose of the present paper is to give a description of the exact rate of strong convergence of the uniform deviation of $\hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x}) - \tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x})]$ over $\mathbf{x} \in \mathbf{I}$. The convergence rate we shall obtain takes the form of uniform law of the logarithm and are stated in Section 2. The methods of proof that we will use to establish these results can be applied to treat a number of other kernel-type estimators, including local polynomial estimators (see, e.g., Fan and Gijbels [7] and Ruppert and Wand [19]). Section 3 contains the one-dimensional extension to the local linear least squares kernel estimator of m_ψ , higher-order polynomial fits and derivative estimation. Section 4 is devoted to the proofs. The main idea underlying our arguments is that by using the techniques from abstract empirical process theory and probability on Banach spaces, one can prove almost sure limit laws for a large class of nonparametric estimators of regression functions.

2. Uniform laws of the logarithm for estimators of the regression derivatives

First, we treat the special case $q = 1$ (that is, when $\psi(\mathbf{Y}) \in \mathbb{R}$).

Theorem 2.1 *Under (F.1-3), (H.1-3), (K.1-4) we have, as $n \rightarrow \infty$,*

$$\left| \left\{ \frac{nh_n^{2k+p}}{2 \log(h_n^{-p})} \right\}^{1/2} \sup_{\mathbf{x} \in \mathbf{I}} \pm \left\{ \hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x}) - \tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x})] \right\} - \sigma_\psi(\mathbf{I}) \right| = o(1), \quad \text{almost surely}, \quad (2.1)$$

where

$$\sigma_\psi(\mathbf{I}) = \sup_{\mathbf{x} \in \mathbf{I}} \left\{ \frac{\text{Var}[\psi(\mathbf{Y})|\mathbf{X} = \mathbf{x}]}{f(\mathbf{x})} \int_{\mathbb{R}^p} [\mathbf{K}^{(\mathbf{k})}(\mathbf{u})]^2 d\mathbf{u} \right\}^{1/2}. \quad (2.2)$$

For a related ‘in probability’ theorem, refer to Theorem 1.1 in [2].

The strictly multivariate case: $\psi(\mathbf{Y}) \in \mathbb{R}^q$, $q > 1$

The extension to several dimensions for $\psi(\mathbf{Y})$ is straightforward. We just need a suitable preliminary normalization of the stochastic deviation $\hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x}) - \tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x})]$. This enables us to characterize precisely the corresponding limit set, by using a standard argument developed by Finkelstein [8]. Let $\Sigma_\psi(\mathbf{x})$ denotes the conditional variance-covariance matrix of $\psi(\mathbf{Y})$ given that $\mathbf{X} = \mathbf{x}$, which is assumed definite

positive without loss of generality. We will make use of the following asymptotic variance-covariance matrix

$$\mathbf{V}_{\mathbf{x}} = \mathbf{V}_{\psi}(\mathbf{x}) := \left\{ \frac{1}{f(\mathbf{x})} \int_{\mathbb{R}^p} [\mathbf{K}^{(\mathbf{k})}(\mathbf{t})]^2 d\mathbf{t} \right\} \times \Sigma_{\psi}(\mathbf{x}),$$

which is properly defined for $\mathbf{x} \in \mathbf{J}$ and definite positive under (F.1-3). Below, suprema of q -vectors are meant with respect to the maximum norm $|\cdot|_+$ on \mathbb{R}^q , where $|\mathbf{v}|_+ = \max_{i \leq q} |v_i|$.

Theorem 2.2 *If (F.1-3), (H.2-4) and (K.1-4) hold, then, the sequence of normalized q -vectors*

$$\left\{ \frac{nh_n^{2k+p}}{2 \log(h_n^{-p})} \right\}^{1/2} \sup_{\mathbf{x} \in \mathbf{I}} \pm \left\{ \mathbf{V}_{\mathbf{x}}^{-1/2} \{ \hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x}) - \tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x})] \} \right\}$$

is almost surely relatively compact in \mathbb{R}^q with limiting set

$$\mathcal{S}_q = \{ \mathbf{v} \in \mathbb{R}^q : \mathbf{v}^T \mathbf{v} = 1 \}, \quad \text{the } q\text{-dimensional unit sphere.}$$

Notice that Theorem 2.2 is a simple consequence of Theorem 2.1, when combined with the proof of Lemma 2 of Finkelstein (1971) [8]. To see how Theorem 2.2 follows from Theorem 2.1, observe that the normalized deviation (properly rescaled)

$$\left\{ \mathbf{V}_{\mathbf{x}}^{-1/2} \{ \hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x}) - \tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{(\mathbf{k})}(\mathbf{x})] \} \right\}$$

is asymptotically normally distributed with an identity covariance matrix.

