KYBERNETIKA — VOLUME 47 (2011), NUMBER 6, PAGES 855–865

ONE BOOTSTRAP SUFFICES TO GENERATE SHARP UNIFORM BOUNDS IN FUNCTIONAL ESTIMATION

Paul Deheuvels

We consider, in the framework of multidimensional observations, nonparametric functional estimators, which include, as special cases, the Akaike–Parzen–Rosenblatt kernel density estimators ([1, 18, 20]), and the Nadaraya–Watson kernel regression estimators ([16, 22]). We evaluate the sup-norm, over a given set $\mathbf{I}$, of the difference between the estimator and a non-random functional centering factor (which reduces to the estimator mean for kernel density estimation). We show that, under suitable general conditions, this random quantity is consistently estimated by the sup-norm over $\mathbf{I}$ of the difference between the original estimator and a bootstrapped version of this estimator. This provides a simple and flexible way to evaluate the estimator accuracy, through a single bootstrap. The present work generalizes former results of Deheuvels and Derzko [4], given in the setup of density estimation in $\mathbb{R}$.

Keywords: nonparametric functional estimation, density estimation, regression estimation, bootstrap, resampling methods, confidence regions, empirical processes

Classification: 62G05, 62G08, 62G09, 62G15, 62G20, 62G30

1. KERNEL DENSITY AND REGRESSION ESTIMATORS

Let $(\mathbf{X}, \mathbf{Y})$ be a random vector, with $\mathbf{X} = (X_1, \ldots, X_p) \in \mathbb{R}^p$ and $\mathbf{Y} = (Y_1, \ldots, Y_q) \in \mathbb{R}^q$. We denote the joint distribution function [df] of $(\mathbf{X}, \mathbf{Y})$ by $F(\mathbf{x}, \mathbf{y}) = \mathbb{P}(\mathbf{X} \leq \mathbf{x}, \mathbf{Y} \leq \mathbf{y})$, for $\mathbf{x} \in \mathbb{R}^p$ and $\mathbf{y} \in \mathbb{R}^q$. Here, for $\mathbf{v}' = (v_1', \ldots, v_r') \in \mathbb{R}^r$ and $\mathbf{v}'' = (v_1'', \ldots, v_r'') \in \mathbb{R}^r$, we set $\mathbf{v}' \leq \mathbf{v}''$ whenever $v_j' \leq v_j''$ for $j = 1, \ldots, r$. Moreover, we set, when $\mathbf{v} = (v_1, \ldots, v_r) \in \mathbb{R}^r$,

$$(-\infty, \mathbf{v}] = \prod_{j=1}^{r} (-\infty, v_j].$$

Let $\psi : \mathbb{R}^q \to \mathbb{R}$ be a measurable function. In this paper, we are primarily concerned with the estimation of the conditional expectation (or regression) of $\mathcal{Y} := \psi(\mathbf{Y})$ given $\mathbf{X} = \mathbf{x}$,

$$m_\psi(\mathbf{x}) = \mathbb{E}\big(\mathcal{Y} | \mathbf{X} = \mathbf{x}\big) = \mathbb{E}\big(\psi(\mathbf{Y}) | \mathbf{X} = \mathbf{x}\big), \tag{1}$$

whenever it exists. We will work under the following notation, borrowed from Deheuvels and Mason [5]. We denote by $\mathbf{I}$ and $\mathbf{J}$ two fixed subsets of $\mathbb{R}^p$ such that

$$\mathbf{I} = \prod_{j=1}^{p} [a_j, b_j] \subset \mathbf{J} = \prod_{j=1}^{p} [c_j, d_j] \subset \mathbb{R}^p, \tag{2}$$

where

$$-\infty < c_j < a_j < b_j < d_j < \infty \quad \text{for} \quad j = 1, \ldots, p. \tag{3}$$

We assume that $(\mathbf{X}, \mathbf{Y})$ has a joint density $f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = \frac{\partial^{p+q}}{\partial x_1 \ldots \partial x_p \partial y_1 \ldots \partial y_q} F(\mathbf{x}, \mathbf{y})$ on $\mathbf{J} \times \mathbb{R}^q$, with respect to the Lebesgue measure $d\mathbf{x} \times d\mathbf{y}$, and denote by

$$f_{\mathbf{X}}(\mathbf{x}) = \int_{\mathbb{R}^q} f(\mathbf{x}, \mathbf{y}) \, \mathrm{d}\mathbf{y} \quad \text{for} \quad \mathbf{x} \in \mathbf{J}, \tag{4}$$

the density of $\mathbf{X}$ (which is only assumed to exist on $\mathbf{J}$). We denote by $\{\mathcal{M}(x) : x \geq 0\}$ a nonnegative continuous function, increasing on $[0, \infty)$, and such that, for some $\rho > 2$, ultimately as $x \uparrow \infty$,

$$(i) \quad x^{-\rho}\mathcal{M}(x) \downarrow; \quad (ii) \quad x^{-1}\mathcal{M}(x) \uparrow . \tag{5}$$

For each $t \geq \mathcal{M}(0)$, we define $\mathcal{M}^{\mathrm{inv}}(t) \geq 0$ by $\mathcal{M}(\mathcal{M}^{\mathrm{inv}}(t)) = t$. We assume further that:

$(F.1)$ $f_{\mathbf{X},\mathbf{Y}}$ is continuous on $\mathbf{J} \times \mathbb{R}^q$;

$(F.2)$ $f_{\mathbf{X}}$ is continuous and bounded away from 0 on $\mathbf{J}$;

$(F.3)$ $\beta_{\mathcal{M}}(\psi) := \sup_{\mathbf{x} \in \mathbf{J}} \mathbb{E}\Big(\mathcal{M}\big(|\psi(\mathbf{Y})|\big)\big|\mathbf{X} = \mathbf{x}\Big) < \infty$.

