Ann. I. H. Poincaré – PR 38, 6 (2002) 907–921

S0246-0203(02)01128-7/FLA

RATES OF STRONG UNIFORM CONSISTENCY FOR MULTIVARIATE KERNEL DENSITY ESTIMATORS

VITESSE DE CONVERGENCE UNIFORME PRESQUE SÛRE POUR DES ESTIMATEURS À NOYAUX DE DENSITÉS MULTIVARIÉES

Evarist GINÉ [a,1], **Armelle GUILLOU** [b]

[a] *Departments of Mathematics and Statistics, University of Connecticut, Storrs, CT 06269, USA*
[b] *Université Paris VI, L.S.T.A., tour 45-55 E3, 4 place Jussieu, 75252 Paris cedex 05, France*

Received October 2000, revised December 2001

ABSTRACT. – Let f_n denote the usual kernel density estimator in several dimensions. It is shown that if $\{a_n\}$ is a regular band sequence, K is a bounded square integrable kernel of several variables, satisfying some additional mild conditions ((K_1) below), and if the data consist of an i.i.d. sample from a distribution possessing a bounded density f with respect to Lebesgue measure on $\mathbf{R}^d$, then

$$\limsup_{n\to\infty} \sqrt{\frac{na_n^d}{\log a_n^{-1}}} \sup_{t\in\mathbf{R}^d} |f_n(t) - Ef_n(t)| \leqslant C\sqrt{\|f\|_\infty \int K^2(x)\,\mathrm{d}x} \quad \text{a.s.}$$

for some absolute constant C that depends only on d. With some additional but still weak conditions, it is proved that the above sequence of normalized suprema converges a.s. to $\sqrt{2d\|f\|_\infty \int K^2(x)\,\mathrm{d}x}$. Convergence of the moment generating functions is also proved. Neither of these results require f to be strictly positive. These results improve upon, and extend to several dimensions, results by Silverman [13] for univariate densities.

MSC: primary 62G07, 62G20, 60F15; secondary 62G30

Keywords: Uniform almost sure rates; Non-parametric density estimation; Kernel density estimators

RÉSUMÉ. – Soit f_n l'estimateur à noyau d'une densité multivariée. Nous démontrons dans cet article que si $\{a_n\}$ est une suite régulière, K un noyau multivarié borné de carré intégrable,

E-mail addresses: gine@uconnvm.uconn.edu (E. Giné), guillou@ccr.jussieu.fr (A. Guillou).
[1] Research supported in part by NSF Grant No. DMS-0070382.

satisfaisant quelques faibles conditions additionnelles ((K_1) ci-dessous), et si les données sont i.i.d. de densité f bornée relativement à la mesure de Lebesgue sur $\mathbf{R}^d$, alors

$$\limsup_{n\to\infty} \sqrt{\frac{na_n^d}{\log a_n^{-1}}} \sup_{t\in\mathbf{R}^d} |f_n(t) - Ef_n(t)| \leqslant C\sqrt{\|f\|_\infty \int K^2(x)\,\mathrm{d}x} \quad \text{p.s.}$$

où C est une constante dépendant uniquement de d. Sous de faibles hypothèses supplémentaires, nous démontrons que le suprémum normalisé ci-dessus converge p.s. vers $\sqrt{2d\|f\|_\infty \int K^2(x)\,\mathrm{d}x}$. Nous démontrons aussi la convergence des fonctions génératrices des moments. Aucun des résultats précédents ne requiert la stricte positivité de f. Ces résultats améliorent et étendent au cas multivarié les résultats de Silverman [13] pour des densités univariées.

1. Introduction

Let f be a probability density with respect to Lebesgue measure on $\mathbf{R}^d$, let X, X_i, $i \in \mathbf{N}$, be independent identically distributed $\mathbf{R}^d$-valued random variables with density f, and let K be a bounded square integrable kernel (a measurable function on $\mathbf{R}^d$). Let $a_n \searrow 0$, $na_n \to \infty$. The kernel density estimators of f based on the observations X_i, with kernel K and bandwidths $\{a_n\}$, are defined as

$$f_n(t) = \frac{1}{na_n^d} \sum_{i=1}^n K\left(\frac{t - X_i}{a_n}\right) \tag{1.1}$$

for all $n \in \mathbf{N}$ (Rosenblatt [12]). Stute [14], Theorem 3.1, obtained the exact rate at which the deviation of f_n with respect to its mean, weighted by its standard deviation, tends to zero uniformly over compact parallellepipeds. The object of this note is to complement Stute's already classical result by obtaining the exact rate of a.s. convergence to zero of the supremum over all of $\mathbf{R}^d$ of the deviation of f_n with respect to its mean, that is, of

$$\|f_n - \bar{f}_n\|_\infty := \sup_{t\in\mathbf{R}^d} \bigl|f_n(t) - \bar{f}_n(t)\bigr|, \tag{1.2}$$

where

$$\bar{f}_n(t) = Ef_n(t) = \frac{1}{a_n^d} \int_{\mathbf{R}}^{d} K\left(\frac{t-x}{a_n}\right) f(x)\,\mathrm{d}x, \quad n \in \mathbf{N}. \tag{1.3}$$

We will see that when the sup is over the whole space we cannot divide by the standard deviation (which is proportional to $\sqrt{f}$) but, on the other hand, f is not required to be non-zero. In the case $d = 1$, Silverman [13] obtained an approximate rate under assumptions on f, K and the bandwidths that are more restrictive than the assumptions we will impose.

Our first result will only be approximate: it consists of an upper bound for (1.2), exact only up to a multiplicative constant. Its interest rests upon the facts that the assumptions on K and f are much weaker than is usual, that the interval of uniformity of the bound consists of all of $\mathbf{R}^d$, and that it will be part of the proof of a more exact result. In fact our second result gives the exact rate in (1.2) under slightly stronger assumptions on K and f.

The first result, which has an extremely simple proof, is based on direct application of an exponential bound for empirical processes indexed by VC classes of functions from Giné and Guillou [7] which is just a reformulation of results of Talagrand [15,16] in a form suitable for our purposes. Einmahl and Mason [6] use a similar inequality.

We obtain the second and main result, Theorem 3.3 below, which is asymptotically exact, by combining the first one with a result that can be inferred with little effort from the proof of the theorem in Einmahl and Mason [6]. We complement the a.s. convergence in Theorem 3.3 with moment bounds and with convergence of the moment generating functions.