Remark 2.1 *With obvious changes of notation, Theorems 2.1 and 2.2 and their proofs remain valid when the regression function m_{ψ} is replaced by*

$$m_{\psi}^{\xi}(\mathbf{x}) = \mathbb{E}[\psi(\mathbf{Y}) | \xi(\mathbf{X}) = \mathbf{x}],$$

where $\xi : \mathbb{R}^p \rightarrow \mathbb{R}^r$ denotes an auxiliary continuous function and $\mathbf{x} \in \mathbb{R}^r$.

3. Uniform law of the logarithm for local polynomial kernel estimators

In this paper, we are mainly focused on the stochastic part of the usual deviation. The deterministic part or the so-called bias depends essentially on smoothness properties of the regression function m_{ψ} and can be easily estimated via techniques from analysis. In the literature, it is a well-known fact that the asymptotic bias of the Nadaraya-Watson estimator has a bad form. Namely, if we use a standard kernel of order two, the bias of the Nadaraya-Watson estimator depends on the intrinsic part m_{ψ}'' interplaying with the artifact $m_{\psi}' f' / f$. Thus, even in the situation of a linear regression, the bias of the NW estimator is large. The same phenomenon occurs for the Nadaraya-Watson estimators of the regression derivatives. To overcome this problem, it exists many alternative estimators. For example, if we assume some regularity conditions on the regression function, one can use the local polynomial regression techniques of Fan and Gijbels [7]. Our next task will be to extend the preceding results to the local polynomial least squares smoothers. For ease of presentation, we restrict ourselves first to the univariate case and to the local linear least squares estimator. At the end of this section, we present also some extensions to higher-order polynomial fits and derivative estimation.

Let $(X, Y), (X_1, Y_1), (X_2, Y_2), \dots$, be independent and identically distributed random couples in \mathbb{R}^2 . Now K denotes a real valued kernel function defined on \mathbb{R} , ψ is a Borel function bounded on each compact subinterval of \mathbb{R} and $I = [a, b], J = [a', b'] \supset I$ are two fixed compact intervals in \mathbb{R} . In the univariate setting, assumptions (F.1-3), (K.1-3) and (H.1-3) reduce to the followings:

(F.1) $g(\cdot, \cdot)$ is continuous on $J \times \mathbb{R}$;

(F.2) $f(\cdot)$ is continuous and strictly positive on J ;

(F.3) $Y \mathbb{1}_{\{X \in J\}}$ is bounded on \mathbb{R} ;

(K.1) K is a right-continuous function with bounded variation on \mathbb{R} ;

(K.2) K is compactly supported;

(K.3) $\int_{\mathbb{R}} K(u) du = 1$;

(H.1) $h_n \searrow 0$ and $nh_n \nearrow \infty$, as $n \rightarrow \infty$;

(H.2) $nh_n^{2k+1} / \log(h_n^{-1}) \rightarrow \infty$, as $n \rightarrow \infty$;

(H.3) $\log(h_n^{-1}) / \log_2 n \rightarrow \infty$, as $n \rightarrow \infty$.

For a kernel K fulfilling (K.1–2), we set for any integer $r \geq 0$,

$$\mu_r = \mu_r(K) = \int_{\mathbb{R}} u^r K(u) du.$$

It is common to work with a kernel function $K(\cdot)$ such that

(K.4) $\mu_1(K) = 0$ and $\mu_2(K) \neq 0$.

Throughout this section, we assume also that $m_\psi(\cdot)$ is two times continuously differentiable over the compact interval J (see, e.g. (F.4–5) below). Our first aim will be to establish the strong uniform consistency of the local linear estimator of the regression, defined by

$$\hat{m}_{\psi;n}^{LL}(x) := \frac{\hat{r}_{\psi;n}(x) \hat{f}_{n,2}(x) - \hat{r}_{\psi;n,1}(x) \hat{f}_{n,1}(x)}{\hat{f}_n(x) \hat{f}_{n,2}(x) - \hat{f}_{n,1}(x) \hat{f}_{n,1}(x)}, \quad (3.1)$$

where

$$\begin{aligned} \hat{f}_{n,j}(x) &:= \frac{1}{nh_n} \sum_{i=1}^n \left\{ \frac{x - X_i}{h_n} \right\}^j K\left(\frac{x - X_i}{h_n}\right), \quad j = 1, 2, \\ \hat{r}_{\psi;n,1}(x) &:= \frac{1}{nh_n} \sum_{i=1}^n \psi(Y_i) \left\{ \frac{x - X_i}{h_n} \right\} K\left(\frac{x - X_i}{h_n}\right). \end{aligned}$$

