AUSTRIAN JOURNAL OF STATISTICS
Volume 35 (2006), Number 2&3, 105–120

Uniform in Bandwidth Consistency of Local Polynomial Regression Function Estimators

Julia Dony[1], Uwe Einmahl[1], and David M. Mason[2]
[1]Vrije Universiteit Brussel, Belgium
[2]University of Delaware, U.S.A.

Abstract: We generalize a method for proving uniform in bandwidth consistency results for kernel type estimators developed by the two last named authors. Such results are shown to be useful in establishing consistency of local polynomial estimators of the regression function.

Keywords: Consistent Estimators, Kernels, Density, Local Polynomial Approach.

1 Introduction

Let $X, X_1, X_2, \ldots$ be i.i.d. $\mathbb{R}^d$ ($d \geq 1$) valued random variables and assume that the common distribution function of these variables has a Lebesgue density function, which we shall denote by f_X. A kernel K will be any measurable function which satisfies the following conditions:

(K.i) $$\int_{\mathbb{R}^d} K(s)ds = 1\,, \quad \text{and}$$

(K.ii) $$\|K\|_\infty := \sup_{x \in \mathbb{R}^d} |K(x)| = \kappa < \infty\,.$$

The kernel density estimator of f_X based upon the sample $X_1, \ldots, X_n$ and bandwidth $0 < h < 1$ is

$$\widehat{f}_{n,h}(x) := \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h^{1/d}}\right)\,, \quad x \in \mathbb{R}^d\,.$$

It is well known that if one chooses a suitable bandwidth sequence $h_n \to 0$ and the density f_X is continuous, one obtains a strongly consistent estimator $\widehat{f}_n := \widehat{f}_{n,h_n}$ of f_X, i.e. one has with probability 1, $\widehat{f}_n(x) \to f_X(x)$, $x \in \mathbb{R}^d$. It is also natural to investigate other modes of convergence, for instance uniform convergence and to ask what convergence rates are feasible.

For proving such results, one usually writes the difference $\widehat{f}_n(x) - f_X(x)$ as the sum of a probabilistic term $\widehat{f}_n(x) - \mathbb{E}\widehat{f}_n(x)$ and a deterministic term $\mathbb{E}\widehat{f}_n(x) - f_X(x)$, the so-called bias. The order of the bias depends on smoothness properties of f_X only, whereas the first (random) term can be studied via empirical process techniques as has been pointed out by Stute (1982a), Stute (1982b), Stute (1984) and Pollard (1984), among other authors.

Giné and Guillou (2002) (see also Deheuvels, 2000 for the one-dimensional case) have shown that if K is a regular kernel, the density function f_X is bounded and h_n satisfies the regularity conditions $h_n \searrow 0$, h_n/h_{2n} is bounded, and

$$\log(1/h_n)/\log\log n \to \infty \quad \text{and} \quad nh_n/\log n \to \infty\,,$$

one has with probability 1,

$$\|\widehat{f}_n - \mathbb{E}\widehat{f}_n\|_\infty = O\left(\sqrt{|\log h_n|/nh_n}\right) ,$$

where $||\cdot||_\infty$ denotes the supremum norm on $\mathbb{R}^d$. Moreover, this rate cannot be improved. Interestingly one does not need continuity of f_X for this result. (Continuity of f_X is of course needed for controlling the bias.)

Recently, Einmahl and Mason (2005) have provided a uniform in h version of this result, i.e., they have proved that

$$\limsup_{n\to\infty} \sup_{\frac{c\log n}{n}\leq h\leq 1} \frac{\sqrt{nh}\|\widehat{f}_{n,h} - \mathbb{E}\widehat{f}_{n,h}\|_\infty}{\sqrt{|\log h_n| \vee \log\log n}} =: K(c) < \infty\,. \tag{1}$$

This result implies that if one chooses the bandwidth depending on the data and/or the location x, as is usually done in practice, one has the same order of convergence as in the case of a deterministic bandwidth sequence.

Now let $Y, Y_1, Y_2, \ldots$ be a sequence of r-dimensional random vectors ($r \geq 1$) so that the random vectors $(X,Y),(X_1,Y_1),\ldots$ are i.i.d. with common joint Lebesgue density function f. In this case it is also of great interest to estimate $\mathbb{E}\left[\psi(Y)|X = x\right]$, where $\psi : \mathbb{R}^r \to \mathbb{R}$ is a suitable mapping. A possible kernel type estimator which reduces to the classical Nadaraya-Watson estimator if $r = 1$, $\psi(y) = y$, is given by

$$\widehat{m}_n(x,\psi) = \frac{\sum_{i=1}^n \psi(Y_i)K((x - X_i)/h_n)}{\sum_{i=1}^n K((x - X_i)/h_n)}\,. \tag{2}$$

Likewise by setting in the one-dimensional case for $t \in \mathbb{R}$, $\psi_t(y) = I_{]-\infty,t]}(y)$, $y \in \mathbb{R}$, we obtain the kernel estimator of the conditional empirical function

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

given by

$$\widehat{F}_n(t|x) := \frac{\sum_{i=1}^n 1(Y_i \leq t)K((x - X_i)/h_n)}{\sum_{i=1}^n K((x - X_i)/h_n)}\,.$$

This kernel estimator is called the conditional empirical distribution function and was first extensively studied by Stute (1986). Exact convergence rates uniformly on compact subsets of $\mathbb{R}^d$ have been obtained for both Nadaraya-Watson type estimators as in (2) and the conditional empirical distribution function by Einmahl and Mason (2000) in the case of deterministic bandwidth sequences. Recently, Einmahl and Mason (2005) have established uniform in bandwidth results for these estimators which are of a similar type as result (1). The proof of these results requires establishing a suitable version of a result of type (1) for processes of the form

$$\frac{1}{nh}\sum_{i=1}^n \left\{\varphi(Y_i)K\left(\frac{x - X_i}{h^{1/d}}\right) - \mathbb{E}\left[\varphi(Y)K\left(\frac{x - X_i}{h^{1/d}}\right)\right]\right\} ,$$

where $x \in I$ (I a compact subset of $\mathbb{R}^d$ or $I = \mathbb{R}^d$) and $\varphi \in \Phi$, where Φ is a suitable class of functions.

For certain applications, however, this class of processes could be too small. One of the purposes of this paper is to establish such uniform in bandwidth consistency results for a larger class of processes. As an application of our results, we shall prove uniform in bandwidth consistency of local polynomial regression estimators. Such estimators are generalizations of the classic Nadaraya-Watson estimator (see, especially, Fan and Gijbels, 1996 and Tsybakov, 2004). In Section 2 we will state two general consistency results, one of which will be proved in Section 3. In Section 4 we treat the local polynomial regression estimators. In an appendix we gather together some facts needed in our proofs.