After the present article had been completed and circulated, we learned from P. Deheuvels that he had also recently obtained Theorem 3.3 in the particular case of $d = 1$ (Deheuvels [2], part of Theorem 3). Our result, which extends his to several dimensions, was obtained independently and the proofs are different.

2. The general upper bound

We begin by describing the single most important ingredient in the proofs that follow, which is Talagrand's [15,16] remarkable exponential inequality for general empirical processes, complemented by a moment inequality for empirical processes indexed by classes of functions of Vapnik–Červonenkis type (Talagrand [15] for classes of sets, Giné and Guillou [7] for classes of functions). Let $(S, \mathcal{S})$ be a measurable space and let $\mathcal{F}$ be a uniformly bounded collection of measurable functions on it. We say that $\mathcal{F}$ is a bounded measurable *VC* class of functions if the class $\mathcal{F}$ is separable or is image admissible Suslin (Dudley [4, Section 5.3]) and if there exist positive numbers A and v such that, for every probability measure P on $(S, \mathcal{S})$ and every $0 < \tau < 1$,

$$N\big(\mathcal{F}, L_2(P), \tau\|F\|_{L_2(P)}\big) \leqslant \left(\frac{A}{\tau}\right)^v, \tag{2.1}$$

where $N(T, d, \tau)$ denotes the τ-covering number of the metric space (T, d), that is, the smallest number of balls of radius not larger than τ and centers in T needed to cover T. In the above inequality, d is the $L_2(P)$ distance. We will refer to numbers A and v for which the inequality holds for all P as a set of *VC* characteristics of the class $\mathcal{F}$, and we will assume in what follows, without further mention, that $A \geqslant 3\sqrt{\mathrm{e}}$ and $v \geqslant 1$. These definitions are made only because different authors use slightly different notations and definitions. Set $\|\Phi\|_{\mathcal{F}} := \sup_{f\in\mathcal{F}} |\Phi(f)|$. Let P be any probability measure on $(S, \mathcal{S})$ and let $\xi_i : S^{\mathbf{N}} \mapsto S$, $i \in \mathbf{N}$, be the coordinate functions. Then,

THEOREM 2.1 (Talagrand [15,16]; in this form, Giné and Guillou [7]). – *Let $\mathcal{F}$ be a measurable uniformly bounded VC class of functions, and let σ^2 and U be any numbers such that $\sigma^2 \geqslant \sup_{f\in\mathcal{F}} \mathrm{Var}_P f$, $U \geqslant \sup_{f\in\mathcal{F}} \|f\|_\infty$ and $0 < \sigma \leqslant U$. Then, there exist a* *universal constant B and constants C and L, depending only on the VC characteristics A and v of the class $\mathcal{F}$, such that*

$$E\left\|\sum_{i=1}^{n}\big(f(\xi_i)-Ef(\xi_1)\big)\right\|_{\mathcal{F}} \leqslant B\left[vU\log\frac{AU}{\sigma}+\sqrt{v}\sqrt{n\sigma^2\log\frac{AU}{\sigma}}\right], \tag{2.2}$$

and

$$\Pr\left\{\left\|\sum_{i=1}^{n}\big(f(\xi_i)-Ef(\xi_1)\big)\right\|_{\mathcal{F}}>t\right\} \leqslant L\exp\left\{-\frac{1}{L}\frac{t}{U}\log\left(1+\frac{tU}{L\big(\sqrt{n}\,\sigma+U\sqrt{\log\frac{AU}{\sigma}}\big)^2}\right)\right\} \tag{2.3}$$

whenever

$$t\geqslant C\left[U\log\frac{AU}{\sigma}+\sqrt{n}\,\sigma\sqrt{\log\frac{AU}{\sigma}}\right]. \tag{2.4}$$

If $\sigma\leqslant cU$ for some $c<1$ then $\log(AU/\sigma)$ in this proposition can be replaced by $\log(U/\sigma)$ at the price of changing the constants L and C (that now depend on c as well). With some abuse of notation we will continue denoting them as C and L when $c=1/2$. We single out inequality (2.3) for 'the Gaussian range':

COROLLARY 2.2 (Talagrand [15,16]; in this form, Giné and Guillou [7]). – *Under the assumptions of Theorem* 2.1, *if moreover*

$$0<\sigma<U/2 \quad \textit{and} \quad \sqrt{n}\,\sigma\geqslant U\sqrt{\log\frac{U}{\sigma}}, \tag{2.5}$$

there exist positive constants L and C depending only on A and v such that for all $\lambda\geqslant C$ and t satisfying

$$C\sqrt{n}\,\sigma\sqrt{\log\frac{U}{\sigma}}\leqslant t\leqslant\lambda\frac{n\sigma^2}{U}, \tag{2.6}$$

$$\Pr\left\{\left\|\sum_{i=1}^{n}\big(f(\xi_i)-Ef(\xi_1)\big)\right\|_{\mathcal{F}}>t\right\}\leqslant L\exp\left(-\frac{1}{L}\frac{\log(1+\lambda/(4L))}{\lambda}\frac{t^2}{n\sigma^2}\right). \tag{2.7}$$

In particular, if

$$t=C_2\sqrt{n}\,\sigma\sqrt{\log\frac{U}{\sigma}}$$

with $C_2\geqslant C$ then

$$\Pr\left\{\left\|\sum_{i=1}^{n}\big(f(\xi_i)-Ef(\xi_1)\big)\right\|_{\mathcal{F}}>C_2\sigma\sqrt{n}\sqrt{\log\frac{U}{\sigma}}\right\} \leqslant L\exp\left\{-\frac{C_2\log(1+C_2/(4L))}{L}\log\frac{U}{\sigma}\right\}. \tag{2.8}$$

In fact, Theorem 2.1 and Corollary 2.2 hold under the weaker condition: for every probability measure P on $(S,\mathcal{S})$ and every $0<\tau<1$,

$$N\bigl(\mathcal{F},L_2(P),\tau\|F\|_\infty\bigr)\leqslant\left(\frac{A}{\tau}\right)^v. \tag{2.1$'$}$$

When condition (2.5) holds, inequality (2.2) is optimal except for constants (see e.g. Remark 3.6 below). Note also that the set of t's given by (2.6), for which the Gaussian type inequality (2.7) holds, is precisely, up to multiplicative constants, the interval between our bound for the mean (assuming (2.5)) and the break point in the one-dimensional Bernstein's inequality for i.i.d. random variables bounded by U and with variance σ^2. Thus, this range is optimal up to constants. Whereas the main thrust in Alexander [1], Massart [9] and Talagrand [15] consists in finding the right constant L in the exponent of (2.3), the size of L is not important for us here, but what we require is a range of t as large as possible for the validity of (2.7).