Notice that, in the multivariate setting, such simple explicit formula for the local linear estimator is not available. This estimator is better than the Nadaraya-Watson estimator when the design is random and has the favorable property to reproduce polynomial of degree one. In particular, Fan [6] shows that the local linear estimator has an important asymptotic minimax property. Precisely, the local linear estimator has a high minimax efficiency among all possible estimators, including nonlinear smoothers such as median regression. For the centering terms, we proceed as in (1.3) and (1.4) and set

$$\tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{LL}(x)] := \frac{r_{\psi;n}(x) f_{n,2}(x) - r_{\psi;n,1}(x) f_{n,1}(x)}{f_n(x) f_{n,2}(x) - f_{n,1}(x) f_{n,1}(x)},$$

where

$$\mathbb{E}[\hat{f}_{n,j}(x)] = f_{n,j}(x), \quad \mathbb{E}[\hat{r}_{\psi;n,j}(x)] = r_{\psi;n,j}(x), \quad j \in \mathbb{N}.$$

We obtain the following uniform law of the logarithm concerning the local linear smoother.

Theorem 3.1 Under (F.1-3), (H.1-3), (K.1-4) we have, as $n \rightarrow \infty$,

$$\left| \left\{ \frac{nh_n}{2 \log(h_n^{-1})} \right\}^{1/2} \sup_{x \in I} \pm \left\{ \hat{m}_{\psi;n}^{LL}(x) - \tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{LL}(x)] \right\} - \sigma_\psi(I) \right| = o(1), \quad \text{almost surely,}$$

where

$$\sigma_\psi(I) = \sup_{x \in I} \left\{ \frac{\text{Var}[\psi(Y)|X=x]}{f(x)} \int_{\mathbb{R}} [K(u)]^2 du \right\}^{1/2}.$$

We now present an interesting application for the function $\psi(\cdot)$. For some $t \in \mathbb{R}$, by setting $\psi(y) = \mathbb{I}_{\{y \leq t\}}$ in (3.1), we obtain a new estimator of the conditional distribution function

$$F(t|x) := \mathbb{P}\{Y \leq t|X = x\},$$

defined by

$$\hat{F}_n^{LL}(t|x) := \frac{\hat{r}_n(x; t) \hat{f}_{n,2}(x) - \hat{r}_{n,1}(x; t) \hat{f}_{n,1}(x)}{\hat{f}_n(x) \hat{f}_{n,2}(x) - \hat{f}_{n,1}(x) \hat{f}_{n,1}(x)}.$$

where

$$\hat{r}_n(x; t) := \frac{1}{nh_n} \sum_{i=1}^n \mathbb{I}_{\{Y_i \leq t\}} K\left(\frac{x - X_i}{h_n}\right) \quad \text{and} \quad \hat{r}_{n,1}(x; t) := \frac{1}{nh_n} \sum_{i=1}^n \mathbb{I}_{\{Y_i \leq t\}} \left\{ \frac{x - X_i}{h_n} \right\} K\left(\frac{x - X_i}{h_n}\right).$$

In the next corollary we get the exact rate of strong uniform consistency of the *local linear conditional empirical distribution function*. Let

$$F_n^{LL}(t|x) := \frac{r_n(x; t) f_{n,2}(x) - r_{n,1}(x; t) f_{n,1}(x)}{f_n(x) f_{n,2}(x) - f_{n,1}(x) f_{n,1}(x)}.$$

Corollary 3.1 Under (F.1-3), (H.1-3) and (K.1-4), we have, as $n \rightarrow \infty$,

$$\left| \left\{ \frac{nh_n}{2 \log(h_n^{-1})} \right\}^{1/2} \sup_{x \in I} \pm \left\{ \hat{F}_n^{LL}(t|x) - F_n^{LL}(t|x) \right\} - \sigma_{F,t}(I) \right| \stackrel{a.s.}{=} o(1),$$

where

$$\sigma_{F,t}(I) = \sup_{x \in I} \left\{ \frac{F(t|x) \{1 - F(t|x)\}}{f(x)} \int_{\mathbb{R}} K^2(u) du \right\}^{1/2}.$$