2 General Consistency Results

We shall begin by stating a result proved in Einmahl and Mason (2005), which will be instrumental in establishing uniform in bandwidth consistency of local polynomial regression function estimators. Let Φ denote a class of measurable functions on $\mathbb{R}^r$ with a finite valued measurable envelope function F,

$$F(y) \geq \sup_{\varphi \in \Phi} |\varphi(y)| , \quad y \in \mathbb{R}^r .$$

Further assume that Φ is pointwise measurable and satisfies (A.2) in the Appendix with $\mathcal{G}$ replaced by Φ. (For the definition of pointwise measurable also refer to the Appendix.) Consider the following class of functions

$$\mathcal{K} = \left\{ K((x - \cdot)/h^{1/d}) : h > 0 \, , \; x \in \mathbb{R}^d \right\} , \tag{3}$$

and assume that $\mathcal{K}$ is pointwise measurable and satisfies (A.2) with $\mathcal{G}$ replaced by $\mathcal{K}$. Introduce the class of continuous functions on a compact subset J of $\mathbb{R}^d$ indexed by Φ:

$$\mathcal{C} := \{ c_\varphi : \varphi \in \Phi \} .$$

We shall always assume that the class $\mathcal{C}$ is relatively compact with respect to the sup-norm topology, which by the Arzela-Ascoli theorem is equivalent to being uniformly bounded and uniformly equicontinuous.

For any $\varphi \in \Phi$ and continuous functions c_φ on a compact subset J of $\mathbb{R}^d$, set for $x \in J$,

$$\eta_{\varphi,n,h}(x) = \sum_{i=1}^{n} c_\varphi(x) \varphi(Y_i) K\left(\frac{x - X_i}{h^{1/d}} \right) ,$$

where K is a kernel with *support contained in* $[-1/2, 1/2]^d$ such that (K.i) and (K.ii) hold. The following result was proved in Einmahl and Mason (2005), where it is stated as Proposition 2. ($\|\cdot\|_I$ denotes the supremum norm on I.)

Theorem 1. *Let I be a compact subset of $\mathbb{R}^d$ such that $J = I^\eta$, for some $0 < \eta < 1$. Also assume that*

$$f \text{ is continuous and strictly positive on } J. \tag{4}$$

Further assume that the envelope function F of the class Φ satisfies

$$\exists M > 0 : F(Y) 1\{X \in J\} \leq M , \quad a.s. \tag{5}$$

or for some $p > 2$

$$\alpha := \sup_{z \in J} \mathbb{E}[F^p(Y)|X = z] < \infty \,. \tag{6}$$

Then we have for any $c > 0$ *and* $0 < h_0 < (2\eta)^d$*, with probability 1,*

$$\limsup_{n\to\infty} \sup_{c(\log n/n)^\gamma \le h \le h_0} \frac{\sup_{\varphi\in\Phi} \|\eta_{\varphi,n,h} - \mathbb{E}\eta_{\varphi,n,h}\|_I}{\sqrt{nh\,(|\log h| \vee \log\log n)}} =: Q(c) < \infty \,,$$

where $\gamma = 1$ *in the bounded case (5) and* $\gamma = 1 - 2/p$ *under assumption (6).*

The next result generalizes Theorem 1 in the bounded case. Its proof is illustrative of how that of Theorem 1 goes using an empirical process approach based upon an inequality of Talagrand coupled with a moment bound for the supremum of the empirical process. These basic tools are stated in the Appendix.

In the following, $|| \cdot ||_\infty$ denotes the supremum norm on $\mathbb{R}^d$ or $\mathbb{R}^{d+r}$, whichever is appropriate. Let $\mathcal{G}$ denote a class of measurable real valued functions g of $(u,t) \in \mathbb{R}^d \times \mathbb{R}^r = \mathbb{R}^{d+r}$. We shall assume that $\mathcal{G}$ satisfies:

(G.i) $\sup_{g\in\mathcal{G}} ||g||_\infty =: \kappa < \infty$;

(G.ii) $\sup_{g\in\mathcal{G}} \int_{\mathbb{R}^{d+r}} g^2(x,y)dxdy =: L < \infty$.

Denote by $\mathcal{F}_\mathcal{G}$ the class of functions of $(s,t) \in \mathbb{R}^{d+r}$ formed from $\mathcal{G}$ as follows:

$$\mathcal{F}_\mathcal{G} = \left\{ g(z - s\lambda, t) : \lambda \ge 1\,, z \in \mathbb{R}^d \text{ and } g \in \mathcal{G} \right\} \,.$$

We shall also assume that the class of functions $\mathcal{F}_\mathcal{G}$ satisfies the following uniform entropy condition:

(F.i) for some $C_0 > 0$ and $\nu_0 > 0$, $N(\epsilon, \mathcal{F}_\mathcal{G}) \le C_0 \epsilon^{-\nu_0}$, $0 < \epsilon < 1$.

Finally, to avoid using outer probability measures in all of our statements, we impose the measurability assumption:

(F.ii) $\mathcal{F}_\mathcal{G}$ is a pointwise measurable class.

(For the definitions of pointwise measurable and of $N(\epsilon, \mathcal{F}_\mathcal{G})$ see the Appendix below, where we use κ as our envelope function.)

For any $g \in \mathcal{G}$ and $0 < h < 1$ define,

$$g_{n,h}(x) := (nh)^{-1} \sum_{i=1}^n g\left(\frac{x - X_i}{h^{1/d}}\,, Y_i\right)\,, \quad x \in \mathbb{R}^d \,.$$

Theorem 2. *Assuming (G.i), (G.ii), (F.i), (F.ii), and* f *(the joint density of* (X,Y)*) bounded, we have for* $c > 0$ *and* $0 < h_0 < 1$*,*

$$\limsup_{n\to\infty} \sup_{\frac{c\log n}{n} \le h \le h_0} \sup_{g\in\mathcal{G}} \frac{\sqrt{nh}\, \|g_{n,h} - \mathbb{E}g_{n,h}\|_\infty}{\sqrt{|\log h| \vee \log\log n}} =: G(c) < \infty \,. \tag{7}$$

Remark. Theorem 2 is still valid for $r = 0$. In this case, $g : \mathbb{R}^d \to \mathbb{R}^d$ and condition (G.ii) should be read as $\sup_{g\in\mathcal{G}} \int_{\mathbb{R}^d} g^2(x)dx =: L < \infty$.

