IMS Collections
From Probability to Statistics and Back: High-Dimensional Models and Processes
Vol. 9 (2013) 128–149

DOI: 10.1214/12-IMSCOLL910

On the estimation of smooth densities by strict probability densities at optimal rates in sup-norm

Evarist Giné[1,*] **and Hailin Sang**[2,†]

University of Connecticut and University of Mississippi

Abstract: It is shown that the variable bandwidth density estimators proposed by McKay [*Canad. J. Statist.* **21** (1993) 367–375; Variable kernel methods in density estimation (1993) Queen's University] following earlier findings by Abramson [*Ann. Statist.* **10** (1982) 1217–1223] approximate density functions in $C^4(\mathbb{R}^d)$ at the minimax rate in the supremum norm over bounded sets where the preliminary density estimates on which they are based are bounded away from zero. A somewhat more complicated estimator proposed by Jones, McKay and Hu [*Ann. Inst. Statist. Math.* (1994) **46** 521–535] to approximate densities in $C^6(\mathbb{R})$ can also be shown to attain minimax rates in sup norm over the same kind of sets. These estimators are strict probability densities.

1. Introduction and statement of results

Let $X_i, i \in \mathbb{N}$, be independent identically distributed (i.i.d.) observations with density function $f(t)$, $t \in \mathbb{R}$ (to be replaced below by $t \in \mathbb{R}^d$). Setting K to be a symmetric probability kernel satisfying some smoothness and differentiability properties, Abramson [1] proposed the following 'ideal' or 'oracle' variable bandwidth kernel density estimator:

$$f_A(t;h_n) = \frac{1}{nh_n}\sum_{i=1}^{n}\gamma(t,X_i)K\big(h_n^{-1}\gamma(t,X_i)(t-X_i)\big), \tag{1.1}$$

where, $\gamma(t,s) = (f(s) \vee f(t)/10)^{1/2}$, which is made into a 'real' estimator by replacing f with a preliminary estimator. In words, in Abramson's estimator the window-width about each observation X_i is inversely proportional to the square root of the density f at X_i unless $f(X_i)$ is too small, in which case the modification $\gamma(t,X_i)$ prevents against the possibility that the observation X_i will exert too much influence on the estimate of $f(t)$ if it is far from t. This estimator adapts to the local density of the data and Abramson showed that, while the variance of his

*This author is grateful for the hospitality of the MIT Mathematics Department, where he carried out most of his research on this article during a sabbatical leave.

†Most of this author's research on this article was done at the Department of Mathematics of the University of Cincinnati, and he acknowledges its hospitality.

[1]Department of Mathematics, U-3009, University of Connecticut, Storrs, CT 06269, e-mail: gine@math.uconn.edu

[2]Department of Mathematics, University of Mississippi, P.O. Box 1848, University, MS 38677, e-mail: sang@olemiss.edu

AMS 2000 subject classifications: Primary 62G07

Keywords and phrases: kernel density estimator, variable bandwidth, clipping filter, square root law, sup-norm loss, spatial adaptation, rates of convergence.

estimator is pointwise of the same order as that of the regular kernel density estimator, its bias is asymptotically of the order of h_n^4, assuming f has four uniformly continuous derivatives and $f(t) \neq 0$ (recall that the bias achieved by a symmetric non-negative kernel on such densities is of the order of only h_n^2). This ideal estimator is non-negative but it does not integrate to 1. Terrell and Scott [21] and McKay [16] constructed different examples showing that Abramson's ideal estimator without the 'clipping filter' $(f(t)/10)^{1/2}$ on $f^{1/2}(X_i)$, which is a true probability density, may have a bias of order much larger than h_n^4, and in fact their examples show that clipping is necessary for such a bias reduction. Hall, Hu and Marron [10] then proposed the ideal estimator

$$f_{HHM}(t;h_n) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{t-X_i}{h_n} f^{1/2}(X_i)\right) f^{1/2}(X_i) I\big(|t-X_i| < h_n B\big), \tag{1.2}$$

where B is a fixed constant; see also Novak [18] for a similar estimator. This estimator is non-negative and achieves the desired bias reduction but, like Abramson's, it does not integrate to 1.