Moreover, we have, with probability one,

$$\left| \left\{ \frac{nh_n}{2 \log(h_n)} \right\}^{1/2} \sup_{t \in \mathbb{R}} \sup_{x \in I} \pm \left\{ \hat{F}_n^{LL}(t|x) - F_n^{LL}(t|x) \right\} - \sigma_F(I) \right| = o(1), \quad (3.2)$$

where

$$\sigma_F(I) = \sup_{t \in \mathbb{R}} \sigma_{F,t}(I) = \frac{\|K\|_2}{2 \inf_{x \in I} f(x)}.$$

Now, we look after the bias term of this new estimator $\hat{F}_n^{LL}(t|x)$. One needs additional smoothness assumptions on $f(\cdot)$ and $g(\cdot, \cdot)$, namely

(F.4) f has continuous derivatives of order 1, 2, 3 on J ;

(F.5) g is three times continuously differentiable on $J \times \mathbb{R}$.

Proposition 3.1 Assume (K.1-4) and (F.1-5). When $h_n = h \rightarrow 0$, we have, for all $x \in I$,

$$F_n^{LL}(t|x) - F(t|x) = \frac{h^2}{2} F''(t|x) \mu_2(K) (1 + o(1)).$$

Note that Proposition 3.1 and Corollary 3.1 allow us to derive explicit convergence rates for the standard deviation $\hat{F}_n^{LL}(t|x) - F(t|x)$. We can also derive the optimal theoretical bandwidth, namely

$$h_n^{opt}(K) = \left\{ \frac{\log n}{n} \right\}^{1/5} \left\{ \frac{\sup_{x \in I} \left\{ \frac{\sigma^2(x)}{f_X(x)} \right\} \mu_0(K^2)}{\sup_{x \in I} \{m''(x)\}^2 [\mu_2(K)]^2} \right\}^{1/5}.$$

We investigate now general polynomial fits and derivative estimation. Let p be the degree of the local polynomial fit. Throughout the end of this section, we will therefore assume that $m_\psi(\cdot)$ is p times differentiable over the compact interval J and $(p+1)$ times differentiable over the compact interval I .

Set

$$\mathbf{X}_x = \begin{pmatrix} 1 & (X_1 - x) & \dots & (X_1 - x)^p \\ \vdots & \vdots & & \vdots \\ 1 & (X_n - x) & \dots & (X_n - x)^p \end{pmatrix}, \quad \mathbf{W}_x = \text{diag} \left\{ K \left(\frac{X_i - x}{h_n} \right) \right\} \quad \text{and} \quad \Psi(\mathbf{y}) = \begin{pmatrix} \psi(Y_1) \\ \vdots \\ \psi(Y_n) \end{pmatrix}.$$

According to the definition of Ruppert and Wand [17], for $k \leq p$, the local polynomial kernel estimator of the k -th derivative of $m_\psi(x)$ is defined by,

$$\tilde{m}_\psi^{(k)}(x) = \tilde{m}_{\psi;n}^{(k)}(x; p) := k! e_{k+1}^T (\mathbf{X}_x^T \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_x \Psi(\mathbf{y}),$$

where e_{k+1} is the $(p+1) \times 1$ vector having 1 in the $(k+1)$ -th entry and zeros elsewhere. It is convenient to set, for each $j \in \mathbb{N}$,

$$\tilde{f}_j(x) = \frac{1}{nh_n} \sum_{i=1}^n \left(\frac{X_i - x}{h_n} \right)^j K \left(\frac{X_i - x}{h_n} \right) \quad \text{and} \quad \tilde{r}_j(x) = \frac{1}{nh_n} \sum_{i=1}^n \psi(Y_i) \left(\frac{X_i - x}{h_n} \right)^j K \left(\frac{X_i - x}{h_n} \right).$$

Observe that

$$\tilde{m}_\psi^{(k)}(x) = k! e_{k+1}^T \times \begin{pmatrix} \tilde{f}_0(x) & \tilde{f}_1(x) & \dots & \tilde{f}_p(x) \\ \vdots & \vdots & & \vdots \\ \tilde{f}_p(x) & \tilde{f}_{p+1}(x) & \dots & \tilde{f}_{2p}(x) \end{pmatrix}^{-1} \times \begin{pmatrix} \tilde{r}_0(x) \\ \tilde{r}_1(x) \\ \vdots \\ \tilde{r}_p(x) \end{pmatrix}.$$