Under $(F.1$–2–$3)$, the regression function $m_\psi(\mathbf{x})$ exists for $\mathbf{x} \in \mathbf{J}$, and fulfills the equality

$$m_\psi(\mathbf{x}) \quad = \quad \frac{1}{f_{\mathbf{X}}(\mathbf{x})} \int_{\mathbb{R}^q} \psi(\mathbf{y}) f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}) \, \mathrm{d}\mathbf{y} = \frac{r_\psi(\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})}, \tag{6}$$

where we set, for $\mathbf{x} \in \mathbf{J}$,

$$r_\psi(\mathbf{x}) \quad = \quad \int_{\mathbb{R}^q} \psi(\mathbf{y}) f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}) \, \mathrm{d}\mathbf{y}. \tag{7}$$

When combined with (5), $(F.3)$ implies finiteness of the conditional variance $\sigma^2_\psi(\mathbf{x})$ of $\psi(\mathbf{Y})$ given that $\mathbf{X} = \mathbf{x}$, for $\mathbf{x} \in \mathbf{J}$. We have the relations, for $\mathbf{x} \in \mathbf{J}$,

$$\begin{aligned} \sigma^2_\psi(\mathbf{x}) \quad &= \quad \mathrm{Var}\big(\psi(\mathbf{Y})|\mathbf{X} = \mathbf{x}\big) \\ &= \quad \frac{1}{f_{\mathbf{X}}(\mathbf{x})} \int_{\mathbb{R}^q} \Big\{\psi(\mathbf{y}) - r_\psi(\mathbf{x})\Big\}^2 f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}) \, \mathrm{d}\mathbf{y}. \end{aligned} \tag{8}$$

Remark 1.1. The introduction of the function $\psi(\cdot)$ is motivated by $(F.1\text{–}2\text{–}3)$, which impose the existence of a density of $(\mathbf{X}, \mathbf{Y})$ on $\mathbf{J} \times \mathbb{R}^q$. Since $\psi(\cdot)$ is possibly discontinuous, the existence of a density of $(\mathbf{X}, \mathcal{Y}) = (\mathbf{X}, \psi(\mathbf{Y}))$ will not be required in our results.

We now introduce a *kernel function* $\{K(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^p\}$, fulfilling the conditions $(K.1\text{–}2)$ below (see, e. g., §3.2 in the sequel, and p. 227 in Deheuvels and Mason [5]). Below, we denote by $|\mathbf{v}|$ the Euclidian norm of $\mathbf{v} \in \mathbb{R}^r$ for arbitrary $r \geq 1$.

$(K.1)$ $\int_{\mathbb{R}^p} K(\mathbf{t})\, \mathrm{d}\mathbf{t} = 1$.

$(K.2)$ (i) $K(\cdot)$ is of bounded variation on $\mathbb{R}^p$ in the sense of Hardy and Krause, meaning that $K(d\mathbf{t})$ defines a totally bounded Lebesgue-Stieltjes signed measure on $\mathbb{R}^p$.

$(K.2)$ (ii) We may decompose $K(d\mathbf{t}) = \mu_+ - \mu_-$ into the difference of two orthogonal, compactly supported, non-negative (bounded) Radon measures on $\mathbb{R}^p$, such that

$$\|dK\| := \int_{\mathbb{R}^p} |K(\mathrm{d}\mathbf{t})| = \mu_+(\mathbb{R}^p) + \mu_-(\mathbb{R}^p) < \infty.$$

$(K.2)$ (iii) We may write $K = K^+ - K^-$, where $K^+(\mathbf{t}) = \mu_+((-\infty, \mathbf{t}])$ and $K^-(\mathbf{t}) = \mu_-((-\infty, \mathbf{t}])$ are right-continuous on $\mathbb{R}^p$.

The relevance of $(K.1\text{–}2)$ with respect to our theorems will be discussed in §3.2.

We next consider a sequence $\{(\mathbf{X}_i, \mathbf{Y}_i) : i \geq 1\}$ of independent and identically distributed [iid] random copies of the random vector [rv] $(\mathbf{X}, \mathbf{Y})$. For each $n \geq 1$, and for each choice of the bandwidth $h > 0$, we define the kernel estimators

$$f_{\mathbf{X};n}(\mathbf{x}; h) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h^{1/p}}\right), \tag{9}$$

$$r_{\psi;n}(\mathbf{x}; h) = \frac{1}{nh} \sum_{i=1}^{n} \psi(\mathbf{Y}_i) K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h^{1/p}}\right), \tag{10}$$

and set

$$m_{\psi;n}(\mathbf{x}; h) = \begin{cases} \dfrac{r_{\psi;n}(\mathbf{x}; h)}{f_{\mathbf{X};n}(\mathbf{x}; h)} & \text{for} \quad f_{\mathbf{X};n}(\mathbf{x}; h) \neq 0, \\ \frac{1}{n} \sum_{i=1}^{n} \psi(\mathbf{Y}_i) & \text{for} \quad f_{\mathbf{X};n}(\mathbf{x}; h) = 0. \end{cases} \tag{11}$$