3 Proof of Theorem 2

Let α_n be the empirical process based on the sample $(X_1, Y_1), \dots, (X_n, Y_n)$, i.e. if $\varphi : \mathbb{R}^d \times \mathbb{R}^r \to \mathbb{R}$, we have

$$\alpha_n(\varphi) = \sum_{i=1}^{n} \Big(\varphi(X_i, Y_i) - \mathbb{E}\varphi(X, Y) \Big) / \sqrt{n} \,.$$

Notice that in this notation

$$g_{n,h}(x) - \mathbb{E}g_{n,h}(x) = \frac{1}{h\sqrt{n}} \alpha_n \left(g \left(\frac{x - \cdot}{h^{1/d}}, \cdot \right) \right) , \quad x \in \mathbb{R}^d$$

so we get that

$$\sup_{g \in \mathcal{G}} \frac{\sqrt{nh} \, \|g_{n,h} - \mathbb{E}g_{n,h}\|_\infty}{\sqrt{|\log h| \vee \log\log n}} = \sup_{g \in \mathcal{G}} \sup_{x \in \mathbb{R}^d} \frac{\left| \sqrt{n} \alpha_n \left(g \left(\frac{x - \cdot}{h^{1/d}}, \cdot \right) \right) \right|}{\sqrt{nh \left(|\log h| \vee \log\log n \right)}} ,$$

where $g((x - \cdot)/h^{1/d}, \cdot)$ denotes the function $(s, t) \to g((x - s)/h^{1/d}, t)$. We first note that by (G.ii) and the assumption that $||f||_\infty < \infty$,

$$\begin{aligned} \mathbb{E}\left[g^2 \left(\frac{x - X}{h^{1/d}}, Y \right) \right] &= h \int_{\mathbb{R}^d} \int_{\mathbb{R}^r} h^{-1} g^2 \left(\frac{x - s}{h^{1/d}}, t \right) f(s, t) ds dt \\ &\leq h ||f||_\infty L \,. \end{aligned}$$

Set for $j \geq 0$ and $c > 0$,

$$h_{j,n} := \left(2^j c \log n \right) / n$$

and

$$\mathcal{F}_{j,n} = \left\{ g((x - \cdot)/h^{1/d}, \cdot) : g \in \mathcal{G} \,, h_{j,n} \leq h \leq h_{j+1,n} \,, x \in \mathbb{R}^d \right\} \,.$$

Clearly for $h_{j,n} \leq h \leq h_{j+1,n}$,

$$\mathbb{E}\left[g^2 \left(\frac{x - X}{h^{1/d}}, Y \right) \right] \leq 2 h_{j,n} ||f||_\infty L =: D_0 h_{j,n} =: \sigma_{j,n}^2 \,.$$

We shall use Proposition A.1 in the Appendix to bound $\mathbb{E}\| \sum_{i=1}^n \epsilon_i \varphi(X_i, Y_i) \|_{\mathcal{F}_{j,n}}$. To that end we note that each $\mathcal{F}_{j,n}$ satisfies (A.1) of the proposition with $G = \beta = \kappa$ and (A.3) with $\sigma^2 = \sigma_{j,n}^2$. Further, since $\mathcal{F}_{j,n} \subset \mathcal{F}_{\mathcal{G}}$, we see by (F.i) that each $\mathcal{F}_{j,n}$ also fulfills (A.2). Finally (A.4) holds for large enough n and all $j \geq 0$. Now by applying Proposition A.1 we get for all large enough n and $j \geq 0$,

$$\mathbb{E}\| \sum_{i=1}^{n} \epsilon_i \varphi(X_i, Y_i) \|_{\mathcal{F}_{j,n}} \leq D_1 \sqrt{n h_{j,n} \left| \log \left(D_2 h_{j,n} \right) \right|} \,,$$

for some $D_1 > 0$ and $D_2 > 0$. Let for large enough n

$$l_n := \max \left\{ j : h_{j,n} \leq 2h_0 \right\} \,,$$

then a little calculation shows that

$$l_n \sim \frac{\log\left(\frac{nh_0}{c\log n}\right)}{\log 2} \,. \tag{8}$$

For $k \geq 1$, set $n_k = 2^k$, and let

$$c_{j,k} := \sqrt{n_k h_{j,n_k}\left(|\log D_2 h_{j,n_k}| \vee \log\log n_k\right)}\,, \quad j \geq 0\,.$$

Applying Inequality A.1 in the Appendix with

$$M = \kappa \quad \text{and} \quad \sigma^2_{\mathcal{G}} = \sigma^2_{\mathcal{F}_{j,n_k}} \leq D_0 h_{j,n_k}\,,$$

we get for any $t > 0$,

$$\mathbb{P}\left\{\max_{n_{k-1}\leq n\leq n_k} ||\sqrt{n}\alpha_n||_{\mathcal{F}_{j,n_k}} \geq A_1(D_1 c_{j,k} + t)\right\}$$
$$\leq 2\left[\exp\left(-A_2 t^2/\left(D_0 n_k h_{j,n_k}\right)\right) + \exp(-A_2 t/\kappa)\right]\,.$$

Set for any $\rho > 1$, $j \geq 0$ and $k \geq 1$,

$$p_{j,k}(\rho) := \mathbb{P}\left\{\max_{n_{k-1}\leq n\leq n_k} ||\sqrt{n}\alpha_n||_{\mathcal{F}_{j,n_k}} \geq A_1\left(D_1 + \rho\right)c_{j,k}\right\}\,.$$

As we have $c_{j,k}/\sqrt{n_k h_{j,n_k}} \geq \sqrt{\log\log n_k}$, we readily obtain for $j \geq 0$,

$$p_{j,k}(\rho) \leq 2\left[\exp\left(-\frac{\rho^2 A_2}{D_0}\log\log n_k\right) + \exp\left(-\frac{\sqrt{c}\rho A_2}{\kappa}\sqrt{\log n_k \log\log n_k}\right)\right]\,,$$

which for $\gamma = A_2/D_0 \wedge \sqrt{c}A_2/\kappa$ implies

$$p_{j,k}(\rho) \leq 4\exp\left(-\rho\gamma\log\log n_k\right)\,.$$

Thus

$$P_k(\rho) := \sum_{j=0}^{l_{n_k}-1} p_{j,k}(\rho) \leq 4 l_{n_k}\left(\log n_k\right)^{-\rho\gamma}\,,$$

which by (8), for all large k and large enough $\rho > 1$

$$P_k(\rho) \leq 8\left(\log n_k\right)^{1-\rho\gamma} = 8\left(\frac{1}{k\log 2}\right)^{\rho\gamma-1} \leq k^{-2}\,.$$

Notice that by definition of l_n, for large k

$$2h_{l_{n_k},n_k} = h_{l_{n_k}+1,n_k} \geq 2h_0\,,$$

which implies that we have for $n_{k-1} \leq n \leq n_k$

$$\left[\frac{c\log n}{n}, h_0\right] \subset \left[\frac{c\log n_k}{n_k}, h_{l_{n_k},n_k}\right]\,.$$