It is worth mentioning at this point that the second tool we will require in the proofs below is Montgomery-Smith's [10] maximal inequality (cf. de la Peña and Giné [3]):

$$\Pr\left\{\max_{k\leqslant n}\left\|\sum_{i=1}^{k}\bigl(f(\xi_i)-Ef(\xi_i)\bigr)\right\|_{\mathcal{F}}>t\right\}\leqslant 9\Pr\left\{\left\|\sum_{i=1}^{n}\bigl(f(\xi_i)-Ef(\xi_i)\bigr)\right\|_{\mathcal{F}}>\frac{t}{30}\right\} \tag{2.9}$$

for all $t>0$.

Next we describe the hypotheses on K, f and $\{a_n\}$ for our first result.

The hypothesis on K, taken from Giné, Koltchinskii and Zinn [8], is as follows:

(K_1) K is a bounded, square integrable function in the linear span (the set of finite linear combinations) of functions $k\geqslant 0$ satisfying the following property: the subgraph of k, $\{(s,u)\colon\ k(s)\geqslant u\}$, can be represented as a finite number of Boolean operations among sets of the form $\{(s,u)\colon\ p(s,u)\geqslant\varphi(u)\}$, where p is a polynomial on $\mathbf{R}^d\times\mathbf{R}$ and φ is an arbitrary real function.

In particular this is satisfied by $K(x)=\phi(p(x))$, p being a polynomial and ϕ a bounded real function of bounded variation (e.g., Nolan and Pollard [11]). Also, e.g., if the graph of K is a pyramid (truncated or not), or if $K=I_{[-1,1]^d}$, etc.

The above condition seems awkward, but it is quite general. It is imposed because, if K satisfies (K_1), then the class of functions

$$\mathcal{F}=\left\{K\left(\frac{t-\cdot}{a}\right)\colon\ t\in\mathbf{R}^d,\ a\in\mathbf{R}^d\setminus\{0\}\right\} \tag{2.10}$$

is a bounded *VC* class of measurable functions, that is, satisfies (2.1) for some A and v and all probability measures P: this is a consequence of Theorems 4.2.1 and 4.2.4 of Dudley [4] because the family of sets

$$\bigl\{\bigl\{(s,u)\colon\ p\bigl((t-s)/h,u\bigr)\geqslant\varphi(u)\bigr\}\colon\ t\in\mathbf{R}^d,\ h>0\bigr\}$$

is contained in the family of positivity sets of a finite dimensional space of functions. We should also note that, since the map $(t,x,a)\mapsto K((t-x)/a)$ is jointly measurable, the class $\mathcal{F}$ is image admissible Suslin, hence measurable (Dudley [4, pp. 186–189]). Thus, under (K_1), the class $\mathcal{F}$ defined by (2.10) is a bounded measurable *VC* class of (measurable) functions.

We will only assume, on the density f, that it is bounded, and the conditions on the bandwidths will be those of Stute, with some regularity added, concretely,

$$a_n \searrow 0, \quad \frac{na_n^d}{|\log a_n|} \to \infty, \quad \frac{|\log a_n|}{\log\log n} \to \infty \quad \text{and} \quad a_n^d \leqslant c a_{2n}^d \tag{2.11}$$

for some $c > 0$.

THEOREM 2.3. – *Assuming* (K_1), (2.11) *and that* f *is a bounded density on* $\mathbf{R}^d$, *we have*:

$$\limsup_{n\to\infty} \sqrt{\frac{na_n^d}{\log a_n^{-1}}} \|f_n - \bar{f}_n\|_\infty = C \quad a.s., \tag{2.12}$$

where $C^2 \leqslant M^2 c\|f\|_\infty \|K\|_2^2$ *for a constant* M *that depends only on the VC characteristics of* $\mathcal{F}$.

Proof. – Monotonicity of $\{a_n\}$ (hence of $a_n \log a_n^{-1}$ once $a_n < \mathrm{e}^{-1}$) and Montgomery-Smith's maximal inequality (2.9) imply

$$\Pr\left\{\max_{2^{k-1}<n\leqslant 2^k} \sqrt{\frac{na_n^d}{\log a_n^{-1}}} \|f_n - \bar{f}_n\|_\infty > \lambda\right\} \leqslant 9\Pr\left\{\frac{1}{\sqrt{2^{k-1}a_{2^k}^d \log a_{2^k}^{-1}}} \times \sup_{\substack{t\in\mathbf{R}^d \\ a_{2^k}\leqslant a\leqslant a_{2^{k-1}}}} \left|\sum_{i=1}^{2^k}\left[K\left(\frac{t-X_i}{a}\right) - EK\left(\frac{t-X_i}{a}\right)\right]\right| > \frac{\lambda}{30}\right\} \tag{2.13}$$

for any $\lambda > 0$. As mentioned above, the assumptions on K imply that the class of functions $\mathcal{F}$ defined in (2.10) is a bounded measurable *VC* class of functions. In particular, we can apply Corollary 2.2 to the subclasses

$$\mathcal{F}_k = \left\{K\left(\frac{t-\cdot}{a}\right):\ t \in \mathbf{R}^d,\ a_{2^k} \leqslant a \leqslant a_{2^{k-1}}\right\} \tag{2.10$'$}$$

with the same L and C_2 for all k. For $\mathcal{F}_k$, since

$$\int_{\mathbf{R}^d} K^2\left(\frac{t-x}{a}\right) f(x)\,\mathrm{d}x = a^d \int_{\mathbf{R}^d} K^2(u) f(t-ua)\,\mathrm{d}u \leqslant a^d \|f\|_\infty \|K\|_2^2,$$

we can take

$$U_k = \|K\|_\infty \quad \text{and} \quad \sigma_k^2 = a_{2^{k-1}}^d \|f\|_\infty \|K\|_2^2. \tag{2.14}$$

Since $a_{2^k} \searrow 0$ and $na_n^d/\log a_n^{-1} \to \infty$, there exists $k_0 < \infty$ such that, for all $k \geqslant k_0$,

$$\sigma_k < U_k/2 \quad \text{and} \quad \sqrt{2^k}\,\sigma_k \geqslant U_k \sqrt{\log \frac{U_k}{\sigma_k}},$$

conditions required in order to apply inequality (2.8). Moreover, there exists $k_1 < \infty$ such that, for all $k \geqslant k_1 < \infty$,

$$\sigma_k\sqrt{2^k}\sqrt{\log\frac{U_k}{\sigma_k}} \leqslant \sqrt{2dca_{2^k}^d 2^{k-1}\|K\|_2^2\|f\|_\infty \log a_{2^{k-1}}^{-1}}.$$