For example, if $(k, p) = (0, 1)$, we get clearly the local linear estimator defined in (3.1), namely

$$\tilde{m}_{\psi;n}^{(0)}(x; 1) = (1, 0) \times \begin{pmatrix} \tilde{f}_0(x) & \tilde{f}_1(x) \\ \tilde{f}_1(x) & \tilde{f}_2(x) \end{pmatrix}^{-1} \times \begin{pmatrix} \tilde{r}_0(x) \\ \tilde{r}_1(x) \end{pmatrix} = \frac{\tilde{r}_0(x) \tilde{f}_2(x) - \tilde{r}_1(x) \tilde{f}_1(x)}{\tilde{f}_0(x) \tilde{f}_2(x) - \tilde{f}_1(x) \tilde{f}_1(x)}.$$

Explicit formulae of local polynomial estimators became large as p grows due to matrix complexity. When $(k, p) = (1, 2)$, we get that

$$\tilde{m}_\psi^{(1)}(x; 2) = \frac{\tilde{r}(x)(\tilde{f}_2(x)\tilde{f}_3(x) - \tilde{f}_1(x)\tilde{f}_4(x)) + \tilde{r}_1(x)(\tilde{f}(x)\tilde{f}_4(x) - \tilde{f}_2^2(x)) + \tilde{r}_2(x)(\tilde{f}_1(x)\tilde{f}_2(x) - \tilde{f}(x)\tilde{f}_3(x))}{\tilde{f}(x)\tilde{f}_2(x)\tilde{f}_4(x) + 2\tilde{f}_1(x)\tilde{f}_2(x)\tilde{f}_3(x) - \tilde{f}_2^3(x) - \tilde{f}(x)\tilde{f}_3^2(x) - \tilde{f}_1^2(x)\tilde{f}_4(x)}.$$

Let N_p be the $(p+1) \times (p+1)$ matrix having (i, j) th entry equal to $\mu_{i+j-2}(K)$, $1 \leq i, j \leq p+1$ and let $M_{k,p}(u)$ be the same as N_p , but with the $(k+1)$ -th column replaced by $(1, u, \dots, u^p)^T$. This notation allows

us to introduce the kernel corresponding to the estimation of the k th derivative with local polynomial fitting of degree p (see, e.g., [7] or [19] pp. 135-137):

$$K_{(k,p)}(u) := k! \times \frac{|M_{k,p}(u)|}{|N_p|} \times K(u).$$

For $(k,p) = (1,2)$ and $(k,p) = (2,3)$, we obtain

$$\frac{|M_{1,2}(u)|}{|N_3|} = \frac{u}{\mu_2(K)} \quad \text{and} \quad \frac{|M_{2,3}(u)|}{|N_3|} = \frac{u^2 - \mu_2(K)}{\mu_4(K) - \mu_2^2(K)}.$$

Thus, under (K.3-4),

$$K_{(1,2)}(u) := \frac{uK(u)}{\mu_2(K)} \quad \text{and} \quad K_{(2,3)}(u) := 2 \times \frac{\{u^2 - \mu_2(K)\}K(u)}{\mu_4(K) - \mu_2^2(K)}$$

are kernels of order (1,3) and (2,4) respectively. More generally, it is easily established that $K_{(k,p)}$ satisfies

$$\int_{\mathbf{R}} u^j K_{(k,p)}(u) du = \begin{cases} 0, & 0 \leq j \leq p, \quad j \neq q \\ k!, & j = q \\ \beta_{k,p} \neq 0, & j = p+1. \end{cases}$$

Therefore $(-1)^k K_{(k,p)}$ is an order $(k, p+1)$ kernel as defined by Gasser et al. [9]. Such kernels are tailored for estimating derivatives of functions such as regression or density functions. Following Ruppert and Wand [17] or Wand and Jones [19], some routine analysis gives us the following important fact

$$\text{Var}[\tilde{m}_{\psi;n}^{(k)}(x)] = \frac{1}{nh_n^{2k+1}} \times \left\{ \frac{\text{Var}[\psi(Y)|X=x]}{f(x)} \int_{\mathbf{R}} [K_{(k,p)}(u)]^2 du \right\} (1 + o_{\mathbb{P}}(1)).$$

By repeating the same arguments as in the proof of Theorem 3.1, we can formulate a uniform limit law for each couple (k,p) .