McKay [15, 16] discovered a smooth clipping procedure which solves the problem of obtaining a non-negative ideal estimator that integrates to 1 and that has a bias of the order of h_n^4 for densities with four continuous derivatives. He used in (1.1) a function $\gamma(t,s) = \gamma(s)$ not dependent on t, of the form

$$\gamma(s) := \alpha(f(s)) := cp^{1/2}(f(s)/c^2), \tag{1.3}$$

where the function p is at least four times differentiable and satisfies $p(x) \geq 1$ for all x and $p(x) = x$ for all $x \geq t_0$ for some $0 < t_0 < \infty$, and $0 < c < \infty$ is a fixed number. Then, McKay's ideal estimator is

$$f_{McK}(t;h_n) = \frac{1}{nh_n}\sum_{i=1}^{n} \alpha(f(X_i)) K\big(h_n^{-1}\alpha(f(X_i))(t-X_i)\big). \tag{1.4}$$

McKay [16] and Jones, McKay and Hu [11] also show that, using $\gamma(s,h) = \alpha(f(s)) \times (1 + h^2\beta(s))$, with α as above and a convenient function β that depends on f, f' and f'', a bias of the order of h_n^6 can be achieved on densities that are six times differentiable. This new estimator may be much less practical than McKay's since, in order to implement it, one has to obtain preliminary estimates not only of f but also of its first two derivatives; moreover, these authors claim that preliminary simulations with the ideal estimators do not show significant gains by this new estimator over (1.4).

Samiuddin and El-Sayyad [19] achieved the same results by shifting the centers of the windows by random quantities. See Jones, McKay and Hu [11] who show that, by combining the two methods one can obtain an infinite number of bias reducing estimators. However, they argue that among these, the most practical is McKay's modification of Abramson's estimator based on (1.4), followed, at a distance, by the one for six times differentiable functions just mentioned above (see (1.12) below), and we will pay attention only to these two estimators in this article.

The estimators $\hat{f}(t)$ resulting from these two ideal estimators (see (1.6) and (1.12)) are non-linear and it is difficult to measure their discrepancy from $f(t)$. After Hall and Marron [9], this task is divided into two parts, (a) the study of the ideal estimator, and (b) the study of the discrepancy between the ideal and the real estimators. The literature emphasizes the bias part of the ideal estimators, and the work of McKay [15, 16] and Jones, McKay and Hu [11] on this is final. Regarding the variance part of the ideal estimators, only Giné and Sang [7] consider the uniform closeness of the (ideal) estimator to its mean, and this only for the estimator (1.2) of Hall, Hu and Marron [10]. The discrepancy between the ideal and the corresponding real estimators turns out to be exactly of the same order as the difference between the ideal and the true density f, not less, and this discrepancy was first considered in detail by Hall and Marron [9] and Hall, Hu and Marron [10], who proved that it is asymptotically of the order of $n^{-4/9}$, *pointwise* and *in probability* for bounded densities with four bounded derivatives. McKay [16] adapted their method of proof and corrected some inaccuracies from Hall and Marron [9] to show that this discrepancy for the multidimensional analogue of (1.4) is of the order of $n^{-4/(8+d)}$, also pointwise and in probability, and for dimension $d < 6$. Giné and Sang [7] showed that, in the case of the Hall, Hu and Marron estimator and in dimension 1, the discrepancy is of the order of $((\log n)/n)^{4/9}$ uniformly almost surely, as well as uniformly over densities with fixed but arbitrary bounds on their sup norm and on the sup norms of their first four derivatives, and (unnecessary) undersmoothing of the preliminary estimator was used in order to simplify several arguments. In this article we prove similar results without undersmoothing, both for the McKay [16] estimator based on the generalization of (1.4) to $\mathbb{R}^d$, for any dimension $d < \infty$, and for the estimator (1.12) below (this last, only in dimension 1). In order to obtain these results we use empirical process theory, particularly and repeatedly, Talagrand's [20] exponential inequality for empirical processes and, also at an important instance, an exponential inequality of Major [12] for canonical U-processes, tools that were not available to previous authors, and that were introduced in density estimation respectively by Einmahl and Mason [3] and Giné and Mason [6]. We now describe our results.

We first consider the real estimator corresponding to the multidimensional version of (1.4),

$$f_{McK}(t; h_{2,n}) = \frac{1}{nh_{2,n}^d} \sum_{i=1}^{n} K\left(\frac{t - X_i}{h_{2,n}} \alpha(f(X_i))\right) \alpha^d(f(X_i)) \tag{1.5}$$

with $\alpha(x) = cp^{1/2}(c^{-2}x)$, $x \geq 0$, as in (1.3), that is

$$\hat{f}(t; h_{1,n}, h_{2,n}) = \frac{1}{nh_{2,n}^d} \sum_{i=1}^{n} K\left(\frac{t - X_i}{h_{2,n}} \alpha(\hat{f}(X_i; h_{1,n}))\right) \alpha^d(\hat{f}(X_i; h_{1,n})), \tag{1.6}$$

where $\hat{f}(x; h_{1,n})$ is the classical kernel density estimator

$$\hat{f}(t; h_{1,n}) = \frac{1}{nh_{1,n}^d} \sum_{i=1}^{n} K\left(\frac{t - X_i}{h_{1,n}}\right). \tag{1.7}$$