If we take λ in (2.13) to be

$$\lambda = 30C_2\sqrt{2dc}\|K\|_2\|f\|_\infty^{1/2},$$

where C_2 is as in inequality (2.8), this inequality and (2.13) give

$$\Pr\left\{\max_{2^{k-1}<n\leqslant 2^k}\sqrt{\frac{na_n^d}{\log a_n^{-1}}}\|f_n-\bar f_n\|_\infty > 30C_2\sqrt{2dc}\|K\|_2\|f\|_\infty^{1/2}\right\}$$
$$\leqslant 9L\exp\left\{-\frac{D}{L}\log\frac{U_k}{\sigma_k}\right\}. \tag{2.15}$$

Since

$$\frac{\log(U_k/\sigma_k)}{\log k} = \frac{\log(\|K\|_\infty^2/(\|K\|_2^2\|f\|_\infty a_{2^{k-1}}^d))}{2\log k} \to \infty$$

by (2.11), it follows that the probabilities in (2.15) are summable. Now, the theorem follows by Borel–Cantelli and the zero-one law. □

Although it is not of great interest to us here, we note that, if the condition $a_n^d \leqslant ca_{2n}^d$ is replaced by $na_n^d \nearrow \infty$ in Theorem 2.3, then one can replace the subsequence $n_k = 2^k$ in the above proof by the subsequence $n_k = [\lambda^k]$ and then let λ tend to 1, to conclude that (2.12) holds with $c = 1$ in the constant C.

3. The exact bound

The assumptions on K, f and $\{a_n\}$ in this section are slightly stronger than in the previous one:

- (K_2) K satisfies (K_1) and moreover it is compactly supported (without loss of generality, with support contained in the unit cube of $\mathbf{R}^d$), and is either nonnegative or satisfies $\int_{\mathbf{R}^d} K(s)\,\mathrm{d}s = 1$;
- (D_2) the density f is bounded and uniformly continuous on $\mathbf{R}^d$;
- (W_2) the sequence $\{a_n\}$ satisfies conditions (2.11) with the last condition there strengthened to $na_n^d \nearrow \infty$.

In what follows, for $x = (x_1,\dots,x_d) \in \mathbf{R}^d$, $|x|$, the norm of x, will denote the maximum length of the coordinates, $|x| := \max_{i\leqslant d}|x_i|$. (It is irrelevant what norm we take in $\mathbf{R}^d$, but this one is more convenient.) Also, we set $\|h\|_D := \sup_{x\in D}|h(x)|$ for any set $D \subset \mathbf{R}^d$ and function h on $\mathbf{R}^d$. Let f be a probability density which is uniformly continuous on all of $\mathbf{R}^d$, and let $B_f := \{x\colon f(x) > 0\} \subseteq (\operatorname{supp} f)^\circ$, the interior of the support of f. Set

$$D_\varepsilon := \{x\colon f(x) > \varepsilon, |x| < \varepsilon^{-1}\}, \quad \varepsilon > 0.$$

Then there is $\varepsilon_0 > 0$ such that $D_\varepsilon \neq \emptyset$ for $0 < \varepsilon < \varepsilon_0$ and, by uniform continuity,

$$\lim_{\varepsilon \searrow 0} D_\varepsilon = B_f, \quad \lim_{\varepsilon \searrow 0} \| f \|_{D_\varepsilon} = \lim_{\varepsilon \searrow 0} \| f \|_{\bar{D}_\varepsilon} = \| f \|_{B_f} = \| f \|_\infty \quad \text{and} \quad \lim_{\varepsilon \searrow 0} \| f \|_{D_\varepsilon^c} = 0, \tag{3.1}$$

where $\bar{D}$ is the closure of D; actually, $\| f \|_{D_\varepsilon} = \| f \|_\infty$ for all ε small enough. To prove our second result we will examine $f_n - Ef_n$ on D_ε and on D_ε^c, and then let $\varepsilon \to 0$.

In what follows, 'a cube' means a closed hypercube of $\mathbf{R}^d$ with sides parallel to the axes, that is, a closed ball for the ℓ_d^∞ distance of $\mathbf{R}^d$.

The following proposition is basically contained in Einmahl and Mason [6]. We sketch parts of the proof for the reader's convenience since their result is given explicitly only in one dimension, and then, for more general objects.

PROPOSITION 3.1. – *Under hypotheses* (K_2), (D_2) *and* (W_2), *if D is a bounded open set and* $D \cap B_f \neq \emptyset$, *we have*:

$$\lim_{n\to\infty} \sqrt{\frac{n a_n^d}{2 \log a_n^{-d}}} \| f_n - \bar{f}_n \|_D = \| K \|_2 \| f \|_D^{1/2} \quad a.s. \tag{3.2}$$

(*where D can be replaced by its closure $\bar{D}$ in either side of the identity*).

Proof (Sketch). – To obtain the exact upper bound, the Einmahl–Mason idea consists of using Bernstein's inequality, which is more exact than Talagrand's (it has the right multiplicative constant for the 'Gaussian' part of the tail probabilities), to estimate from above the sup of $|f_n - Ef_n|$ on the nodes of a discrete grid, and then use an inequality similar to (2.3) in Theorem 2.1 (here we will use Corollary 2.2) to estimate the difference between the original empirical process and its values over the nodes of the grid.

Note that, if D is as in the statement of the proposition, then $\| f \|_D > 0$ and the diameter of D, $\operatorname{diam}(D)$, is finite. Given $\lambda > 1$ define $n_k = [\lambda^k]$, $k \in \mathbf{N}$. Let δ be a positive number. Since D is contained in a cube with side of length $\operatorname{diam}(D) < \infty$, it follows that D (and $\bar{D}$) can be covered with ℓ_k cubes $c_{k,i}$, each of side length δa_{n_k}, with

$$\ell_k \leqslant \left[\frac{\operatorname{diam}(D)}{\delta a_{n_k}} + 1 \right]^d \leqslant \left(\frac{2 \operatorname{diam}(D)}{\delta a_{n_k}} \right)^d,$$

where the last inequality holds for all k large enough (we will use the expression 'all k large enough' to mean 'all $k \geqslant k_0$ for some $k_0 < \infty$'). Let us choose points $z_{k,i} \in c_{k,i} \cap D$, $1 \leqslant i \leqslant \ell_k$.