The above Akaike–Parzen–Rosenblatt kernel estimator $f_{\mathbf{X};n}(\mathbf{x}; h)$ of $f_{\mathbf{X}}(\mathbf{x})$ ([1, 18, 20]), and Nadaraya–Watson estimator $m_{\psi;n}(\mathbf{x}; h)$ of the regression function $m_{\psi}(\mathbf{x})$ ([16, 22]), have been extensively studied in the literature (see, e. g., Deheuvels and Mason [5], and the references therein). Introduce the non-random centering func-

tions

$$\mathbb{E}f_{\mathbf{X};n}(\mathbf{x};h) = \frac{1}{h}\,\mathbb{E}\left\{K\left(\frac{\mathbf{x}-\mathbf{X}}{h^{1/p}}\right)\right\}, \tag{12}$$

$$\mathbb{E}r_{\psi;n}(\mathbf{x};h) = \frac{1}{h}\,\mathbb{E}\left\{\psi(\mathbf{Y})K\left(\frac{\mathbf{x}-\mathbf{X}}{h^{1/p}}\right)\right\}, \tag{13}$$

$$\widehat{\mathbb{E}}m_{\psi;n}(\mathbf{x};h) = \begin{cases} \dfrac{\mathbb{E}r_{\psi;n}(\mathbf{x};h)}{\mathbb{E}f_{\mathbf{X};n}(\mathbf{x};h)} & \text{for} \quad \mathbb{E}f_{\mathbf{X};n}(\mathbf{x};h) \neq 0, \\ \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\psi(\mathbf{Y}_i) & \text{for} \quad \mathbb{E}f_{\mathbf{X};n}(\mathbf{x};h) = 0. \end{cases} \tag{14}$$

Under (F.1–2–3) and (K.1–2) (see, e. g., p. 1383 in [8]), for all $\mathbf{x} \in \mathbf{J}$, as $h \to 0$,

$$\mathbb{E}f_{\mathbf{X};n}(\mathbf{x};h) \to f_{\mathbf{X}}(\mathbf{x}), \quad \mathbb{E}r_{\psi;n}(\mathbf{x};h) \to r_{\psi}(\mathbf{x}), \tag{15}$$

and

$$\widehat{\mathbb{E}}m_{\psi;n}(\mathbf{x};h) \to m_{\psi}(\mathbf{x}). \tag{16}$$

The rate of convergence in (15)–(16) is a purely analytical problem which we will not consider here. We will concentrate on the study of the random sup-norm deviations

$$\mathcal{D}_n(\mathbf{I};h) := \sup_{\mathbf{x}\in\mathbf{I}}\left|f_{\mathbf{X};n}(\mathbf{x};h) - \mathbb{E}f_{\mathbf{X};n}(\mathbf{x};h)\right|, \tag{17}$$

$$\mathcal{D}_{r;\psi;n}(\mathbf{I};h) := \sup_{\mathbf{x}\in\mathbf{I}}\left|r_{\psi;n}(\mathbf{x};h) - \mathbb{E}r_{\psi;n}(\mathbf{x};h)\right|, \tag{18}$$

$$\mathcal{D}_{m;\psi;n}(\mathbf{I};h) := \sup_{\mathbf{x}\in\mathbf{I}}\left|m_{\psi;n}(\mathbf{x};h) - \widehat{\mathbb{E}}m_{\psi;n}(\mathbf{x};h)\right|. \tag{19}$$

Theorem 1.2 below is due to Deheuvels and Mason [5] for $p = q = 1$. The extension of this result to arbitrary $p \geq 1$ and $q \geq 1$ will be discussed in the forthcoming § 3. Introduce the following assumptions on the limiting behavior of $h = h_n$ as $n \to \infty$.

(H.1) $h_n \to 0$ as $n \to \infty$;

(H.2) $nh_n/\log n \to \infty$ as $n \to \infty$;

(H.3) $nh_n\left\{\frac{\log n}{\mathcal{M}^{\mathrm{inv}}(n)^2}\right\} \to \infty$ as $n \to \infty$.

Theorem 1.2. Under (F.1–2–3), (K.1–2) and (H.1–2–3), we have, as $n \to \infty$

$$\mathcal{D}_n(\mathbf{I};h_n) = (1+o_{\mathbb{P}}(1)) \times \left\{\frac{2\log(1/h_n)}{nh_n}\right\}^{1/2}\left\{\sup_{\mathbf{x}\in\mathbf{I}} f_{\mathbf{X}}(\mathbf{x})\int_{\mathbb{R}^p} K(\mathbf{t})^2\mathrm{d}\mathbf{t}\right\}^{1/2}, \tag{20}$$

$$\mathcal{D}_{r;\psi;n}(\mathbf{I};h_n) = (1+o_{\mathbb{P}}(1)) \times \left\{\frac{2\log(1/h_n)}{nh_n}\right\}^{1/2}\left\{\sup_{\mathbf{x}\in\mathbf{I}} f_{\mathbf{X}}(\mathbf{x})\sigma_{\psi}^2(\mathbf{x})\int_{\mathbb{R}^p} K(\mathbf{t})^2\mathrm{d}\mathbf{t}\right\}^{1/2}, \tag{21}$$

$$\mathcal{D}_{m;\psi;n}(\mathbf{I};h_n) = (1+o_{\mathbb{P}}(1)) \times \left\{\frac{2\log(1/h_n)}{nh_n}\right\}^{1/2}\left\{\sup_{\mathbf{x}\in\mathbf{I}} \frac{\sigma_{\psi}^2(\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})}\int_{\mathbb{R}^p} K(\mathbf{t})^2\mathrm{d}\mathbf{t}\right\}^{1/2}. \tag{22}$$

The right-hand sides of (20), (21) and (22) are difficult to evaluate in practice. Not only they depend upon unknown distributional factors, but also upon the scale-dependent coefficient $\log(1/h_n)$. This motivates the resampling approach to this problem presented in § 2.