Thus for all large enough k and $n_{k-1} \le n \le n_k$,

$$A_k(\rho) := \left\{ \max_{n_{k-1}\le n\le n_k} \sup_{g\in\mathcal{G}} \sup_{\frac{c\log n}{n}\le h\le h_0} \frac{\sqrt{nh}\|g_{n,h} - \mathbb{E}g_{n,h}\|_\infty}{\sqrt{|\log h| \vee \log\log n}} > 2A_1(D_1+\rho) \right\}$$

$$\subset \bigcup_{j=0}^{l_{n_k}-1} \left\{ \max_{n_{k-1}\le n\le n_k} ||\sqrt{n}\alpha_n||_{\mathcal{F}_{j,n_k}} \ge A_1(D_1+\rho)c_{j,k} \right\} .$$

It follows now for large enough ρ that

$$\mathbb{P}\left\{A_k(\rho)\right\} \le P_k(\rho) \le k^{-2} ,$$

which by the Borel-Cantelli lemma implies our theorem. □

4 Application to Local Polynomial Regression Function Estimators

In this section we shall always assume that the assumptions of Theorem 1 hold (in particular, that K has support contained in $[-1/2, 1/2]$) and I is a fixed compact interval in $\mathbb{R}$. We shall also assume that $K \ge 0$.

4.1 Estimating the Regression Function by Local Polynomials

Let $(X,Y), (X_1,Y_1), \ldots, (X_n,Y_n)$ be i.i.d. two-dimensional random vectors and write

$$g(x) := \mathbb{E}\left[Y||X = x\right]$$

for the regression function. Suppose that $g(x)$ is $(p+1)$ times differentiable on $J = I^\eta$, then we can approximate $g(x)$ locally around $x_0 \in I$ by a polynomial of order p (Taylor):

$$g(x) \approx g(x_0) + g^t(x_0)(x-x_0) + \ldots + \frac{g^{(p)}(x_0)}{p!}(x-x_0)^p .$$

Then consider the weighted least-squares regression problem (WLS)

$$\text{argmin}_{\boldsymbol{\beta}\in\mathbb{R}^{p+1}} \frac{1}{nh} \sum_{i=1}^{n} \left[Y_i - \sum_{j=0}^{p} \beta_j (X_i - x_0)^j \right]^2 K\left(\frac{x_0 - X_i}{h}\right) . \tag{9}$$

It is clear that if $\hat{\boldsymbol{\beta}} \in \mathbb{R}^{p+1}$ is the solution of the WLS problem in (9), we obtain an estimator $\hat{g}_{n,h}^{(p)}(x_0)$ of $g(x_0)$ by taking it be $\hat{\beta}_0$, the first component of $\hat{\boldsymbol{\beta}}$. At the same time we obtain estimators of the derivatives of the regression function up to order p. To solve (9), first note that it can be written in a matrix notation:

$$\text{argmin}_{\boldsymbol{\beta}\in\mathbb{R}^{p+1}} \left(\mathbf{Y} - \mathbf{X}_{x_0}\boldsymbol{\beta}\right)^t \mathbf{W}_{x_0} \left(\mathbf{Y} - \mathbf{X}_{x_0}\boldsymbol{\beta}\right) ,$$

where $\mathbf{W}_{x_0} = (nh)^{-1}\text{diag}\left(K\left((x_0 - X_i)/h\right)\right) \in \mathbb{R}^{n\times n}$, and $\mathbf{X}_{x_0} \in \mathbb{R}^{n\times(p+1)}$, $\mathbf{Y} \in \mathbb{R}^{n\times 1}$ and $\boldsymbol{\beta} \in \mathbb{R}^{(p+1)\times 1}$ are defined as

$$\mathbf{X}_{x_0} := \begin{pmatrix} 1 & (X_1 - x_0) & \cdots & (X_1 - x_0)^p \\ \vdots & \vdots & & \vdots \\ 1 & (X_n - x_0) & \cdots & (X_n - x_0)^p \end{pmatrix}, \quad \mathbf{Y} := \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad \boldsymbol{\beta} := \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix}.$$

If we set

$$L(x_0) := \frac{1}{nh}\sum_{i=1}^{n}\left[Y_i - \sum_{j=0}^{p}\beta_j (X_i - x_0)^j\right]^2 K\left(\frac{x_0 - X_i}{h}\right),$$

it is not too difficult to see that for $k = 0, \ldots, p$, the partial derivatives can be written as

$$\frac{\partial L(x_0)}{\partial \beta_k} = -2(\mathbf{Y} - \mathbf{X}_{x_0}\boldsymbol{\beta})^t \mathbf{W}_{x_0}\mathbf{X}_{x_0}\mathbf{e}_k^t,$$

where $\mathbf{e}_k$ is the k-th unit vector in $\mathbb{R}^{p+1}$. So by setting the partial derivatives equal to zero, we obtain that the solution $\hat{\boldsymbol{\beta}}$ of the WLS problem (9) must satisfy

$$\mathbf{Y}^t\mathbf{W}_{x_0}\mathbf{X}_{x_0} = \hat{\boldsymbol{\beta}}^t\mathbf{X}_{x_0}^t\mathbf{W}_{x_0}\mathbf{X}_{x_0}.$$

Assuming that

$$\mathcal{S}_{x_0} := \mathbf{X}_{x_0}^t\mathbf{W}_{x_0}\mathbf{X}_{x_0},$$

is invertible, we can compute the solution by

$$\hat{\boldsymbol{\beta}}_{x_0} = \left(\mathbf{X}_{x_0}^t\mathbf{W}_{x_0}\mathbf{X}_{x_0}\right)^{-1}\mathbf{X}_{x_0}^t\mathbf{W}_{x_0}\mathbf{Y}.$$