Theorem 3.2 *Under (F.1-3), (H.1-3), (K.1-3) we have, as $n \rightarrow \infty$,*

$$\left| \left\{ \frac{nh_n^{2k+1}}{2 \log(h_n^{-1})} \right\}^{1/2} \sup_{x \in I} \pm \left\{ \tilde{m}_{\psi;n}^{(k)}(x) - \tilde{\mathbb{E}}[\tilde{m}_{\psi;n}^{(k)}(x)] \right\} - \sigma_{(k,p)}(I) \right| = o(1), \quad \text{almost surely,}$$

where

$$\sigma_{(k,p)}(I) = \sup_{x \in I} \left\{ \frac{\text{Var}[\psi(Y)|X=x]}{f(x)} \int_{\mathbf{R}} [K_{(k,p)}(u)]^2 du \right\}^{1/2}.$$

4. Proofs

The proofs of our theorems follow the same line and can be inferred with little effort from the proof of the Theorem 1 in Einmahl and Mason [4]. For ease of conciseness, we decide only to prove Theorem 3.1. In order to study the asymptotic behavior of the local linear estimator we introduce a general local empirical process. For any $j \in \mathbb{N}$ and continuous real valued functions $c(\cdot)$ and $d(\cdot)$ on the compact interval J , set for $x \in J$,

$$W_{n,j}(x, \psi) = \sum_{i=1}^n \left(c(x)\psi(Y_i) + d(x) \right) K_j \left(\frac{x - X_i}{h_n} \right) - n \mathbb{E} \left\{ \left(c(x)\psi(Y) + d(x) \right) K_j \left(\frac{x - X}{h_n} \right) \right\}, \quad (4.1)$$

where $K_j(u) = u^j K(u)$, $u \in \mathbb{R}$. Notice that

$$W_{n,j}(x, \psi) = \sqrt{n} \alpha_n(\vartheta_x),$$

where α_n denotes the bivariate empirical process based upon $(X_1, Y_1), \dots, (X_n, Y_n)$ and indexed by the class of functions

$$\vartheta_x(u, v) := \left(c(x)\psi(v) + d(x) \right) K_j \left(\frac{x-u}{h_n} \right), \quad x \in I.$$

Theorem 4.1 *Under (F.1-3), (H.1-3) and (K.1-3), as $n \rightarrow \infty$, with probability one,*

$$\left| \left\{ 2nh_n \log(h_n^{-1}) \right\}^{-1/2} \sup_{x \in I} \left\{ \pm W_{n,j}(x, \psi) \right\} - \sigma_W(I) \right| = o(1),$$

where

$$\sigma_W^2(I) = \sup_{x \in I} \mathbb{E} \left[(c(x)\psi(Y) + d(x))^2 | X = x \right] f(x) \int_{\mathbb{R}} [K_j(u)]^2 du.$$

Proof. The methodology of proof is exactly the same as in Einmahl et Mason (2000), Theorem 1, p .4. To obtain the upper bound, the authors use a maximal form of the Bernstein's inequality to estimate the supremum of the deviation on the nodes of a discrete grid and then use the Talagrand's inequality combined with a moment bound to control the difference between the original empirical processes and its values over the nodes of the grid. The main argument of the proof is based on Talagrand's exponential equality [18] for general empirical processes combined with an useful moment inequality for empirical processes indexed by classes of functions of Vapnik-Cervonenkis type (see, e.g., [4], [5], [10] and [11]). Regarding the inner bound part, they apply poissonization techniques in the border of the limiting set $\sigma_W^2(I)$. In our case, we notice that the function $K_j(u) = u^j [K(u)]$, $j \geq 0$, is clearly a function with bounded variation on \mathbb{R} . Thus K_j fulfills also (K.1-2). By following the steps of their Theorem 1 and by replacing K by K_j , we obtain easily Theorem 4.1. Details are omitted. \square

The deviation $\hat{m}_{\psi;n}^{LL}(x) - \tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{LL}(x)]$ cannot be expressed as a linear functional of the bivariate empirical process. However we are going to show that the stochastic deviation behaves asymptotically like its linearized version which corresponds to a special case of the general process in (4.1). Set

$$\epsilon_n := \sup_{x \in I} \left\{ \left\{ \hat{m}_{\psi;n}^{LL}(x) - \tilde{\mathbb{E}}[\hat{m}_{\psi;n}^{LL}(x)] \right\} - \frac{1}{f(x)} \left\{ \left\{ \hat{r}_{\psi;n}(x) - r_{\psi;n}(x) \right\} - m_{\psi}(x) \left\{ \hat{f}_n(x) - f_n(x) \right\} \right\} \right\}.$$

Notice that the right-hand side above is exactly the process $W_{n,j}(x, \psi)$ defined in (4.1), where we have chosen $c(x) = 1/f(x)$ and $d(x) = -m_{\psi}(x)/f(x)$ which are uniformly continuous on J via (F.1-3) in conjunction with Scheffé's lemma. The proof of Theorem 3.1 boils down to the following lemma:

Lemma 1 *Under the assumptions of Theorem 3.1, we have, with probability one,*

$$\epsilon_n = o \left(\left\{ \frac{nh_n}{2 \log(h_n^{-1})} \right\}^{-1/2} \right).$$

Proof. Applying Theorem 4.1 with $c(x) = 0$, $d(x) = 1$, $j = 0, 1, 2$, and then with $c(x) = 1$, $d(x) = 0$ and $j = 0, 1$, we get that

$$\left\{ \frac{nh_n}{2 \log(h_n^{-1})} \right\}^{1/2} \sup_{x \in I} \pm \{ \hat{f}_n(x) - f_n(x) \} \stackrel{a.s.}{=} O(1), \quad (4.2)$$

$$\left\{ \frac{nh_n}{2 \log(h_n^{-1})} \right\}^{1/2} \sup_{x \in I} \pm \{ \hat{f}_{n,1}(x) - f_{n,1}(x) \} \stackrel{a.s.}{=} O(1), \quad (4.3)$$

$$\left\{ \frac{nh_n}{2 \log(h_n^{-1})} \right\}^{1/2} \sup_{x \in I} \pm \{ \hat{f}_{n,2}(x) - f_{n,2}(x) \} \stackrel{a.s.}{=} O(1), \quad (4.4)$$

$$\left\{ \frac{nh_n}{2 \log(h_n^{-1})} \right\}^{1/2} \sup_{x \in I} \pm \{ \hat{r}_{\psi;n}(x) - r_{\psi;n}(x) \} \stackrel{a.s.}{=} O(1), \quad (4.5)$$

$$\left\{ \frac{nh_n}{2 \log(h_n^{-1})} \right\}^{1/2} \sup_{x \in I} \pm \{ \hat{r}_{\psi;n,1}(x) - r_{\psi;n,1}(x) \} \stackrel{a.s.}{=} O(1). \quad (4.6)$$

Under (F.1–3), (K.1–4) and (H.1), we obtain, via Bochner's Lemma,

$$f_n(x) = f(x) + o(1), \quad (4.7)$$

$$f_{n,1}(x) = f(x)\mu_1(K) + o(1) = o(1), \quad (4.8)$$

$$f_{n,2}(x) = f(x)\mu_2(K) + o(1), \quad (4.9)$$

$$r_{\psi;n}(x) = r_\psi(x) + o(1), \quad (4.9)$$

$$r_{\psi;n,1}(x) = r_\psi(x)\mu_1(K) + o(1) = o(1). \quad (4.10)$$

We leave temporarily the dependence on x and ψ ,

$$\begin{aligned} \hat{m}_n^{LL} - \tilde{\mathbb{E}}[\hat{m}_n^{LL}] &= \frac{\{\hat{r}_n \hat{f}_{n,2} - \hat{r}_{n,1} \hat{f}_{n,1}\}}{\hat{f}_n \hat{f}_{n,2} - \hat{f}_{n,1}^2} - \frac{\{r_n f_{n,2} - r_{n,1} f_{n,1}\}}{f_n f_{n,2} - f_{n,1}^2} \\ &= \frac{\{\hat{r}_n \hat{f}_{n,2} - r_n f_{n,2} + r_{n,1} f_{n,1} - \hat{r}_{n,1} \hat{f}_{n,1}\}}{\hat{f}_n \hat{f}_{n,2} - \hat{f}_{n,1}^2} + \frac{\{r_n f_{n,2} - r_{n,1} f_{n,1}\} \{f_n f_{n,2} - f_{n,1}^2 - \hat{f}_n \hat{f}_{n,2} + \hat{f}_{n,1}^2\}}{\{\hat{f}_n \hat{f}_{n,2} - \hat{f}_{n,1}^2\} \times \{f_n f_{n,2} - f_{n,1}^2\}} \\ &\stackrel{a.s.}{=} \left\{ \frac{1}{\hat{f}_n \hat{f}_{n,2}} \times \left\{ \{\hat{r}_n - r_n\} \hat{f}_{n,2} + r_n \{\hat{f}_{n,2} - f_{n,2}\} + \{r_{n,1} - \hat{r}_{n,1}\} f_{n,1} + \hat{r}_{n,1} \{f_{n,1} - \hat{f}_{n,1}\} \right\} \right. \\ &\quad \left. + \frac{r_n}{\hat{f}_n \hat{f}_{n,2} f_n} \times \left\{ \{f_n - \hat{f}_n\} \hat{f}_{n,2} + f_n \{f_{n,2} - \hat{f}_{n,2}\} + \{\hat{f}_{n,1} - f_{n,1}\} \{\hat{f}_{n,1} + f_{n,1}\} \right\} \right\} \\ &\quad + o\left(\left\{ \frac{nh_n}{2 \log(h_n^{-1})} \right\}^{-1/2} \right) \\ &\stackrel{a.s.}{=} \frac{\hat{r}_n - r_n}{\hat{f}_n} + \frac{f_n - \hat{f}_n}{\hat{f}_n f_n} \times r_n + o\left(\left\{ \frac{nh_n}{2 \log(h_n^{-1})} \right\}^{-1/2} \right) \quad [(4.6)+(4.8) \text{ and } (4.3)+(4.10)] \\ &\stackrel{a.s.}{=} \frac{1}{f} \times \{\hat{r}_n - r_n\} - \frac{m}{f} \times \{\hat{f}_n - f_n\} + o\left(\left\{ \frac{nh_n}{2 \log(h_n^{-1})} \right\}^{-1/2} \right) \quad [(4.7)+(4.9)], \end{aligned}$$