CLAIM 1. – *For* $\lambda > 1$ *and* $\delta > 0$,

$$\limsup_{k\to\infty} \frac{1}{\sqrt{2 n_k a_{n_k}^d \log a_{n_k}^{-d}}} \max_{n_{k-1} < n \leqslant n_k} \max_{1 \leqslant i \leqslant \ell_k} \left| \sum_{r=1}^n \left[K\left(\frac{z_{k,i} - X_r}{a_{n_k}} \right) - EK\left(\frac{z_{k,i} - X}{a_{n_k}} \right) \right] \right|$$
$$\leqslant (1+\delta) \| f \|_D^{1/2} \| K \|_2 \quad a.s. \tag{3.3}$$

The proof of this claim follows directly from the maximal version of Bernstein's inequality (Einmahl and Mason [5, Lemma 2.2]), just like in the proof of (2.16), Einmahl and Mason [6], but using, instead of their estimates, the variance estimate

$$EK^2\left(\frac{z_{k,i}-X}{a_{n_k}}\right) = a_{n_k}^d \int_{\mathbf{R}^d} K^2(u) f(z_{k,i}-ua_{n_k})\,\mathrm{d}u \leqslant a_{n_k}^d \|f\|_D \|K\|_2^2 (1+\delta_k)$$

for some $\delta_k \to 0$, which follows by the continuity of f: The maximal form of Bernstein's inequality gives

$$\begin{aligned}
&\Pr\left\{\max_{1\leqslant i\leqslant \ell_k}\max_{n_{k-1}<n\leqslant n_k}\left|\sum_{r=1}^n\left[K\left(\frac{z_{k,i}-X_r}{a_{n_k}}\right)-EK\left(\frac{z_{k,i}-X}{a_{n_k}}\right)\right]\right|\right.\\
&\qquad\left.> (1+\delta)\sqrt{2n_k a_{n_k}^d (\log a_{n_k}^{-d})\|f\|_D\|K\|_2^2}\right\}\\
&\qquad\leqslant 2\ell_k \exp\Big(-2(1+\delta)^2 n_k a_{n_k}^d (\log a_{n_k}^{-d})\|f\|_D\|K\|_2^2 \Big/ \Big(2(1+\delta_k)n_k a_{n_k}^d\|f\|_D\|K\|_2^2\\
&\qquad\quad + \tfrac{4}{3}\|K\|_\infty\sqrt{2(1+\delta)^2 n_k a_{n_k}^d (\log a_{n_k}^{-d})\|f\|_D\|K\|_2^2}\Big)\Big)
\end{aligned}$$

and, since $\sqrt{n_k a_{n_k}^d (\log a_{n_k}^{-d})}/n_k a_{n_k}^d \to 0$, this bound is dominated by

$$2^{d+1}(\mathrm{diam}\ D)^d \delta^{-d} a_{n_k}^{-d} \exp\left(-\frac{(1+\delta)^2(\log a_{n_k}^{-d})}{1+\delta_k+\eta_k}\right)$$

for some $\eta_k \to 0$, in particular, for all k large enough, by $2^{d+1}(\mathrm{diam}\ D)^d\delta^{-d}a_{n_k}^{\tau}$ for some $\tau > 0$. The claim now follows because $\sum a_{n_k}^{\tau} < \infty$ for all $\tau > 0$, by (2.11) and the definition of n_k.

Let now

$$\mathcal{G}'_{k,i}(\lambda) = \left\{K\left(\frac{z_{k,i}-\cdot}{a_{n_k}}\right) - K\left(\frac{z-\cdot}{a_n}\right):\ z\in c_{k,i}\cap D,\ n_{k-1}<n\leqslant n_k\right\},\quad 1\leqslant i\leqslant \ell_k.$$

$\mathcal{G}'_{k,i}(\lambda)$ is a measurable *VC* class of functions because its elements are differences of functions belonging to two *VC* classes (proving this involves only a simple estimate of covering numbers). Moreover, there are *VC* characteristics A and v for this class that do not depend on k, i or λ, since the same is true for $\mathcal{F}$ in (2.10).

CLAIM 2. – *There exists an absolute constant C and, given $\varepsilon > 0$, there exist $\delta_\varepsilon > 0$ and $\lambda_\varepsilon > 1$ such that, if $0 < \delta \leqslant \delta_\varepsilon$, $1 < \lambda \leqslant \lambda_\varepsilon$ and $n_k = [\lambda^k]$, $k \in \mathbf{N}$, then*

$$\begin{aligned}
&\limsup_{k\to\infty}\max_{1\leqslant i\leqslant \ell_k}\max_{n_{k-1}<n\leqslant n_k}\frac{1}{\sqrt{2n_k a_{n_k}^d\log a_{n_k}^{-d}}}\left\|\sum_{r=1}^n(\delta_{X_r}-P)\right\|_{\mathcal{G}'_{k,i}(\lambda)}\\
&\qquad\leqslant C\sqrt{\varepsilon\|f\|_D}\quad a.s.
\end{aligned}\tag{3.4}$$

We will apply Proposition 2.1 to $\|\sum_{r=1}^n(\delta_{X_r}-P)\|_{\mathcal{G}'_{k,i}(\lambda)}$. To this end, we see that we can take $U = 2\|K\|_\infty$ and we must find a good candidate for σ. Consider

$$\int\left[K\left(\frac{z_{k,i}-x}{a_{n_k}}\right)-K\left(\frac{z-x}{a_n}\right)\right]^2 f(x)\,\mathrm{d}x$$
$$=a_{n_k}^d\int\left[K(u)-K\left(\frac{a_{n_k}}{a_n}\left(\frac{z-z_{k,i}}{a_{n_k}}+u\right)\right)\right]^2 f(z_{k,i}-ua_{n_k})\,\mathrm{d}u$$
$$\leqslant a_{n_k}^d\sup_{|u|\leqslant 1} f(z_{k,i}-a_{n_k}u)\int\left[K(u)-K\left(\frac{a_{n_k}}{a_n}\left(\frac{z-z_{k,i}}{a_{n_k}}+u\right)\right)\right]^2\mathrm{d}u.$$

By uniform continuity of f, since $z_{k,i}\in D$,

$$\limsup_{k\to\infty}\max_{i\leqslant \ell_k}\sup_{|u|\leqslant 1} f(z_{k,i}-a_{n_k}u)\leqslant \|f\|_D.$$