2. RESAMPLED ESTIMATORS

Bootstrap and resampling methods have been used for decades in nonparametric functional estimation, starting with Efron [7], Härdle and Bowman [13], Härdle and Marron [12], Hall [11], and Li and Datta [14] (see, e. g., Claeskens and van Keilegom [3]). Deheuvels and Derzko [4] have introduced a bootstrap methodology to override the difficulty of evaluating $\mathcal{D}_n(\mathbf{I}; h_n)$. We will extend their results to $\mathcal{D}_{r;\psi;n}(\mathbf{I}; h_n)$ and $\mathcal{D}_{m;\psi;n}(\mathbf{I}; h_n)$. We define random weights $\{W_{i,n} : 1 \leq i \leq n\}$, independent of $\{(\mathbf{X}_i, \mathbf{Y}_i) : 1 \leq i \leq n\}$, via either one of the resampling schemes $(R.1)$ or $(R.2)$ below.

$(R.1)$ As in the Mason and Newton [15] bootstrap, we introduce an auxiliary iid sequence $Z = Z_1, Z_2, \ldots$ of real-valued rv's, independent of $\{(\mathbf{X}_i, \mathbf{Y}_i) : i \geq 1\}$, and such that

$(A.1)$ $\quad \mathbb{E}(Z) = 1; \quad \mathbb{E}(Z^2) = 2;$
$(A.2)$ $\quad$ For some $\epsilon > 0$, $\mathbb{E}(e^{tZ}) < \infty$ for all $|t| \leq \epsilon$.

Setting $T_n = Z_1 + \ldots + Z_n$ we define the $\{W_{i,n} : 1 \leq i \leq n\}$, by setting, for $i = 1, \ldots, n$,

$$W_{i,n} = \begin{cases} \dfrac{Z_i}{T_n} = \dfrac{Z_i}{\sum_{j=1}^n Z_j} & \text{when} \quad T_n > 0, \\ \dfrac{1}{n} & \text{when} \quad T_n \leq 0. \end{cases}$$

$(R.2)$ As in the Efron [7] multinomial bootstrap, we let $(Z_1, \ldots, Z_n)$ denote a multinomial $\mathrm{Mult}(\frac{1}{n}, \ldots, \frac{1}{n}; n)$ rv, independent of $\{(\mathbf{X}_i, \mathbf{Y}_i) : i \geq 1\}$, and such that, for each n-uple $\{k_1, \ldots, k_n\}$ of nonnegative integers fulfilling $k_1 + \ldots + k_n = n$,

$$\mathbb{P}(Z_1 = k_1, \ldots, Z_n = k_n) = \frac{n^{-n} n!}{k_1! \ldots k_n!}.$$

We then define the $\{W_{i,n} : 1 \leq i \leq n\}$, by setting

$$W_{i,n} = \frac{Z_i}{n} = \frac{Z_i}{\sum_{j=1}^n Z_j} \quad \text{for} \quad i = 1, \ldots, n.$$

Under either $(R.1)$ or $(R.2)$, we define *resampled* or *bootstrapped* versions of $f_{\mathbf{X},n}(\mathbf{x}; h)$,

$r_{\psi;n}(\mathbf{x};h)$ and $m_{\psi;n}(\mathbf{x};h)$, by setting, for each $h>0$ and $\mathbf{x}\in\mathbb{R}^p$,

$$f^*_{\mathbf{X};n}(\mathbf{x};h) = \frac{1}{h}\sum_{i=1}^n W_{i,n}K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h^{1/p}}\right), \tag{23}$$

$$r^*_{\psi;n}(\mathbf{x};h) = \frac{1}{h}\sum_{i=1}^n W_{i,n}\psi(\mathbf{Y}_i)K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h^{1/p}}\right), \tag{24}$$

$$m^*_{\psi;n}(\mathbf{x};h) = \begin{cases} \dfrac{r^*_{\psi;n}(\mathbf{x};h)}{f^*_{\mathbf{X};n}(\mathbf{x};h)} & \text{for} \quad f^*_{\mathbf{X};n}(\mathbf{x};h)\neq 0, \\ \sum_{i=1}^n W_{i,n}\psi(\mathbf{Y}_i) & \text{for} \quad f^*_{\mathbf{X};n}(\mathbf{x};h)=0. \end{cases} \tag{25}$$

We next consider the resampled sup-norm deviations

$$\mathcal{D}^*_n(\mathbf{I};h) := \sup_{\mathbf{x}\in\mathbf{I}}\left|f^*_{\mathbf{X};n}(\mathbf{x};h)-f_{\mathbf{X};n}(\mathbf{x};h)\right|, \tag{26}$$

$$\mathcal{D}^*_{r;\psi;n}(\mathbf{I};h) := \sup_{\mathbf{x}\in\mathbf{I}}\left|r^*_{\psi;n}(\mathbf{x};h)-r_{\psi;n}(\mathbf{x};h)\right|, \tag{27}$$

$$\mathcal{D}^*_{m;\psi;n}(\mathbf{I};h) := \sup_{\mathbf{x}\in\mathbf{I}}\left|m^*_{\psi;n}(\mathbf{x};h)-m_{\psi;n}(\mathbf{x};h)\right|. \tag{28}$$

The following result extends Theorem 13.1.1 of Deheuvels and Derzko [4] (which yields the limiting statement (29) when $p=q=1$).