Convergence will be uniform over the regions

$$\mathcal{D}_r = \mathcal{D}_r(f) := \{t \in \mathbb{R}^d : f(t) > r > t_0 c^2, \|t\| < 1/r\}, \tag{1.8}$$

and

$$\hat{D}_r^n(f) = \{t : \hat{f}(t; h_{1,n}) > 2r > t_0 c^2, \|t\| < 1/r\}, \tag{1.9}$$

where, c and t_0 are the constants that appear in the clipping function γ in (1.3).

The following notation will be convenient: $\mathcal{P}_C$ will denote the set of all probability densities on $\mathbb{R}^d$ that are uniformly continuous and are bounded by $C < \infty$, and $\mathcal{P}_{C,k}$ will denote the set of densities on $\mathbb{R}^d$ which together with their partial derivatives of order k or lower are bounded by $C < \infty$ and are uniformly continuous. The dependence on the dimension d will be left implicit both for the regions $\mathcal{D}_r$ and for the sets of densities $\mathcal{P}_{C,k}$.

Here is our first theorem:

Theorem 1. *Assume that the kernel K on $\mathbb{R}^d$ is non-negative, integrates to 1 and has the form $K(t) = \Phi(\|t\|^2)$ for some real twice boundedly differentiable even function Φ with support contained in $[-T, T]$, $T < \infty$. Let $\alpha(f(x))$ be defined by (1.3) for a nondecreasing clipping function $p(s)$ ($p(s) \geq 1$ for all s and $p(s) = s$ for all $s \geq t_0 \geq 1$) with five bounded and uniformly continuous derivatives, and constant $c > 0$. Set $h_{2,n} = ((\log n)/n)^{1/(8+d)}$ and $h_{1,n} = ((\log n)/n)^{1/(4+d)}$, $n \in \mathbb{N}$. Then, the estimator $\hat{f}(t; h_{1,n}, h_{2,n})$ given by (1.6) and (1.7) with the kernel, bandwidths and function α just described, satisfies*

$$\sup_{t \in \mathcal{D}_r(f)} \left|\hat{f}(t; h_{1,n}, h_{2,n}) - f(t)\right| = O_{\text{a.s.}}\left(\left(\frac{\log n}{n}\right)^{4/(8+d)}\right) \quad \textit{uniformly in } f \in \mathcal{P}_{C,4} \tag{1.10}$$

and

$$\sup_{t \in \hat{\mathcal{D}}_r^n} \left|\hat{f}(t; h_{1,n}, h_{2,n}) - f(t)\right| = O_{\text{a.s.}}\left(\left(\frac{\log n}{n}\right)^{4/(8+d)}\right) \quad \textit{uniformly in } f \in \mathcal{P}_{C,4}. \tag{1.11}$$

We should recall that, given measurable functions $Z_{n,f}(X_1, \ldots, X_n)$, X_i the coordinate functions of $(\mathbb{R}^d)^{\mathbb{N}}$, $f \in \mathcal{D}$, $\mathcal{D}$ a collection of densities on $\mathbb{R}^d$, we say that the collection of random variables $Z_{n,f}(X_1, \ldots, X_n)$ is asymptotically a.s. of the order of a_n *uniformly in* $f \in \mathcal{D}$ if there exists $C < \infty$ such that

$$\lim_{k \to \infty} \sup_{f \in \mathcal{D}} (P_f)^{\mathbb{N}} \left\{ \sup_{n \geq k} \frac{1}{a_n} |Z_{n,f}(X_1, \ldots, X_n)| > C \right\} = 0,$$

where $dP_f(x) = f(x)\,dx$, and that it is $o_{\text{a.s.}}(a_n)$ uniformly in f if this limit holds for every $C > 0$. In the text we will use $\Pr_f$ for $(P_f)^{\mathbb{N}}$, or even $\Pr$ if f is understood from the context.

Here is an example of a five times differentiable clipping function p for which $t_0 = 2$:

$$p(t) = \begin{cases} 1 + \frac{t^6}{64}(1 - 2(t-2) + \frac{9}{4}(t-2)^2 - \frac{7}{4}(t-2)^3 + \frac{7}{8}(t-2)^4) & \text{if } 0 \leq t \leq 2, \\ t & \text{if } t \geq 2, \\ 1 & \text{if } t \leq 0. \end{cases}$$

This is based on McKay's [16] example of a four times differentiable clipping function. Other examples of such functions are possible, and in particular see McKay [16] for an infinitely differentiable one.