We shall show that *asymptotically* the inverse matrix of $\mathcal{S}_{x_0}$ always exists. To see this, consider for $0 \le j \le 2p$ the functions

$$H^{(j)}(u) := (-u)^j K(u).$$

Since we assume K to be bounded with support contained in $[-1/2, 1/2]$, we see that each $H^{(j)} \in L_1(\mathbb{R})$ and has support contained in $[-1/2, 1/2]$. Now for each $j \ge 0$ define the bounded function

$$\phi_j(u) = (-u)^j 1\left\{u \in [-1/2, 1/2]\right\}.$$

Since this function is of bounded variation, the class

$$\{\phi_j((x - \cdot)/h) : h > 0,\ x \in \mathbb{R}\}$$

satisfies (A.2). (See Lemma 22 of Nolan and Pollard (1987).) Thus the class $\mathcal{K}$, as defined in (3) is assumed to be pointwise measurable and satisfies (A.2). By Lemma A.1 in the Appendix, for each $j = 0, \ldots, 2p$, the class

$$\mathcal{G}_j := \left\{H^{(j)}((x - \cdot)/h) : h > 0, x \in \mathbb{R}\right\}$$

also fulfills (A.2). Moreover, it is easily checked that each $\mathcal{G}_j$ is pointwise measurable. Hence the assumptions of Theorem 2 hold and we can infer that for each $0 \le j \le 2p$, and sequence a_n satisfying

$$a_n \searrow 0 \quad \text{and} \quad na_n/\log n \to \infty\,, \tag{10}$$

we have

$$\sup_{x_0 \in I} \sup_{a_n \le h \le h_0} \left| H_{n,h}^{(j)}(x_0) - \mathbb{E}H_{n,h}^{(j)}(x_0) \right| \to 0\,, \quad a.s.\,, \tag{11}$$

where

$$H_{n,h}^{(j)}(x_0) := \frac{1}{nh} \sum_{i=1}^{n} H^{(j)} \left(\frac{x_0 - X_i}{h} \right).$$

Notice that

$$\mathbb{E}H_{n,h}^{(j)}(x_0) = \frac{1}{h} \int_{\mathbb{R}} H^{(j)} \left(\frac{x_0 - t}{h} \right) f(t)dt =: f * H_h^{(j)}(x_0)\,,$$

and since f is continuous on $J = I^\eta$ with I being a compact interval, we can use Lemma A.2 in the Appendix to get that as $h \searrow 0$,

$$\sup_{x_0 \in I} \left| \mathbb{E}H_{n,h}^{(j)}(x_0) - f(x_0) \int_{\mathbb{R}} (-u)^j K(u)du \right| \to 0\,. \tag{12}$$

Hence, it follows immediately by (11) and (12), that uniformly in $x_0 \in I$ and for $a_n < b_n$ with a_n satisfying (10) and $b_n \searrow 0$,

$$\sup_{a_n \le h \le b_n} \left| H_{n,h}^{(j)}(x_0) - f(x_0) \int_{\mathbb{R}} (-u)^j K(u)du \right| \to 0\,, \quad a.s.$$

Next consider the Hilbert space $\mathcal{L}(\mathbb{R}, Kd\lambda)$ consisting of all the measurable functions $\phi : \mathbb{R} \to \mathbb{R}$ such that

$$\int_{\mathbb{R}} \phi^2(u)K(u)du < \infty\,.$$

As usual, $\phi_1 = \phi_2$ if $\int_{\mathbb{R}} (\phi_1 - \phi_2)^2(u)K(u)du = 0$; that is, each $\phi \in \mathcal{L}(\mathbb{R}, Kd\lambda)$ represents an equivalence class of functions. Now let

$$G := \left(\int_{\mathbb{R}} (-u)^{j+k} K(u)du \right)_{j=0,k=0}^{p,p}\,,$$

then G is the Gramian matrix of the set of functions $\{\varphi_j : \varphi_j(x) = (-x)^j,\, j = 0, \ldots, p\}$ and these functions belong to $\mathcal{L}(\mathbb{R}, Kd\lambda)$ since K has compact support. It is known that G is nonsingular if the functions are linearly independent. Hence, in our case, G will always be invertible. (Here we use $K \ge 0$ and $0 < \int_{\mathbb{R}} K(u)du < \infty$.) To see that $\mathcal{S}_{x_0}$ is invertible as well, recall that the function $M \to \det M$ with $M \in \mathcal{M}_{p+1}(\mathbb{R})$ is continuous, and that by (11) and (12), with probability one, the components of

$$A_{x_0} := \left(H_{n,h}^{(j+k)}(x_0) \right)_{j=0,k=0}^{p,p}\,,$$

converge uniformly in $x_0 \in I$ and $a_n \le h \le b_n$ with $b_n \searrow 0$ to those of $f(x_0)G$. Hence, since we assume f to be strictly positive on $J = I^\eta$, for n large enough, uniformly in $x_0 \in I$, we have $\det A_{x_0} > 0$. Now let $\mathcal{H}_p := \text{diag}\{1, h, \ldots, h^p\}$, note that

$$\mathcal{S}_{x_0} = \mathcal{H}_p A_{x_0} \mathcal{H}_p \,,$$

and observe that $\det \mathcal{S}_{x_0} = h^{p(p+1)} \det A_{x_0}$, so for n large enough, uniformly in $x_0 \in I$ and $a_n \le h \le b_n$, $\mathcal{S}_{x_0}$ will have a positive determinant, showing that *asymptotically*, $\mathcal{S}_{x_0}$ is nonsingular and invertible.

From the above it follows that with probability one, for all large n, uniformly in $x_0 \in I$ and $a_n \le h \le b_n$, the local polynomial regression estimator of $g(x_0)$ is given by

$$\hat{g}_{n,h}^{(p)}(x_0) = \mathbf{e}_1 \mathcal{S}_{x_0}^{-1} \mathbf{X}_{x_0}^t \mathbf{W}_{x_0} \mathbf{Y} \,.$$

The difficulty is to determine $\mathcal{S}_{x_0}^{-1}$ explicitly, especially when p becomes large. Moreover, it is not possible to find a nice general formula for $\hat{g}_{n,h}^{(p)}(x_0)$, since the calculation of $\mathcal{S}_{x_0}^{-1}$ and $\hat{g}_{n,h}^{(p)}(x_0)$ becomes more complex as p increases. However, we shall see in the next section that $\hat{g}_{n,h}^{(p)}(x_0)$ can be easily computed for $p = 0, 1, 2$.