Also, since $|(1-\frac{a_{n_k}}{a_n})u|\leqslant(1-\lambda^{-1}+\lambda^{-k})|u|$ and, for $z\in c_{k,i}$, $\frac{a_{n_k}}{a_n}\frac{|z-z_{k,i}|}{a_{n_k}}\leqslant\delta$, the two arguments of K in the above integral can be made arbitrarily small just by taking λ close enough to 1, δ small enough and k large enough, which implies, by square integrability of K, that the integral itself can be made arbitrarily small. So, by uniform continuity of f and square integrability of K, given $\varepsilon>0$ there are $\lambda_\varepsilon>1$, $\delta_\varepsilon>0$ and $k_0=k_0(\lambda,\varepsilon)>0$, such that, for $0<\delta\leqslant\delta_\varepsilon$, $1<\lambda\leqslant\lambda_\varepsilon$ and $k\geqslant k_0(\lambda,\varepsilon)$, the above integral is dominated by $\varepsilon a_{n_k}^d\|f\|_D$. It then follows that we can take $\sigma^2=\varepsilon a_{n_k}^d\|f\|_D$. Obviously $\sigma\leqslant U/2$ and $\sqrt{n_k}\,\sigma\geqslant U\sqrt{\log(U/\sigma)}$ for all k large enough because $a_{n_k}\to 0$ and $n_k a_{n_k}^d/\log a_{n_k}^{-1}\to\infty$. Then, there exists $k_1<\infty$ such that, for $k\geqslant k_1$, both

$$\sigma\sqrt{n_k}\sqrt{\log\frac{U}{\sigma}}=\sqrt{\varepsilon\|f\|_D a_{n_k}^d n_k}\sqrt{\frac{1}{2}\log\frac{4\|K\|_\infty^2}{\varepsilon\|f\|_D a_{n_k}^d}}\leqslant\sqrt{\varepsilon\|f\|_D a_{n_k}^d n_k\log a_{n_k}^{-d}},$$

and

$$\log\frac{U}{\sigma}\geqslant\frac{1}{2}\log a_{n_k}^{-d}.$$

So, the exponential and the maximal inequalities from Section 2 (resp. (2.8) and (2.9)), give

$$\Pr\left\{\max_{1\leqslant i\leqslant \ell_k}\max_{n_{k-1}<n\leqslant n_k}\frac{1}{\sqrt{n_k a_{n_k}^d\log a_{n_k}^{-d}}}\left\|\sum_{r=1}^n(\delta_{X_r}-P)\right\|_{\mathcal{G}'_{k,i}(\lambda)}>30C_2\sqrt{\varepsilon\|f\|_D}\right\}$$
$$\leqslant 9L\left(\frac{2\,\mathrm{diam}(D)}{\delta a_{n_k}}\right)^d\exp\left(-\frac{C_3}{2L}\log a_{n_k}^{-d}\right)$$
$$\leqslant 9\cdot 2^d L(\mathrm{diam}\ D)^d\delta^{-d}a_{n_k}^{(C_3/(2L)-1)d}$$

for all $k\geqslant k_0\vee k_1$, where $C_3=C_2\log(1+C_2/(4L))$. We can choose C_2 large enough so that $C_3/(2L)-1>0$, in which case the above is the general term of a convergent series (as $\log a_n^{-1}/\log\log n\to\infty$), proving the claim by Borel–Cantelli.

Since, for $n_{k-1}<n\leqslant n_k$,

$$\limsup_{k\to\infty}\frac{n_k a_{n_k}^d\log a_{n_k}^{-d}}{n a_n^d\log a_n^{-d}}\leqslant\limsup_{k\to\infty}\frac{n_k a_{n_k}^d\log a_{n_k}^{-d}}{n_{k-1}a_{n_{k-1}}^d\log a_{n_{k-1}}^{-d}}\leqslant\lambda,$$

Claims 1 and 2 give that, for all $\varepsilon > 0$, $0 < \delta \leqslant \delta_\varepsilon$ and $1 < \lambda < \lambda_\varepsilon$,

$$\limsup_{n\to\infty} \sqrt{\frac{na_n^d}{2\log a_n^{-d}}} \|f_n - \bar{f}_n\|_D \leqslant \sqrt{\lambda}(1+\delta)\|K\|_2\sqrt{\|f\|_D} + C\sqrt{\lambda}\sqrt{\varepsilon\|f\|_D} \quad \text{a.s.},$$

and therefore,

$$\limsup_{n\to\infty} \sqrt{\frac{na_n^d}{2\log a_n^{-d}}} \|f_n - \bar{f}_n\|_D \leqslant \|K\|_2\sqrt{\|f\|_D} \quad \text{a.s.}$$

Regarding the reverse inequality for the lim inf, we notice first that, given $\varepsilon > 0$, there is a cube $\bar{I} \subset D$ such that both $\inf_{x\in\bar{I}} f(x) \geqslant \|f\|_D(1-\varepsilon/2)$ and $\Pr\{X \in \bar{I}\} \leqslant 1/2$: the intersection with D of any neighborhood of any point in $\bar{D}$ where $\|f\|_D$ is attained contains such a cube by continuity of f. Now we apply Proposition 2 in Einmahl and Mason [6] by following the steps of the proof of their Proposition 3 from Eq. (2.48) on. In the present case, if $\bar{I} = [\bar{a}, \bar{b}] = \{(x_1, \ldots, x_d)\colon \bar{a}_j \leqslant x_j \leqslant \bar{b}_j, j = 1, \ldots, d\}$, we take $z_{i,n} = \bar{a} + 2ia_n$ where $i = (i_1, \ldots, i_d)$, with $i_j = 1, \ldots, [(\bar{b}_j - \bar{a}_j)/(2a_n)] - 1 := \ell_n$, $j = 1, \ldots, d$ (note ℓ_n is independent of j) and $k_n = \ell_n^d$. Then, we replace h_n by a_n^d in the rest of their argument. The hypotheses $K \geqslant 0$ or $\int K(x)\,\mathrm{d}x = 1$ are needed to check the hypotheses on mean and variance in Proposition 2 of Einmahl and Mason [6]. Also, in our case, $c_f = 0$ and $d_f = 1$, which makes for considerably easier expressions. Details are omitted. □

Next, we estimate the a.s. size of the random variable $\|f_n - \bar{f}_n\|_D$, assuming only $D \cap B_f \neq \emptyset$. For this, we can proceed exactly as in Theorem 2.3 with a change in the estimate of the variance: The uniform continuity of f implies that $\lim_{\varepsilon\to 0}\sup_{x:\, d(x,D)<\varepsilon} f(x) = \|f\|_D > 0$ (note that f is not identically zero on D). So, we have that for k large enough depending on D, and $a_{2^k} \leqslant b \leqslant a_{2^{k-1}}$,

$$\int_{\mathbf{R}^d} K^2\left(\frac{t-x}{b}\right) f(x)\,\mathrm{d}x = b^d \int_{\mathbf{R}^d} K^2(u) f(t-ub)\,\mathrm{d}u \leqslant 2b^d\|f\|_D\|K\|_2^2.$$