In $\mathbb{R}$, the ideal estimator with bias $h_n^6 = h_{2,n}^6$ that we will consider has the form

$$f_{JKH}(t; h_{2,n}) = \frac{1}{nh_{2,n}} \sum_{i=1}^{n} K\left(\frac{t - X_i}{h_{2,n}} \gamma_{h_{2,n}}(X_i)\right) \gamma_{h_{2,n}}(X_i), \tag{1.12}$$

where

$$\gamma_{h_{2,n}}(x) = \frac{\alpha(f(x))}{1 + h_{2,n}^2 \beta(x)} \qquad \text{with } \alpha(f(x)) = cp^{1/2}(c^{-2} f(x)),$$
$$\beta(x) = \frac{\tau_4[f''(x)f(x) - 2(f'(x))^2]}{24\tau_2\alpha^6(f(x))}, \qquad \tau_r = \int K(x)|x|^r\,dr, \quad r > 0. \tag{1.13}$$

The true estimator corresponding to (1.12) is

$$\hat f(t; h_{1,n}, h_{2,n}, h_{3,n}, h_{4,n}) = \frac{1}{nh_{2,n}} \sum_{i=1}^n K\left(\frac{t - X_i}{h_{2,n}} \hat\gamma(X_i; h_{1,n}, h_{2,n}, h_{3,n}, h_{4,n})\right) \times \hat\gamma(X_i; h_{1,n}, h_{2,n}, h_{3,n}, h_{4,n}), \tag{1.14}$$

where

$$\hat\gamma(x; h_{1,n}, h_{2,n}, h_{3,n}, h_{4,n}) = \frac{\hat\alpha(x; h_{1,n})}{1 + h_{2,n}^2 \hat\beta(x; h_{1,n}, h_{3,n}, h_{4,n})},$$
$$\hat\alpha(x; h_{1,n}) := \alpha(\hat f(x; h_{1,n})),$$
$$\hat\beta(x; h_{1,n}, h_{3,n}, h_{4,n}) = \frac{\tau_4[f_{G_2}(x; h_{4,n})\hat f(x; h_{1,n}) - 2(f_{G_1}(x; h_{3,n}))^2]}{24\tau_2\hat\alpha^6(x; h_{1,n})}. \tag{1.15}$$

Here $\hat f(x; h_{1,n})$ is the classical kernel density estimator (1.7), and f_{G_1} and f_{G_2} are the estimators of f' and f'' given by

$$f_{G_1}(x; h_{3,n}) = \frac{1}{nh_{3,n}^2} \sum_{i=1}^n G'\left(\frac{x - X_i}{h_{3,n}}\right),$$
$$f_{G_2}(x; h_{4,n}) = \frac{1}{nh_{4,n}^3} \sum_{i=1}^n G''\left(\frac{x - X_i}{h_{4,n}}\right), \tag{1.16}$$

where G is a fourth order kernel, that is, it integrates to one and is orthogonal to x^k for $k = 1, 2, 3$.

Theorem 2. *Assume the kernel K is as in Theorem 1. Assume that the fourth order kernel G is supported by $[-T_G, T_G]$ for some $T_G < \infty$, is twice continuously differentiable, is symmetric about zero and integrates to 1. Let $\alpha(f(x))$ be defined by (1.3) for a nondecreasing clipping function $p(s)$ ($p(s) \geq 1$ for all s and $p(s) = s$ for all $s \geq t_0 \geq 1$) with seven bounded and uniformly continuous derivatives, and constant $c > 0$, and let γ and β be as in (1.13). Set $h_{1,n} = ((\log n)/n)^{1/5}$, $h_{2,n} = h_{4,n} = ((\log n)/n)^{1/13}$ and $h_{3,n} = ((\log n)/n)^{1/11}$, $n \in \mathbb{N}$. Let $\hat\gamma$, $\hat\alpha$, $\hat\beta$ be as in (1.15) for these bandwidths, and let $\hat f$ be defined by (1.14) with the kernels, bandwidths and function $\hat\gamma$ just described. Then we have*

$$\sup_{t \in D_r} \left|\hat f(t; h_{1,n}, h_{2,n}, h_{3,n}, h_{4,n}) - f(t)\right| = O_{\text{a.s.}}\left(\left(\frac{\log n}{n}\right)^{6/13}\right) \quad \text{uniformly in } f \in \mathcal{P}_{C,6}. \tag{1.17}$$

Further, for the region defined in (1.9), we have

$$\sup_{t \in \hat D_r^n} \left|\hat f(t; h_{1,n}, h_{2,n}, h_{3,n}, h_{4,n}) - f(t)\right| = O_{\text{a.s.}}\left(\left(\frac{\log n}{n}\right)^{6/13}\right) \quad \text{uniformly in } f \in \mathcal{P}_{C,6}. \tag{1.18}$$

In this article we only prove the part of this theorem corresponding to the ideal estimator. The discrepancy between ideal and real estimators is similar to the analogous part of the proof of Theorem 1 but it is quite involved, and can be consulted in http://arxiv.org/abs/1006.0971.