Theorem 2.1. Under $(F.1\text{–}2\text{–}3)$, $(K.1\text{–}2)$, $(H.1\text{–}2\text{–}3)$, and, either $(R.1)$ or $(R.2)$, we have, as $n\to\infty$

$$\mathcal{D}^*_n(\mathbf{I};h_n) = (1+o_{\mathbb{P}}(1)) \times \left\{\frac{2\log(1/h_n)}{nh_n}\right\}^{1/2}\left\{\sup_{\mathbf{x}\in\mathbf{I}} f_{\mathbf{X}}(\mathbf{x})\int_{\mathbb{R}^p}K(\mathbf{t})^2\mathrm{d}\mathbf{t}\right\}^{1/2}, \tag{29}$$

$$\mathcal{D}^*_{r;\psi;n}(\mathbf{I};h_n) = (1+o_{\mathbb{P}}(1)) \times \left\{\frac{2\log(1/h_n)}{nh_n}\right\}^{1/2}\left\{\sup_{\mathbf{x}\in\mathbf{I}} f_{\mathbf{X}}(\mathbf{x})\sigma^2_\psi(\mathbf{x})\int_{\mathbb{R}^p}K(\mathbf{t})^2\mathrm{d}\mathbf{t}\right\}^{1/2}, \tag{30}$$

$$\mathcal{D}^*_{m;\psi;n}(\mathbf{I};h_n) = (1+o_{\mathbb{P}}(1)) \times \left\{\frac{2\log(1/h_n)}{nh_n}\right\}^{1/2}\left\{\sup_{\mathbf{x}\in\mathbf{I}} \frac{\sigma^2_\psi(\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})}\int_{\mathbb{R}^p}K(\mathbf{t})^2\mathrm{d}\mathbf{t}\right\}^{1/2}. \tag{31}$$

A direct consequence of Theorems 1.2 and 2.1 is stated in the following corollary.

Corollary 2.2. Under the assumptions of Theorem 2.1, we have, as $n\to\infty$

$$\mathcal{D}_n(\mathbf{I};h_n) = (1+o_{\mathbb{P}}(1))\mathcal{D}^*_n(\mathbf{I};h_n), \tag{32}$$

$$\mathcal{D}_{r;\psi;n}(\mathbf{I};h_n) = (1+o_{\mathbb{P}}(1))\mathcal{D}^*_{r;\psi;n}(\mathbf{I};h_n), \tag{33}$$

$$\mathcal{D}_{m;\psi;n}(\mathbf{I};h_n) = (1+o_{\mathbb{P}}(1))\mathcal{D}^*_{m;\psi;n}(\mathbf{I};h_n). \tag{34}$$

The application of Corollary 2.2 is pretty much obvious. By performing just a single resampling (or bootstrap), one evaluates $\mathcal{D}_n^*(\mathbf{I}; h_n)$ (resp. $\mathcal{D}_{r;\psi;n}^*(\mathbf{I}; h_n)$ or $\mathcal{D}_{m;\psi;n}^*(\mathbf{I}; h_n)$) when one seeks to estimate $f_{\mathbf{X}}(\cdot)$) (resp. $r_\psi(\cdot)$ or $m_\psi(\cdot)$). Making use of either (32), (33), or (34), we so obtain sharp *asymptotic certainty bands*, for the unknown quantities $\mathbb{E}f_{\mathbf{X};n}(\mathbf{x}; h)$, $\mathbb{E}r_{\psi;n}(\mathbf{x}; h)$, or $\widehat{\mathbb{E}}m_{\psi;n}(\mathbf{x}; h)$. This notion will be discussed below, in the setup of density estimation. The applications of this methodology are similar in the case of Nadaraya–Watson regression estimation with the formal change of (32) by (34).

By a *uniform asymptotic certainty band* on $\mathbf{I}$ for $\mathbb{E}f_{\mathbf{X};n}(\mathbf{x}; h)$, is meant a (possibly non-random) statistic $\theta_n = \theta_n(X_1, \ldots, X_n) \geq 0$, such that, for each $0 < \varepsilon < 1$, as $n \to \infty$,

$$P\Big(\sup_{x \in I} |f_{\mathbf{X};n}(\mathbf{x}; h) - \mathbb{E}f_{\mathbf{X};n}(\mathbf{x}; h)| \leq \theta_n(1 + \varepsilon) + \varepsilon \Big) \to 1, \tag{35}$$

and

$$P\Big(\sup_{x \in I} |f_{\mathbf{X};n}(\mathbf{x}; h) - \mathbb{E}f_{\mathbf{X};n}(\mathbf{x}; h)| \leq \theta_n(1 - \varepsilon) - \varepsilon \Big) \to 0. \tag{36}$$

Remark 2.3. The qualification of *certainty bands* is used here, rather than that of *confidence bands*, since the above θ_n does not depend on a confidence level $\alpha \in (0, 1)$. Some authors (see, e. g., Härdle and Marron [12]) have used the concept of *simultaneous error bars*, which is not quite equivalent to the present definitions. Our approach also differs from that of Bickel and Rosenblatt [2] (see, e. g., Giné, Koltchinskii and Sakhanenko [9]), who evaluate limiting distributions for weighted sup-norms of $|f_{\mathbf{X};n}(\mathbf{x}; h) - \mathbb{E}f_{\mathbf{X};n}(\mathbf{x}; h)|$. The convergence of such statistics to their limit laws is, unfortunately, so slow as to render their application problematic for "ordinary" sample sizes. The methodology we follow should appear, therefore, less refined, but more applicable to the "usual" statistical data sets.