from which the lemma follows. \square

Corollary 3.1 is a straightforward consequence of Theorem 3.1. The class of functions $\psi_t(y) = \mathbb{I}_{\{y \leq t\}}$ is clearly a bounded pointwise measurable VC subgraph class of real valued functions. Thus we can equally obtain (3.2), which is a bias improved version of Corollary 2 in Einmahl and Mason [4], p. 6. Finally the proof of Proposition 3.1 follows after some basic analysis.

References

- [1] Collomb, G. (1979). Conditions nécessaires et suffisantes de convergence uniforme d'un estimateur de la régression, estimation des dérivées de la régression. *C. R. Acad. Sci. Paris*, **288**, 161-163.
- [2] Deheuvels, P. et Mason, D. M. (2004). General asymptotic confidence bands based on kernel-type function estimators. *Stat. Infer. Stoc. Processes*, **7.3**, 225-277.
- [3] Dudley, R. M. (1999). *Uniform Central Limit Theorems*. Cambridge university press.
- [4] Einmahl, U., Mason, D.M. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *Journal of Theoretical Probability*, **13.1**, 1-37.
- [5] Einmahl, U. et Mason, D.M. (2005). Uniform in bandwidth consistency of kernel-type functions estimators. *Ann. Statist.*, **33.3**, (to appear).
- [6] Fan, J. (1992). Design-adaptative nonparametric regression. *J. Amer. Statist. Assoc.*, **87**, 998-1004.
- [7] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Monographs on Statistics and Applied Probability, 66. Chapman & Hall, London.
- [8] Finkelstein, H. (1971), The law of the iterated logarithm for empirical distributions. *Ann. Math. Statist.*, **42**, 607-615.
- [9] Gasser, T., Müller, H.G. and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc. B*, **47**, 238-252.
- [10] Giné, E. et Guillou, A. (2001). On consistency of kernel density estimators for randomly censored data: Rates holding uniformly over adaptative intervals. *Ann. Inst. H. Poincaré Probab. Statist.*, **37**, 503-522.
- [11] Giné, E. et Guillou, A. (2002). Rates on strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. H. Poincaré Probab. Statist.*, **38.6**, 907-921.
- [12] Hobson, E.W. (1927), *The Theory of Functions of a Real Variable and the Theory of Fourier Series*. **1**, 3rd. ed. Cambridge Univ. Press.
- [13] Mason, D. M. (2003). A uniform functional law of the logarithm for a local Gaussian process. In *High Dimensional Probability III* (J. Hoffmann-Jørgensen, M. B. Marcus and J. A. Wellner, eds) 135-151. Birkhäuser, Boston
- [14] Mason, D. M. (2004). A uniform functional law of the logarithm for the local empirical process. *Ann. Probab.*, **32.2**, 1391-1418.
- [15] Nadaraya, E.A. (1964), On estimating regression. *Theor. Prob. Appl.*, **9**, 141-142.
- [16] Nolan, D. and Polard, D. (1987). U-processes: rates of convergence. *Ann. of Statist.*, **15.2**, 780-799.
- [17] Ruppert, D. et Wand, M. P. (1994). Multivariate weighted least squares regression. *Ann. Statist.*, **22.3**, 1346-1370.
- [18] Talagrand, M. (1994). Sharper bounds for Gaussian and empirical processes. *Ann. Probab.*, **22.1**, 28-76.
- [19] Wand, M. P. et Jones, M. C. (1995). *Kernel Smoothing*. Chapman and Hall, London.
- [20] Watson, G.S. (1964), Smooth regression analysis. *Sankhya Ser. A*, **26**, 359-372.