2. Bias and variance of the ideal estimator

In this section we consider the ideal estimators f_{McK} and f_{KJH} in several dimensions. We a) describe the bias reduction for the ideal estimators, mainly following McKay [15, 16] (see also Jones, McKay and Hu [11]), and b) show that the uniform rates of concentration of the ideal estimators about their means, not surprisingly, turn out to be the same as for regular kernel density estimators in $\mathbb{R}^d$ (Giné and Guillou [5]; Deheuvels [2] in one dimension).

2.1. Uniform bias expansions

Our ideal estimator is

$$\bar{f}(t;h_n) = \bar{f}_n(t;h_n) = \frac{1}{nh_n^d}\sum_{i=1}^n \gamma_h^d(X_i)K\big(h_n^{-1}\gamma_h(X_i)(t-X_i)\big), \qquad t\in\mathbb{R}^d. \tag{2.1}$$

Since $\gamma = \gamma_h$ may depend on h, in order to handle this carefully it is better to assume that γ depends on another variable δ and eventually have $\delta = h$:

$$\bar{f}_n(t;h,\delta) = \frac{1}{nh^d}\sum_{i=1}^n \gamma_\delta^d(X_i)K\big(h^{-1}\gamma_\delta(X_i)(t-X_i)\big), \qquad t\in\mathbb{R}^d. \tag{2.2}$$

The following proposition and its proof are contained in McKay ([16], Theorems 2.10, 1.1 and 5.13) (see also Hall [8], and particularly Jones, McKay and Hu [11] (Theorem A.1) and McKay [15]). We sketch McKay's proof in the case $d = 1$ for the reader's convenience.

Notation: we say that a function g is in $C^l(\Omega)$ if itself and its first l derivatives are bounded and uniformly continuous on Ω. More notation: for $v = (v_1,\dots,v_d) \in (\mathbb{N}\cup\{0\})^d$, we set $|v| = \sum_{i=1}^d v_i$, $D_v := D_{x_1}^{v_1}\circ\cdots\circ D_{x_d}^{v_d}$, $v! = v_1!\cdots v_d!$ and $\tau_v = \int_{\mathbb{R}^d} u_1^{v_1}\cdots u_d^{v_d}K(u)\,du$.

Proposition 1 (McKay [15, 16])**.** *Let the kernel $K:\mathbb{R}^d\mapsto\mathbb{R}$ be symmetric about zero separately in each coordinate, have bounded support and integrate to 1. Assume the density f is in $C^l(\mathbb{R}^d)$. Assume $\gamma_\delta(t)\ge c>0$ for some $c>0$ and all $t\in\mathbb{R}^d$ and $0\le\delta\le\delta_0$, for some $\delta_0>0$, and that the function $\gamma(t,\delta) := \gamma_\delta(t)$ is in $C^{l+1}(\mathbb{R}^d\times[0,\delta_0])$. Then we have*

$$E\bar{f}_n(t;h,\delta) = \sum_{k=0}^l a_{k,\delta}(t)h^k + o(h^l) \tag{2.3}$$

as $h\to 0$, uniformly in $t\in\mathbb{R}^d$ and $0\le\delta\le\delta_1$ for some $\delta_1>0$, and the set of functions $a_{k,\delta}$, which are uniformly bounded and equicontinuous, are defined as

$$a_{2k+1,\delta}(t) = 0, \quad a_{2k,\delta}(t) = \sum_{|v|=2k}\frac{\tau_v}{v!}D_v\left(\frac{f(t)}{\gamma_\delta^{2k}(t)}\right), \tag{2.4}$$

for $k\le l/2$, in particular, $a_{0,\delta}(t) = f(t)$.