Therefore, in this case we can take $\sigma_k^2 = 2a_{2^{k-1}}^d\|f\|_D\|K\|_2^2$ in (2.14) and we have, as a consequence of the proof of Theorem 2.3, that the following holds:

PROPOSITION 3.2. – *Under hypotheses (K_2), (D_2) and (W_2), if $D \cap B_f \neq \emptyset$, then*

$$\limsup_{n\to\infty} \sqrt{\frac{na_n^d}{\log a_n^{-d}}} \|f_n - \bar{f}_n\|_D \leqslant M\sqrt{\|f\|_D}\|K\|_2 \quad \textit{a.s.}, \tag{3.5}$$

for a constant $M < \infty$ that depends only on the VC characteristics of the class $\mathcal{F}$.

Combining the last two propositions, we obtain the main result of this article:

THEOREM 3.3. – *Under hypotheses (K_2), (D_2) and (W_2), we have*:

$$\lim_{n\to\infty} \sqrt{\frac{na_n^d}{2\log a_n^{-d}}} \|f_n - \bar{f}_n\|_\infty = \|K\|_2\|f\|_\infty^{1/2} \quad \textit{a.s.} \tag{3.6}$$

Proof. – Take $D_\varepsilon = \{x\colon\ f(x) > \varepsilon, |x| < \varepsilon^{-1}\}$ as defined immediately above (3.1). Note that D_ε satisfies the hypotheses on D in Proposition 3.1, and D_ε^c the hypothesis on D in Proposition 3.2. Then, Proposition 3.1 and the limits (3.1) give

$$\liminf_{n\to\infty}\sqrt{\frac{na_n^d}{2\log a_n^{-d}}}\|f_n-\bar f_n\|_\infty \geqslant \lim_{\varepsilon\searrow 0}\lim_{n\to\infty}\sqrt{\frac{na_n^d}{2\log a_n^{-d}}}\|f_n-\bar f_n\|_{D_\varepsilon}$$
$$= \lim_{\varepsilon\searrow 0}\|K\|_2\|f\|_{D_\varepsilon}^{1/2} = \|K\|_2\|f\|_\infty^{1/2} \quad \text{a.s.} \tag{3.7}$$

For the reverse inequality, we note that, by Propositions 3.1 and 3.2,

$$\limsup_{n\to\infty}\sqrt{\frac{na_n^d}{2\log a_n^{-d}}}\|f_n-\bar f_n\|_\infty$$
$$\leqslant \lim_{n\to\infty}\sqrt{\frac{na_n^d}{2\log a_n^{-d}}}\|f_n-\bar f_n\|_{D_\varepsilon} + \limsup_{n\to\infty}\sqrt{\frac{na_n^d}{2\log a_n^{-d}}}\|f_n-\bar f_n\|_{D_\varepsilon^c}$$
$$\leqslant \|K\|_2\|f\|_{D_\varepsilon}^{1/2} + M\|K\|_2\|f\|_{D_\varepsilon^c}^{1/2} \quad \text{a.s.}$$

Letting $\varepsilon \searrow 0$ and using (3.1) we finally obtain

$$\limsup_{n\to\infty}\sqrt{\frac{na_n^d}{2\log a_n^{-d}}}\|f_n-\bar f_n\|_\infty \leqslant \|K\|_2\|f\|_\infty^{1/2} \quad \text{a.s.},$$

which, together with (3.7), proves the theorem. □

Note that the density f in Theorem 3.3 needs not be strictly positive on $\mathbf{R}^d$, not even on the interior of its support. One might ask whether the limit (3.6) can be improved to

$$\lim_{n\to\infty}\sqrt{\frac{na_n^d}{2\log a_n^{-d}}}\left\|\frac{f_n-\bar f_n}{\sqrt f}\right\|_\infty = \|K\|_2 \quad \text{a.s.} \tag{3.8}$$

assuming only the hypotheses of Theorem 3.3 and that, in addition, f is strictly positive on $\mathbf{R}^d$. This is even false for the normal density in $\mathbf{R}$; in fact, if X, X_i, $i \in \mathbf{N}$, are i.i.d. $N(0,1)$ and e.g. $K(0) \neq 0$ and $a_n = n^{-\alpha}$ for some $\alpha \in (0,1)$, then the sequence of random variables in (3.8) is not even stochastically bounded. To see this, we first note that, by Montgomery-Smith's maximal inequality together with the fact that $\sup_{t\in\mathbf{R}} EK((t-X)/a_n)/\sqrt{f(t)} = \mathrm{O}(a_n)$ (which follows by an easy direct computation), it suffices to prove that the sequence

$$\left\{\frac{\max_{1\leqslant i\leqslant n}\left\|\frac{K((t-X_i)/a_n)}{\sqrt{f(t)}}\right\|_\infty}{\sqrt{na_n\log a_n^{-1}}}\right\}$$

is not stochastically bounded. Then, we observe that

$$\max_{1\leqslant i\leqslant n}\left\|\frac{K((t-X_i)/a_n)}{\sqrt{f(t)}}\right\|_\infty \geqslant \frac{K(0)}{\sqrt{f(\max_{i\leqslant n}|X_i|)}}.$$

Finally, given $M > 0$, for all n large enough and for $0 < \alpha' < \alpha < 1$, we have

$$\Pr\left\{ \frac{1}{\sqrt{f(\max_{i\leqslant n}|X_i|)}} > \sqrt{Mna_n \log a_n^{-1}} \right\} \geqslant \Pr\left\{ \max_{i\leqslant n}|X_i| > \sqrt{2(1-\alpha')\log n} \right\},$$

and this last probability tends to 1 because it is larger than $1 - (1 - n^{-(1-\alpha'')})^n$ for $0 < \alpha'' < \alpha'$ and n large enough. Similar estimates show that the sequence in (3.8) is not stochastically bounded if X has the symmetric exponential distribution or if X has a power type density ($f(x) = c/|x|^n$ for large $|x|$). That is, the lack of stochastic boundedness of the sequence (3.8) does not depend on the rate at which $f(x)$ tends to zero when $|x| \to \infty$. A slight modification of these computations also shows that (3.8) with $\sqrt{f}$ replaced by f^β, $0 < \beta \leqslant 1/2$, does not hold either for $1 - 2\beta < \alpha < 1$ in the normal and exponential cases.