4.2 Uniform in Bandwidth Consistency

We shall now discuss uniform in bandwidth consistency of $\hat{g}_{n,h}^{(p)}$ on a compact interval I. Define the functions

$$\tilde{f}_{n,h,j}(x) := \frac{1}{nh} \sum_{i=1}^{n} \left(\frac{X_i - x}{h} \right)^j K\left(\frac{x - X_i}{h} \right) , \quad j = 0, \ldots, 2p \,,$$
$$\tilde{r}_{n,h,j}(x) := \frac{1}{nh} \sum_{i=1}^{n} Y_i \left(\frac{X_i - x}{h} \right)^j K\left(\frac{x - X_i}{h} \right) , \quad j = 0, \ldots, p \,.$$

By Theorem 2,

$$\limsup_{n\to\infty} \sup_{\frac{c \log n}{n} \le h \le h_0} \max_{0 \le j \le 2p} \frac{\sqrt{nh} \left\| \tilde{f}_{n,h,j} - \mathbb{E}\tilde{f}_{n,h,j} \right\|_I}{\sqrt{|\log h| \vee \log\log n}} < \infty \,, \quad a.s.$$

and by Theorem 1 with obvious identifications and K replaced by $H^{(j)}$,

$$\limsup_{n\to\infty} \sup_{c(\log n/n)^\gamma \le h \le h_0} \max_{0 \le j \le p} \frac{\|\tilde{r}_{n,h,j} - \mathbb{E}\tilde{r}_{n,h,j}\|_I}{\sqrt{nh(|\log h| \vee \log\log n)}} < \infty \,, \quad a.s.$$

For $j \ge 0$, set

$$\mu_j := \int_{\mathbb{R}} (-u)^j K(u) du \,,$$

and define

$$f_j(x) := \mu_j f_X(x) \,, \quad j = 0, \ldots, 2p \,,$$
$$r_j(x) := \mu_j \int_{\mathbb{R}} y f(x, y) dy \,, \quad j = 0, \ldots, p \,.$$

Lemma A.2 gives (also see (12)) that for all $0 \le j \le 2p$,

$$\sup_{a_n \le h \le b_n} \|\mathbb{E}\tilde{f}_{n,h,j} - f_j\|_\infty \to 0\,.$$

Now define the function

$$\varphi(x) := \int_{\mathbb{R}} y f(x,y)dy\,, \quad x \in J\,,$$

and introduce the assumption:

$$\forall x \in J\,,\ \lim_{x' \to x} f(x', y) = f(x,y) \text{ for almost every } y \in \mathbb{R}.$$

Then by an argument based on the Lebesgue dominated convergence theorem, using assumptions (4) along with (5) or (6), one readily shows that φ is bounded and continuous on J. Applying Lemma A.2, we get that for all $0 \le j \le p$,

$$\sup_{a_n \le h \le b_n} \|\mathbb{E}\tilde{r}_{n,h,j} - r_j\|_I \to 0\,.$$

From these observations, we easily conclude that for all smooth functions $\Phi : \mathbb{R}^{3p+2} \to \mathbb{R}$ and suitable sequences $0 < a_n < b_n$ depending on Theorem 1 and whether (5) or (6) holds, with probability 1,

$$\sup_{a_n \le h \le b_n} \left\| \Phi\left(\tilde{f}_{n,h,0}, \ldots, \tilde{f}_{n,h,2p}, \tilde{r}_{n,h,0}, \ldots, \tilde{r}_{n,h,p}\right) - \Phi\left(f_0, \ldots, f_{2p}, r_0, \ldots, r_p\right) \right\|_I \to 0\,. \tag{13}$$

When (5) is in force, we assume that a_n satisfies (10), and when (6) holds that $a_n = c(\log n/n)^\gamma$ for $\gamma > 1$.

Calculation for $p = 0$. In this case we get the usual Nadaraya-Watson regression estimator:

$$\hat{g}^{(0)}_{n,h}(x_0) = \frac{\sum_{i=1}^n Y_i K\left(\frac{x_0 - X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x_0 - X_i}{h}\right)} = \frac{\tilde{r}_{n,h,0}(x_0)}{\tilde{f}_{n,h,0}(x_0)}\,.$$

So applying (13) with $\Phi(x_1, x_2) = x_2/x_1$, we get that uniformly in $x_0 \in I$,

$$\sup_{a_n \le h \le b_n} \left\| \hat{g}^{(0)}_{n,h} - g \right\|_I \to 0\,, \quad a.s.\,,$$

proving the uniform in bandwidth consistency of the Nadaraya-Watson estimator.

From now on, for ease of notation we shall omit the subscripts x_0, as well as the argument (x_0) in all the functions that we defined above.

Calculation for $p = 1$. This is the local linear regression estimator, where $\mathcal{S}$ and $\mathbf{X}^t\mathbf{WY}$ are given by

$$\mathcal{S} = \begin{pmatrix} nh\tilde{f}_{n,h,0} & nh^2\tilde{f}_{n,h,1} \\ nh^2\tilde{f}_{n,h,1} & nh^3\tilde{f}_{n,h,2} \end{pmatrix}, \quad \mathbf{X}^t\mathbf{WY} = \begin{pmatrix} nh\tilde{r}_{n,h,0} \\ nh^2\tilde{r}_{n,h,1} \end{pmatrix},$$

such that

$$\mathcal{S}^{-1}\mathbf{X}^t\mathbf{WY} = \frac{1}{\tilde{f}_{n,h,0}\tilde{f}_{n,h,2} - \tilde{f}^2_{n,h,1}} \begin{pmatrix} \tilde{f}_{n,h,2}\tilde{r}_{n,h,0} - \tilde{f}_{n,h,1}\tilde{r}_{n,h,1} \\ \tilde{f}_{n,h,0}\tilde{r}_{n,h,1} - \tilde{f}_{n,h,1}\tilde{r}_{n,h,0} \end{pmatrix}.$$

Hence, the local linear estimator of the regression function is given by

$$\hat{g}^{(1)}_{n,h} = \frac{\tilde{f}_{n,h,2}\tilde{r}_{n,h,0} - \tilde{f}_{n,h,1}\tilde{r}_{n,h,1}}{\tilde{f}_{n,h,0}\tilde{f}_{n,h,2} - \tilde{f}^2_{n,h,1}}.$$

So applying (13) with $\Phi(x_1, \ldots, x_5) = (x_3x_4 - x_2x_5)/(x_1x_3 - x_2^2)$, we obtain after a little algebra based on the definitions of f_j and r_j, the uniform in bandwidth consistency of this local linear estimator:

$$\sup_{a_n \le h \le b_n} \left\| \hat{g}^{(1)}_{n,h} - g \right\|_I \to 0, \quad a.s.$$