The practical interest of asymptotic certainty bands appears through the plots of

$$(f_{\mathbf{X};n}(\mathbf{x}; h) - \theta_n) \vee 0 \quad \text{and} \quad (f_{\mathbf{X};n}(\mathbf{x}; h) + \theta_n) \vee 0,$$

over $\mathbf{x} \in \mathbf{I}$. In view of Corollary 2.2, we recommend the choice of $\theta_n = \mathcal{D}_n^*(\mathbf{I})$, which fulfills (35)–(36). These limiting statements show that the centering factor $\mathbb{E}f_{\mathbf{X};n}(\mathbf{x}; h)$ should, with probability tending to 1, be, for all $\mathbf{x} \in \mathbf{I}$ in between or close to (up to a factor tending to 1, in the sense of (35)–(36)) the *lower* (resp. *upper*) *certainty bound*

$$(f_{\mathbf{X};n}(\mathbf{x}; h) - \theta_n) \vee 0, \quad (\text{resp.} \quad (f_{\mathbf{X};n}(\mathbf{x}; h) + \theta_n) \vee 0).$$

In view of (15), we would like to replace $\mathbb{E}f_{\mathbf{X};n}(\mathbf{x}; h)$ by $f_{\mathbf{X}}(\mathbf{x})$ in the above statements. This, however, would require the non-random bias $f_{\mathbf{X}}(\mathbf{x}) - \mathbb{E}f_{\mathbf{X};n}(\mathbf{x}; h)$ to be negligible with respect to θ_n. As mentioned earlier, this condition imposes additional rate conditions on $h = h_n$ and regularity assumptions on $f_{\mathbf{X}}(\mathbf{x})$ (see, e. g. [5]). The same observation holds for the other kernel functional estimators considered above.

3. PROOFS

3.1. The general setup

When $p = q = 1$, Theorem 1.2 reduces to Theorem 1.1 of Deheuvels and Mason [5]. The extension of this result to arbitrary $p \geq 1$ and $q \geq 1$ is mostly a matter of book-keeping, with minor changes. The proof relies on the study of the oscillations of the process

$$W_{\lambda,n}(\mathbf{x},\psi) := \sum_{i=1}^{n} \big(c(\mathbf{x})\psi(\mathbf{Y}_i) + d(\mathbf{x})\big) K\left(\frac{\mathbf{x}-\mathbf{X}_i}{\lambda h_n^{1/p}}\right) - n\mathbb{E}\left\{\big(c(\mathbf{x})\psi(\mathbf{Y}) + d(\mathbf{x})\big) K\left(\frac{\mathbf{x}-\mathbf{X}}{\lambda h_n^{1/p}}\right)\right\}, \tag{37}$$

where $c(\mathbf{x})$ and $d(\mathbf{x})$ are continuous functions of $\mathbf{x} \in \mathbf{J}$. Theorem 1.1 of [5] turns out to follow from the limit law stated in (38)–(39) below (see, e. g., Theorem 3.1 of [5]). Under the assumptions of the theorem, for any $0 < \lambda_1 \leq \lambda_2 < \infty$, as $n \to \infty$

$$\sup_{\lambda_1 \leq \lambda \leq \lambda_2} \left| \{2h_n \log(1/h_n)\}^{-1/2} \sup_{\mathbf{x}\in\mathbf{I}} \Big\{ \pm W_{\lambda,n}(\mathbf{x},\psi)\Big\} - \sigma(\psi)\right| = o_{\mathbb{P}}(1), \tag{38}$$

where

$$\sigma^2(\psi) = \sup_{\mathbf{x}\in\mathbf{I}} \Big\{ \mathbb{E}\Big(\big(c(\mathbf{x})\psi(\mathbf{Y}) + d(\mathbf{x})\big)^2 \big| \mathbf{X} = \mathbf{x}\Big) f_{\mathbf{X}}(\mathbf{x})\Big\} \int_{\mathbb{R}^p} K^2(\mathbf{t})\,\mathrm{d}\mathbf{t}. \tag{39}$$

To establish (38)–(39) for arbitrary $p \geq 1$ and $q \geq 1$, we rely on the technical arguments in §3 of Einmahl and Mason [8], where the properties of the process (37) are investigated for $\mathbf{x} \in \mathbb{R}^p$ and $\mathbf{y} \in \mathbb{R}^q$. Our assumptions on $K(\cdot)$ are slightly different from that of [8], and this point is discussed in § 3.2 below. Given these remarks, the remaining details of the proof are readily derived from §4 of [5].