Finally, we consider convergence of moments in Theorem 3.3.

COROLLARY 3.4. – *Under the hypotheses of Theorem* 2.3*, for all* $\lambda \in \mathbf{R}$*,*

$$\sup_n E \exp\left\{ \lambda \sqrt{\frac{na_n^d}{\log a_n^{-1}}} \|f_n - \bar{f}_n\|_\infty \right\} < \infty, \tag{3.9}$$

and, under the hypotheses of Theorem 3.3*,*

$$\lim_{n\to\infty} E \exp\left\{ \lambda \sqrt{\frac{na_n^d}{2\log a_n^{-d}}} \|f_n - \bar{f}_n\|_\infty \right\} = \exp\left\{ \lambda \|K\|_2 \|f\|_\infty^{1/2} \right\}, \tag{3.10}$$

for all $\lambda \in \mathbf{R}$*; in particular there is convergence of all moments in* (3.6)*, Theorem* 3.3*.*

Proof. – We can apply inequality (2.8) to the class of functions

$$\mathcal{F}_n = \left\{ K\left((t - \cdot)/a_n\right):\ t \in \mathbf{R}^d \right\}$$

with $U = \|K\|_\infty$ and $\sigma^2 = a_n^d \|f\|_\infty \|K\|_2^2$ (as in (2.14)), and then, taking the form of the constant in the exponent of (2.8) into account, integrate with respect to C_2 the resulting bound on the tail probabilities of the random variables in (3.9). This immediately gives (3.9). Now, (3.10) follows from Theorem 3.3 and the uniform integrability provided by inequality (3.9). □

Remark 3.5. – The continuity condition on f can be somewhat relaxed: the proof of Theorem 3.3, with only formal modifications (e.g. consider only D_ε in Proposition 3.1 and $D_\varepsilon^c \cap B_f$ in Proposition 3.2) yields that, if (K_2) and (W_2) hold and we assume that B_f is open, that f is continuous and bounded on B_f and that $\lim_{|x|\to\infty} f(x) = 0$, then

$$\lim_{n\to\infty} \sqrt{\frac{na_n^d}{2|\log a_n^d|}} \sup_{t\in B_f} \left| f_n(t) - \bar{f}_n(t) \right| = \|K\|_2 \|f\|_\infty^{1/2}.$$

This applies for instance to the exponential distribution.

Remark 3.6. – One may ask how sharp, up to constants, is the bound (2.2) for the expected value of the supremum of an empirical process. With the same choices for $\mathcal{F}_n$, U and σ as in the previous proof, this bound gives the following: if K and f are as in the first paragraph of the introduction, if

$$a_n^{d/2} \leqslant C_1 \min\left(\frac{\|f\|_\infty^{1/2}\|K\|_2}{\|K\|_\infty}, \frac{\|K\|_\infty}{\|f\|_\infty^{1/2}\|K\|_2}\right) \quad \text{and} \quad na_n^d/\log a_n^{-1} \geqslant C_2$$

for some $0 < C_1 < 1$ and $C_2 > 0$, then there exists $C_3 < \infty$ depending only on C_1, C_2, d, $\|K\|_\infty$, $\|K\|_2$ and $\|f\|_\infty$, such that

$$\sqrt{\frac{na_n^d}{\log a_n^{-1}}} E\|f_n - \bar{f}_n\|_\infty \leqslant C_3.$$

By Corollary 3.4, this is exact up to the value of the constant C_3. On the other hand U. Einmahl and one of us have observed that $\frac{AU}{\sigma}$ in equality (2.2) can be replaced (twice) by $\min(AU/\sigma, n)$ (twice). This follows from the proof of (2.2) in [7] or from an inequality in [6].

The moment bounds in Corollary 3.4 and Remark 3.6 apply to minimax risk in density estimation (Massart [9, p. 395]).

Acknowledgements

We thank Vladimir Koltchinskii for several useful conversations on the subject of this article.

REFERENCES

[1] K. Alexander, Probability inequalities for empirical processes and a law of iterated logarithm, Ann. Probab. 12 (1984) 1041–1067.
[2] P. Deheuvels, Uniform limit laws for kernel density estimators on possibly unbounded intervals, in: N. Limnios, M. Nikulin (Eds.), Recent Advances in Reliability Theory: Methodology, Practice and Inference, Birkhauser, Boston, 2000, pp. 477–492.
[3] V. de la Peña, E. Giné, Decoupling, from Dependence to Independence, Springer-Verlag, New York, 1999.
[4] R.M. Dudley, Uniform Central Limit Theorems, Cambridge University Press, Cambridge, UK, 1999.
[5] U. Einmahl, D. Mason, Some universal results on the behavior of increments of partial sums, Ann. Probab. 24 (1996) 1388–1407.
[6] U. Einmahl, D. Mason, An empirical process approach to the uniform consistency of kernel-type function estimators, J. Theoret. Probab. 13 (2000) 1–37.
[7] E. Giné, A. Guillou, On consistency of kernel density estimators for randomly censored data: rates holding uniformly over adaptive intervals, Ann. Inst. Henri Poincaré – PR 37 (2001) 503–522.
[8] E. Giné, V.I. Koltchinskii, J. Zinn, Weighted uniform consistency of kernel density estimators, Preprint, 2001.

[9] P. Massart, Rates of convergence in the central limit theorem for empirical processes, Ann. Inst. Henri Poincaré 22 (1986) 381–423.
[10] S.J. Montgomery-Smith, Comparison of sums of independent identically distributed random vectors, Probab. Math. Statist. 14 (1993) 281–285.
[11] D. Nolan, D. Pollard, U-processes: rates of convergence, Ann. Statist. 15 (1987) 780–799.
[12] M. Rosenblatt, Remarks on some nonparametric estimates of a density function, Ann. Math. Statist. 27 (1956) 832–835.
[13] B.W. Silverman, Weak and strong uniform consistency of the kernel estimate of a density and its derivatives, Ann. Statist. 6 (1978) 177–189.
[14] W. Stute, The oscillation behavior of empirical processes: the multivariate case, Ann. Probab. 12 (1984) 361–379.
[15] M. Talagrand, Sharper bounds for Gaussian and empirical processes, Ann. Probab. 22 (1994) 28–76.
[16] M. Talagrand, New concentration inequalities in product spaces, Invent. Math. 126 (1996) 505–563.