Proof. (For $d = 1$.) We refer to Lemma 2.11 in McKay [16] for the details in any dimensions, whereas here we only consider the case $d = 1$. Since the functions γ_δ are bounded away from zero and their derivatives are bounded (uniformly in δ), there exists $\delta_1 > 0$ such that $\gamma_\delta(t-v) - v\gamma'_\delta(t-v)$ is bounded away from zero for all $t \in \mathbb{R}$, $\delta \in [0, \delta_0]$, and $v \in [-\delta_1, \delta_1]$. Hence, for each $t \in \mathbb{R}$ and $0 \le \delta \le \delta_0$, the function $v \mapsto U_{t,\delta}(v) := v\gamma_\delta(t-v)$ is invertible on the neighborhood $[-\delta_1, \delta_1]$ of $v = 0$. These inverse functions, say $V_{t,\delta}(u)$, are $l+1$ times differentiable with continuous derivatives, with respect to the three variables (this can be seen directly by differentiation, or using the implicit function theorem as in McKay [16] Theorem 2.10 and Lemma 2.11 for $\mathbf{x} = (t, \delta)$). If the support of K is $[-T, T]$ then $K(h^{-1}\gamma_\delta(s)(t-s)) = 0$ unless $|t-s| \le hT/c$. This implies that the change of variables

$$hz = (t-s)\gamma_\delta\big(t-(t-s)\big), \text{ that is } t-s = V_{t,\delta}(hz),$$

in the following integral is valid for all h small enough

$$\begin{aligned} E\bar{f}(t; h, \delta) &= \frac{1}{h}\int \gamma_\delta(s) f(s) K\left(\frac{t-s}{h}\gamma_\delta(s)\right) ds \\ &= -\int \gamma_\delta(t - V_{t,\delta}(hz)) f\big(t - V_{t,\delta}(hz)\big) \frac{dV_{t,\delta}(hz)}{d(hz)} K(z)\, dz. \end{aligned}$$

Now, the first statement in the proposition follows by developing the function $\gamma_\delta(t - V_{t,\delta}(hz)) f(t - V_{t,\delta}(hz)) \frac{dV_{t,\delta}(hz)}{d(hz)}$ into powers of hz and integrating, on account of the compactness of the domain of integration ($z \in [-T, T]$) and the differentiability properties of f and γ_δ. (Note that the presence of $dV(hz)/d(hz)$ in the integrand requires that the function V be $l+1$ times differentiable in order to obtain differentiability of the integrand up to the l-th order, necessary for (2.3).)

Let ψ be an infinitely differentiable function of bounded support. Then, changing variables ($t = s + hu$), developing ψ, changing variables once more ($w = u\gamma_\delta(s)$)) and integrating by parts, we obtain

$$\int \psi(t) E\bar{f}(t; h, \delta)\, dt = \int \psi(s) f(s)\, ds + \sum_{k=1}^{l} (-1)^k \frac{\tau_k h^k}{k!} \int \psi(s) \left(\frac{f(s)}{\gamma_\delta^k(s)}\right)^{(k)} ds + o(h^l),$$

and note that, by symmetry, $\tau_k = 0$ if k is odd. But by (2.3),

$$\int \psi(t) E\bar{f}(t; h, \delta)\, dt = \sum_{k=0}^{l} h^k \int \psi(t) a_{k,\delta}(t)\, dt + o(h^l),$$

and (2.4) follows by comparing the coefficients of h^k in both expansions. □

With a slightly less simple proof, one can replace the bounded support hypothesis on K by $\int (1 + |x|^l) K(x)\, dx < \infty$, as done in the above mentioned references.

Corollary 1 (McKay [15, 16]). *Let f be a density in $C^4(\mathbb{R}^d)$, let p be a clipping function in $C^5(\mathbb{R})$, set $\alpha(f(t)) = cp^{1/2}(c^{-2}f(t))$ for some $c > 0$, and define $\bar{f}(t, h)$ by equation (2.1) with $\gamma(s) = \alpha(f(s))$, that is $\bar{f}(t; h) = f_{McK}(t; h)$ (see (1.5)). Let $\mathcal{D}_r$ be as in (1.8). Then,*

$$Ef_{McK}(t; h) = f(t) + \left(\sum_{|v|=4} \tau_v D_v(1/f)/v!\right) h^4 + o(h^4) = f(t) + O(h^4) \tag{2.5}$$

as $h \to 0$, uniformly on $\mathcal{D}_r$.