We now turn to Theorem 2.1. We first consider the resampling scheme (R.1), and introduce the resampled version of the process (37) given by

$$W^*_{\lambda,n}(\mathbf{x},\psi) := \sum_{i=1}^{n} \big(c(\mathbf{x})Z_i\psi(\mathbf{Y}_i) + d(\mathbf{x})\big) K\left(\frac{\mathbf{x}-\mathbf{X}_i}{\lambda h_n^{1/p}}\right) - n\mathbb{E}\left\{\big(c(\mathbf{x})Z\psi(\mathbf{Y}) + d(\mathbf{x})\big) K\left(\frac{\mathbf{x}-\mathbf{X}}{\lambda h_n^{1/p}}\right)\right\}.$$

We observe that $W^*_{\lambda,n}(\mathbf{x},\psi)$ reduces to a process of the form $W_{\lambda,n}(\mathbf{x},\psi^*)$, for a suitable measurable ψ^*, and after some easy changes. Without loss of generality, we set $Z = Q(U)$ and $Z_i = Q(U_i)$ for $i = 1,\ldots,n$, where U and $U_1,\ldots,U_n$ are independent rv's, with a uniform distribution on $(0,1)$, and independent of $\{(\mathbf{X}_i,\mathbf{Y}_i) : 1 \leq i \leq n\}$. This allows us to define a measurable function $\psi^*(\cdot)$ on $\mathbb{R}^{q+1}$, and a rv $\mathbf{Y}^*$, by

$$\mathbf{Y}^* = \begin{bmatrix} U & Y_1 & \ldots & Y_q \end{bmatrix}' \in \mathbb{R}^{q+1} \quad \text{and} \quad \psi^*(\mathbf{Y}^*) = Q(U)\psi(\mathbf{Y}) = Z\psi(\mathbf{Y}).$$

Letting $(\mathbf{X}_i,\mathbf{Y}^*_i)$, $i = 1,2,\ldots$, denote iid random copies of $(\mathbf{X},\mathbf{Y}^*)$, it is readily checked that $\{(\mathbf{X}_i,\mathbf{Y}^*_i) : 1 \leq i \leq n\}$ and $\psi^*(\cdot)$ fulfill the general assumptions imposed

upon $\{(\mathbf{X}_i, \mathbf{Y}_i) : 1 \leq i \leq n\}$ and $\psi(\cdot)$ in Theorem 1.2. We may therefore write the version of (38)–(39) holding in this case. Namely, for each $0 < \lambda_1 \leq \lambda_2 < \infty$, as $n \to \infty$

$$\sup_{\lambda_1 \leq \lambda \leq \lambda_2} \left| \{2h_n \log(1/h_n)\}^{-1/2} \sup_{\mathbf{x} \in \mathbf{I}} \left\{ \pm W^*_{\lambda,n}(\mathbf{x}, \psi) - \sigma_*(\psi) \right\} \right| = o_{\mathbb{P}}(1), \quad (40)$$

where

$$\sigma^2_*(\psi) = \sup_{\mathbf{x} \in \mathbf{I}} \left\{ \mathbb{E}\Big(\big(c(\mathbf{x}) Z\psi(\mathbf{Y}) + d(\mathbf{x})\big)^2 \big| \mathbf{X} = \mathbf{x} \Big) f_{\mathbf{X}}(\mathbf{x}) \right\} \int_{\mathbb{R}^p} K^2(\mathbf{t})\, \mathrm{d}\mathbf{t}. \quad (41)$$

By recopying the arguments of [5], we infer readily (29)–(31) from (40)–(41), which completes the proof of Theorem 2.1 under $(R.1)$. As in [4], a Poisson approximation argument takes care of the $(R.2)$ resampling scheme. The corresponding details will be given elsewhere.

3.2. Technical conditions on the kernel

Following Einmahl and Mason [8], we set $\mathcal{K} = \{K((x-\cdot)/h^{1/p}) : h > 0, x \in \mathbb{R}^p\}$. One of the major ingredients in our proofs relies on the assessment that $\mathcal{K}$ constitutes an appropriate *VC-class with integrable envelope function* (see, e. g., Appendix A in [5]). In particular, we require $\mathcal{K}$ to fulfill the entropy condition (42) below. As on p. 1381 of [8], for each $\varepsilon > 0$, we define $N(\varepsilon, \mathcal{K})$ as the supremum, over all Borel probability measures Q on $\mathbb{R}^p$, of the minimal number of balls of d_Q-radius ε necessary to cover $\mathcal{K}$, where d_Q denotes the $L^2(Q)$ metric induced by Q. We need to assume (see, e. g., (K.iii), p. 1383 of [8]) the existence of a $C > 0$ and a $\nu > 0$, such that, for all $0 < \varepsilon < 1$,

$$N(\varepsilon, \mathcal{K}) \leq C\varepsilon^{-\nu}. \quad (42)$$

To establish (42) requires analytical arguments which are avoided in most of the papers dealing with these methods.The appropriate details are to be found in §4.7 of Dudley [6], Examples 26 and 38, and Exercise 29 in Pollard [19], Lemma 22 in Nolan and Pollard [17], Definition 2.3.3, Example 2.3.5, Theorem 2.5.2 in van der Vaart and Wellner [21], and the Appendix A in Deheuvels and Mason [5]. As mentioned in the proof of Lemma 22 in Nolan and Pollard, (42) holds whenever $K(\cdot)$ is a linear combination of two multidimensional df's, this condition being implied by $(K.2)(i)$. An example is obtained when $K(\cdot)$ is a *product kernel*, of the form

$$K(\mathbf{t}) = \prod_{j=1}^{p} K_j(t_j),$$

with $K_j(\cdot)$ being, for each $j = 1, \ldots, p$, a function of bounded variation on $\mathbb{R}$. Another example is given p. 1381 of Einmahl and Mason [8], and on pp. 1111–1112 of Giné and Mason [10], who show that $K(\mathbf{t}) = \Psi(\mathcal{P}(\mathbf{t}))$ fulfills (42) when $\Psi(\cdot)$ is a function of bounded variation on $\mathbb{R}$, and $\mathcal{P}(\cdot)$ is either a real polynomial, or the αth power of the absolute value of a real polynomial for some $\alpha > 0$. Our theorems remain valid when $(K.2)(i)$ is replaced by this last assumption, or more generally, when $\mathcal{K}$ fulfills (42).