Calculation for $p = 2$. As we have seen in the case $p = 1$, the main work in deriving $\hat{g}^{(2)}_{n,h}$ is to determine $\mathcal{S}^{-1}$. Now $\mathcal{S}$ is a 3×3-matrix, so we can still write down the inverse without difficulties. After some calculations, we obtain (disregarding nh^j factors):

$$\mathcal{S}^{-1} =$$

$$\frac{1}{\det \mathcal{S}} \begin{pmatrix} \tilde{f}_{n,h,2}\tilde{f}_{n,h,4} - \tilde{f}^2_{n,h,3} & \tilde{f}_{n,h,2}\tilde{f}_{n,h,3} - \tilde{f}_{n,h,1}\tilde{f}_{n,h,4} & \tilde{f}_{n,h,1}\tilde{f}_{n,h,3} - \tilde{f}^2_{n,h,2} \\ \tilde{f}_{n,h,2}\tilde{f}_{n,h,3} - \tilde{f}_{n,h,1}\tilde{f}_{n,h,4} & \tilde{f}_{n,h,0}\tilde{f}_{n,h,4} - \tilde{f}^2_{n,h,2} & \tilde{f}_{n,h,1}\tilde{f}_{n,h,2} - \tilde{f}_{n,h,0}\tilde{f}_{n,h,3} \\ \tilde{f}_{n,h,1}\tilde{f}_{n,h,3} - \tilde{f}^2_{n,h,2} & \tilde{f}_{n,h,1}\tilde{f}_{n,h,2} - \tilde{f}_{n,h,0}\tilde{f}_{n,h,3} & \tilde{f}_{n,h,0}\tilde{f}_{n,h,2} - \tilde{f}^2_{n,h,1} \end{pmatrix},$$

and

$$\mathbf{X}^t\mathbf{WY} = \begin{pmatrix} \tilde{r}_{n,h,0} \\ \tilde{r}_{n,h,1} \\ \tilde{r}_{n,h,2} \end{pmatrix},$$

eventually yielding

$$\hat{g}^{(2)}_{n,h} =$$

$$\frac{(\tilde{f}_{n,h,2}\tilde{f}_{n,h,4} - \tilde{f}^2_{n,h,3})\tilde{r}_{n,h,0} + (\tilde{f}_{n,h,2}\tilde{f}_{n,h,3} - \tilde{f}_{n,h,1}\tilde{f}_{n,h,4})\tilde{r}_{n,h,1} + (\tilde{f}_{n,h,1}\tilde{f}_{n,h,3} - \tilde{f}^2_{n,h,2})\tilde{r}_{n,h,2}}{\tilde{f}_{n,h,0}\tilde{f}_{n,h,2}\tilde{f}_{n,h,4} - \tilde{f}_{n,h,0}\tilde{f}^2_{n,h,3} - \tilde{f}^2_{n,h,1}\tilde{f}_{n,h,4} + 2\tilde{f}_{n,h,1}\tilde{f}_{n,h,2}\tilde{f}_{n,h,3} - \tilde{f}^3_{n,h,2}}.$$

So using the function

$$\Phi(x_1, \ldots, x_8) = \frac{(x_3x_5 - x_4^2)x_6 + (x_3x_4 - x_2x_5)x_7 + (x_2x_4 - x_3^2)x_8}{x_1x_3x_5 - x_1x_4^2 - x_2^2x_5 + 2x_2x_3x_4 - x_3^3}$$

in (13), we infer after some algebra based on the definitions of f_j and r_j, the uniform in bandwidth consistency of this local quadratic regression function estimator.

Calculation for larger p. In principle it is possible to write down an explicit formula for the local polynomial estimator $\hat{g}_n^{(p)}(x_0)$ for any $p \geq 0$, by first computing the inverse of $\mathcal{S}_{x_0}$, multiplying it by $\mathbf{X}_{x_0}^t \mathbf{W}_{x_0} \mathbf{Y}$ and then by taking the first component of the resulting vector. But the difficulty lies in determining $\mathcal{S}_{x_0}^{-1}$.

Remark. It was pointed in Einmahl and Mason (2005) and Blondin et al. (2005) that these methods can be used to study the uniform in bandwidth consistency of local polynomial regression estimators.

5 Appendix

Let $X, X_1, \ldots, X_n$ be i.i.d. from a probability space $(\mathcal{X}, \mathcal{A}, P)$ with common distribution μ. Let $\mathcal{G}$ be a *pointwise measurable class* of real valued functions defined on $\mathcal{X}$, i.e. we assume that there exists a countable subclass $\mathcal{G}_0$ of $\mathcal{G}$ so that we can find for any function g in $\mathcal{G}$ a sequence of functions $\{g_m\}$ in $\mathcal{G}_0$ for which $g_m(x) \to g(x)$, $x \in \mathcal{X}$. (See Example 2.3.4, van der Vaart and Wellner (1996).) Further let $\varepsilon_1, \ldots, \varepsilon_n$ be a sequence of independent Rademacher random variables independent of $X_1, \ldots, X_n$.

The following inequality is essentially due to Talagrand (1994) (see Einmahl and Mason, 2000).

Inequality A.1 *Let $\mathcal{G}$ be a pointwise measurable class of functions satisfying for some $0 < M < \infty$*

$$||g||_\infty \leq M\,, \quad g \in \mathcal{G}\,,$$

then for all $t > 0$ we have for suitable finite constants $A_1, A_2 > 0$,

$$\mathbb{P}\left\{ \max_{1 \leq m \leq n} ||\sqrt{m}\alpha_m||_{\mathcal{G}} \geq A_1 \left(\mathbb{E}|| \sum_{i=1}^n \varepsilon_i g(X_i)||_{\mathcal{G}} + t \right) \right\}$$

$$\leq 2(\exp(-A_2 t^2 / n\sigma_{\mathcal{G}}^2) + \exp(-A_2 t/M))\,,$$

where $\sigma_{\mathcal{G}}^2 = \sup_{g \in \mathcal{G}} Var(g(X))$.

It enables us to reduce many problems on almost sure convergence to investigating the moment quantity

$$\mu_n := \mathbb{E}|| \sum_{i=1}^n \varepsilon_i g(X_i)||_{\mathcal{G}}\,.$$

The following proposition proved in Einmahl and Mason (2000) is very helpful for obtaining bounds on this quantity, when the class $\mathcal{G}$ has a polynomial covering number. Let G be a finite valued measurable function satisfying for all $x \in \mathcal{X}$

$$G(x) \geq \sup_{g \in \mathcal{G}} |g(x)|\,,$$

and define

$$N(\epsilon, \mathcal{G}) := \sup_Q N\left(\epsilon\sqrt{Q(G^2)}, \mathcal{G}, d_Q\right)\,,$$

where the supremum is taken over all probability measures Q on $(\mathcal{X}, \mathcal{A})$ for which $0 < Q(G^2) < \infty$ and d_Q is the $L_2(Q)$-metric. As usual $N(\epsilon, \mathcal{G}, d)$ is the minimal number of balls $\{g : d(g, f) < \epsilon\}$ of d-radius ϵ needed to cover $\mathcal{G}$.