Proof. For $x \in \mathcal{D}_r$, $\gamma(t) = cp^{1/2}(c^{-2}f(t)) = f^{1/2}(t)$, so that, by equation (2.4), $a_2(x) = 0$ on $\mathcal{D}_r$. So, the corollary follows from the previous proposition. □

Corollary 2 (McKay [15], Jones, McKay and Hu [11])**.** *Let f be a density in $C^6(\mathbb{R})$. Let p be a clipping function in $C^7(\mathbb{R})$ and, for some $c > 0$, set $\alpha(f(t)) = cp^{1/2}(c^{-2}f(t))$ and*

$$\beta(t) = \frac{\tau_4[f''(t)f(t) - 2(f'(t))^2]}{24\tau_2\alpha^6(t)}.$$

Define

$$\gamma_\delta(t) = \frac{\alpha(t)}{1 + \delta^2\beta(t)}, \tag{2.6}$$

and, consider $\bar{f}_n(t; h, h)$, the estimator defined by (2.2) with this γ_δ and with $\delta = h$, that is $\bar{f}_n(t; h, h) = f_{JKH}(t; h)$ (see (1.14)). Let $\mathcal{D}_r$ be as in (1.8) for dimension $d = 1$. Then,

$$\begin{aligned} &Ef_{JKH}(t; h) \\ &= f(t) + h^6\left\{\frac{1}{2}\tau_2(\beta^2)''(t) + \frac{1}{6}\tau_4\left(\frac{\beta}{f}\right)^{(4)}(t) + \frac{1}{720}\tau_6\left(\frac{1}{f^2}\right)^{(6)}(t)\right\} + o(h^6) \\ &= f(t) + O(h^6) \end{aligned} \tag{2.7}$$

as $h \to 0$, uniformly on $\mathcal{D}_r$.

Proof. By definition, $\beta(t) = \frac{\tau_4[f''(t)f(t)-2(f'(t))^2]}{24\tau_2 f^3(t)} = -\frac{\tau_4}{24\tau_2}(\frac{1}{f})''(t)$ and $\alpha(f(t)) = f^{1/2}(t)$ on $\mathcal{D}_r$, and the corollary follows by direct application of the previous proposition. □

2.2. Rate of uniform deviation from the mean

The idea of the next proposition and its proof goes back to a result of Giné and Guillou [5] who obtain the almost sure exact discrepancy rate between kernel density estimators in $\mathbb{R}^d$ and their expected values, uniformly on the whole space. A proposition closer to the one below was proved in Giné and Sang [7] for the Hall-Hu-Marron estimator, and Mason and Swanepoel [14] (see also Mason [13]) proved a general theorem that also yields the result, even with uniformity in bandwidth. The proof itself is a direct and straightforward application of the famous Talagrand's exponential inequality, in the version in Einmahl and Mason [3] (inequality A.1 combined with Proposition A.1), and particularly in Giné and Guillou ([4], Proposition 2.2; [5], Corollary 2.2). This version of Talagrand's inequality turns out to be as well the main component in the proofs of all the above mentioned results. Before proving the proposition we collect the assumptions needed and present a key lemma.

Assumptions 1. The sequence h_n has the form

$$h_n = ((\log n)/n)^\eta \tag{2.8}$$

for some $0 < \eta < 1/d$. The kernel K has the form $K(t) = \Phi(\|t\|^2)$, where Φ is bounded, has support on $[0, T]$ for some $T < \infty$ and is of bounded variation and left or right continuous. f is a bounded density function, and $\gamma_h(x) = \alpha(x)/(1+h^2\beta(x))$, where the functions α and β are continuous and bounded, α is bounded away from zero, and $0 < h \leq 1/(2\|\beta\|_\infty)^{1/2}$. The ideal estimator $\bar{f}(t; h_n) = \bar{f}_n(t)$ is defined, with these kernel, function γ_h and bandwidths h_n, as in (2.1).

All these assumptions can be weakened, but this is all we need in this article.

We recall that $N(\mathcal{K},d,\varepsilon)$, the ε covering number of the metric or pseudo-metric space $(\mathcal{K},d)$, is defined as the smallest number of (open) d-balls of radius not exceeding ε needed to cover $\mathcal{K}$. Also, a collection of measurable functions $\mathcal{K}$ on a measurable space $(S,\mathcal{S})$ is of VC type relative to an envelope F (a measurable function F such that $F(s)\ge |f(s)|$ for all $s\in S$ and $f\in\mathcal{K}$) if there exist finite constants A, v such that, for all probability measures Q on $(S,\mathcal{S})$,

$$N(\mathcal{K},L_2(Q),\varepsilon)\le\left(\frac{A\|F\|_{L_2(Q)}}{\varepsilon}\right)^v,\qquad 0<\varepsilon\le 2\sup_{f,g\in\mathcal{K}}\|f-g\|_{L_2(Q)}. \tag{2.9}$$

All but one among the classes of functions that we will consider in this article can be shown to be of VC type using the following lemma, whose proof is omitted because it is a variation on Lemma 4 of Giné and Sang [7], and whose idea comes from Nolan and Pollard [17] (inexplicably, we failed to mention this source in Giné and Sang [7]).