A second condition to be imposed upon $\mathcal{K}$ is that this set must be a *pointwise measurable class* (see, e. g., Appendix A in Deheuvels and Mason [5] and Example 2.3.4 in van der Vaart and Wellner [21]), meaning that there exists a countable $\mathcal{K}_0 \subseteq \mathcal{K}$, such that, for each $g \in \mathcal{K}$, there exists a sequence $\{g_n\} \subseteq \mathcal{K}_0$, with

$$g_n(\mathbf{z}) \to g(\mathbf{z}) \quad \text{for} \quad \mathbf{z} \in \mathbb{R}^d.$$

This condition is fulfilled, under $(K.2)(i\text{-}ii)$ (when $K(\cdot)$ is right-continuous, see, e. g., p. 1382 in Einmahl and Mason [8]). Finally, the integrable envelope function is provided for $\mathcal{K}$ by $(K.2)(ii)$, which requires the kernel $K(\cdot)$ to be uniformly bounded and compactly supported.

Remark 3.1. In view of the preceding arguments, it is quite obvious that our theorems may be written in the setup of *uniform in bandwidth* kernel estimation. This will be considered elsewhere.

ACKNOWLEDGEMENT

We thank the referees for their comments and suggestions.

(Received November 15, 2010)

REFERENCES

[1] H. Akaike: An approximation of the density function. Ann. Inst. Statist. Math. *6* (1954), 127–132.

[2] P. J. Bickel and M. Rosenblatt: On some global measures of the deviations of density functions estimates. Ann. Statist. *1* (1973), 1071–1095.

[3] G. Claeskens and I. van Keilegom: Bootstrap confidence bands for regression curves and their derivatives. Ann. Statist. *31* (2003), 1852–1884.

[4] P. Deheuvels and G. Derzko: Asymptotic certainty bands for kernel density estimators based upon a bootstrap resampling scheme. In: Statistical Models and Methods for Biomedical and Technical systems, Stat. Ind. Technol., Birkhäuser, Boston MA 2008, pp. 171–186.

[5] P. Deheuvels and D. M. Mason: General asymptotic confidence bands based on kernel-type function estimators. Statist. Infer. Stoch. Processes *7* (2004), 225–277.

[6] R. M. Dudley: Uniform Central Limit Theorems. Cambridge University Press, Cambridge 1999.

[7] B. Efron: Bootstrap methods: Another look at the jackknife. Ann. Statist. *7* (1979), 1–26.

[8] U. Einmahl and D. M. Mason: Uniform in bandwidth consistency of kernel-type function estimators. Ann. Statist. *33* (2005), 1380–1403.

[9] E. Giné, V. Koltchinskii, and L. Sakhanenko: Kernel density estimators: convergence in distribution for weighted sup-norms. Probab. Theory Related Fields *130* (2004), 167–198.

[10] E. Giné and D. M. Mason: On local U-statistic processes and the estimation of densities of functions of several sample variables. Ann. Statist. *35* (2007), 1105–1145.

[11] P. Hall: The Bootstrap and Edgeworth Expansion. Springer, New York 1992.

[12] W. Härdle and J. S. Marron: Bootstrap simultaneous error bars for nonparametric regression. Ann. Statist. *19* (1991), 778–796.

[13] W. Härdle and A. W. Bowman: Bootstrapping in nonparametric regression: Local adaptative smoothing and confidence bands. J. Amer. Statist. Assoc. *83* (1988), 102–110.

[14] G. Li and S. Datta: A bootstrap approach to nonparametric regression for right censored data. Ann. Inst. Statist. Math. *53* (2001), 708–729.

[15] D. M. Mason and M. A. Newton: A rank statistics approach to the consistency of a general bootstrap. Ann. Statist. *20* (1992), 1611–1624.

[16] E. A. Nadaraya: On estimating regression. Theor. Probab. Appl. *9* (1964), 141–142.

[17] D. Nolan and D. Pollard: U-processes: Rates of convergence. Ann. Statist. *15* (1987), 780–799.

[18] E. Parzen: On the estimation of probability density and mode. Ann. Math. Statist. *33* (1962), 1065–1076.

[19] D. Pollard: Convergence of Stochastic Processes. Springer, New York 1984.

[20] M. Rosenblatt: Remarks on some nonparametric estimates of a density function. Ann. Math. Statist. *27* (1956), 832–837.

[21] A. W. van der Vaart and J. A. Wellner: Weak Convergence and Empirical Processes with Aplications to Statistics. Springer, New York 1996.

[22] G. S. Watson: Smooth regression analysis. Sankhyā A *26* (1964), 359–372.

Paul Deheuvels, L.S.T.A., Université Pierre et Marie Curie (Paris 6), 7 avenue du Château, 92340 Bourg-la-Reine. France

e-mail: paul.deheuvels@upmc.fr