Proposition A.1 *Let $\mathcal{G}$ be a pointwise measurable class of bounded functions such that for some constants β, ν, $C > 1$, $\sigma \leq 1/(8C)$ and function G as above, the following four conditions hold:*

$$\mathbb{E}[G^2(X)] \leq \beta^2 \, ; \tag{A.1}$$

$$N(\epsilon, \mathcal{G}) \leq C\epsilon^{-\nu} \, , \quad 0 < \epsilon < 1 \, ; \tag{A.2}$$

$$\sigma_0^2 := \sup_{g \in \mathcal{G}} \mathbb{E}[g^2(X)] \leq \sigma^2 \, ; \tag{A.3}$$

$$\sup_{g \in \mathcal{G}} ||g||_\infty \leq \frac{1}{2\sqrt{\nu + 1}} \sqrt{n\sigma^2 / \log(\beta \vee 1/\sigma)} \, . \tag{A.4}$$

Then we have for a universal constant A

$$\mathbb{E}|| \sum_{i=1}^{n} \varepsilon_i g(X_i) ||_{\mathcal{G}} \leq A\sqrt{\nu n \sigma^2 \log(\beta \vee 1/\sigma)} \, .$$

Another version of Proposition A.1 has been proved by Giné and Guillou (2001). For refinements, consult Einmahl and Mason (2005) and Giné and Koltchinskii (2005).

We shall also require the following two lemmas. The first is proved in Einmahl and Mason (2000).

Here is Lemma A.1 of Einmahl and Mason (2000).

Lemma A.1 *Let $\mathcal{F}$ and $\mathcal{G}$ be two classes of real valued measurable functions on $\mathcal{X}$ satisfying*

$$|f(x)| \leq F(x) \, , \quad f \in \mathcal{F} \, , \; x \in \mathcal{X}$$

where F is a finite valued measurable envelope function on $\mathcal{X}$;

$$\| \, g \, \|_\infty \leq M \, , \quad g \in \mathcal{G} \, ,$$

where $M > 0$ is a finite constant. Assume that for all p-measures Q with $0 < Q(F^2) < \infty$,

$$N(\epsilon\sqrt{Q(F^2)}, \mathcal{F}, d_Q) \leq C_1 \epsilon^{-\nu_1} \, , \quad 0 < \epsilon < 1 \, ,$$

and for all p-measures Q,

$$N(\epsilon M, \mathcal{G}, d_Q) \leq C_2 \epsilon^{-\nu_2} \, , \quad 0 < \epsilon < 1 \, ,$$

where $\nu_1, \nu_2, C_1, C_2 \geq 1$ are suitable constants. Then we have for all p-measures Q, with $Q(F^2) < \infty$,

$$N(\epsilon M \sqrt{Q(F^2)}, \mathcal{F}\mathcal{G}, d_Q) \leq C_3 \epsilon^{-\nu_1 - \nu_2} \, , \quad 0 < \epsilon < 1 \, ,$$

for some finite constant $0 < C_3 < \infty$.

The next lemma can be inferred from results in Stein (1970, pp. 62-65).

Lemma A.2. *Let φ be a measurable function on $\mathbb{R}^d$, which for some $\gamma > 0$ is bounded and uniformly continuous on D_γ, where D is a closed subset of $\mathbb{R}^d$ and*

$$D_\gamma = \left\{x \in \mathbb{R}^d : |x - y| \leq \gamma\,, y \in D\right\}\,.$$

Then for any $L_1(\mathbb{R}^d)$ function H, which is equal to zero for $x \notin I^d$

$$\sup_{z\in D} |\varphi * H_h(z) - I(H)\varphi(z)| \to 0\,, \quad \textit{as } h \searrow 0\,,$$

where $I(H) = \int_{\mathbb{R}^d} H(u)du$ and $\varphi * H_h(z) := h^{-1} \int_{\mathbb{R}^d} \varphi(x) H\left(h^{-1/d}(z - x)\right) dx$.

Acknowledgements

David Mason's research was partially supported by an NSF Grant. Julia Dony's research is financed by a PhD grant from the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT Vlaanderen).

References

Blondin, D., Massiani, A., and Ribereau, P. (2005). Vitesses de convergence uniforme presque sûre d'estimateurs non-paramétriques de la régression. *C.R. Acad. Sci. Paris, Ser.*, *I 340*, 525-528.

Deheuvels, P. (2000). Uniform limit laws for kernel density estimators on possibly unbounded intervals. In N. Limnios and M. Nikulin (Eds.), *Recent Advances in Reliability Theory: Methodology, Practice and Inference* (p. 477-492). Basel: Birkhäuser.

Einmahl, U., and Mason, D. M. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *J. Theor. Probab.*, *13*, 1-37.

Einmahl, U., and Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Stat.*, *33*, 1380-1403.

Fan, J., and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications* (Monographs on Statistics and Applied Probability, 66 ed.). London: Chapman & Hall.

Giné, E., and Guillou, A. (2001). On consistency of kernel density estimators for randomly censored data: rates holding uniformly over adaptive intervals. *Ann. Inst. H. Poincaré*, *37*, 503-522.

Giné, E., and Guillou, A. (2002). Rates of strong consistency for multivariate kernel density estimators. *Ann. Inst. H. Poincaré*, *38*, 907-921.

Giné, E., and Koltchinskii, V. (2005). Concentration inequalities and asymptotic results for ratio type empirical processes. *Ann. Probab.*, to appear.

Nolan, D., and Pollard, D. (1987). U-processes: rates of convergence. *Ann. Statist.*, *15*, 780-799.

Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer-Verlag.

Stein, E. M. (1970). *Singular Integrals and Differentiability Properties of Functions*. Princeton, New Jersey: Princeton University Press.

Stute, W. (1982a). The oscillation behavior of empirical processes. *Ann. Probab*, *10*, 86-107.

Stute, W. (1982b). The law of the iterated logarithm for kernel density estimators. *Ann. Probab*, *10*, 414-422.

Stute, W. (1984). The oscillation behavior of empirical processes: the multivariate case. *Ann. Probab*, *12*, 361-379.

Stute, W. (1986). On almost sure convergence of conditional empirical distribution functions. *Ann. Probab*, *14*, 891-901.

Talagrand, M. (1994). Sharper bounds for gaussian and empirical processes. *Ann. Probab*, *22*, 28-76.

Tsybakov, A. B. (2004). Introduction à l'estimation non–paramétrique. In *Mathematics & applications.* Paris: Springer.

van der Vaart, A. W., and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. New York: Springer-Verlag.

Authors' addresses:

Julia Dony and Uwe Einmahl
Department of Mathematics
Vrije Universiteit Brussel
Pleinlaan 2
B-1050 Brussel
Belgium
Tel. +32 2 6293471
Fax +32 2 6293495
E-mail: jdony@vub.ac.be, ueinmahl@vub.ac.be

David M. Mason
Statistics Program
University of Delaware
206 Townsend Hall
Newark, Delaware 19717
USA
Tel. 1-302-831-1188
Fax 1-302-831-6243
E-mail: davidm@UDel.edu