Lemma 1. *Let K, f and γ_h satisfy Assumptions 1. Let $\mathcal{G}$ be a uniformly bounded VC type class of measurable functions on $\mathbb{R}^d$ with respect to a constant envelope G and admitting constants A_1, v_1 in equation (2.9). Let $\mathcal{K}$ be the class of functions*

$$\mathcal{K}=\left\{K\left(\frac{t-\cdot}{h}\gamma_h(\cdot)\right)g(\cdot):t\in\mathbb{R}^d,\ 0<h<1/(2\|\beta\|_\infty)^{1/2},\ g\in\mathcal{G}\right\}. \tag{2.10}$$

Then, there exists a universal constant R such that for every Borel probability measure Q on $\mathbb{R}^d$,

$$N(\mathcal{K},L_2(Q),\varepsilon)\le\left(\frac{(R\vee A_1)\|\Phi\|_V G}{\varepsilon}\right)^{8d+20+v_1} \tag{2.11}$$

where $\|\Phi\|_V$ is the total variation norm of Φ, that is, $\mathcal{K}$ is a bounded class of functions of VC type with envelope $\|\Phi\|_V G$ and admitting characteristic constants $A=R\vee A_1$ and $v=8d+20+v_1$, independent of f.

This lemma is important for us because it allows direct application of a version of Talagrand's [20] inequality e.g. in the form given in Einmahl and Mason [3] or in Giné and Guillou [4, 5], to the effect that, if P is a probability measure on a measurable space $(S,\mathcal{S})$ and $X_i:S^{\mathbb{N}}\mapsto S$ are the coordinate functions of $S^{\mathbb{N}}$, which are i.i.d. with law P, and if a class $\mathcal{F}$ of functions is bounded, countable and of VC type for an envelope F, then there exist $0<C_i<\infty$, $1\le i\le 2$, depending only on v and A such that, for all $\lambda\ge 1\vee 2C_1$ and all t satisfying

$$C_1\sqrt{n}\sigma\sqrt{\log\frac{2\|F\|_\infty}{\sigma}}\le t\le\frac{\lambda n\sigma^2}{\|F\|_\infty}, \tag{2.12}$$

we have

$$\Pr\left\{\sup_{g\in\mathcal{F}}\left|\sum_{i=1}^n(g(X_i)-Pg)\right|>t\right\}\le C_2\exp\left(-\frac{t^2}{C_2\lambda n\sigma^2}\right), \tag{2.13}$$

where $\|F\|_\infty\ge\sigma^2\ge\sup_{g\in\mathcal{F}}\mathrm{Var}_P(g)$. The class $\mathcal{K}$ is not countable, but the continuity properties of the functions defining it imply that the sup over $g\in\mathcal{K}$ of $\left|\sum_{i=1}^k(g(X_i)-Pg)\right|$ is in fact a countable supremum. Whenever this will happen in this article we will say that the class is *measurable.*

Proposition 2. *Under the hypotheses in Assumptions 1,*

$$\sup_{t\in\mathbb{R}^d}|\bar f(t;h_n)-E\bar f(t;h_n)| = \|\bar f_n - E\bar f_n\|_\infty = O_{\text{a.s.}}\left(\sqrt{\frac{\log n}{nh_n^d}}\right)$$

uniformly over all densities f such that $\|f\|_\infty\le C$, for any $0<C<\infty$, that is, there exists $L<\infty$ such that, if $\mathcal{P}_C$ is the set of these densities, then

$$\lim_{k\to 0}\sup_{f\in\mathcal{P}_C}\Pr{}_f\left\{\sup_{n\ge k}\sqrt{\frac{nh_n^d}{\log n}}\|\bar f_n - E\bar f_n\|_\infty > L\right\}=0.$$

Proof. We have:

$$\begin{aligned}\Pr&\left\{\sqrt{\frac{nh_n^d}{\log h_n^{-1}}}\|\bar f_n - E\bar f_n\|_\infty>\lambda\right\}\\ &=\Pr\Bigg\{\sup_{t\in\mathbb{R}^d}\Bigg|\sum_{i=1}^n\Bigg[K\left(\frac{t-X_i}{h}\gamma_h(X_i)\right)\gamma_h^d(X_i)\\ &\qquad -EK\left(\frac{t-X_i}{h}\gamma_h(X_i)\right)\gamma_h^d(X_i)\Bigg]\Bigg|>\lambda\sqrt{nh_n^d\log n}\Bigg\}\end{aligned}\tag{2.14}$$

for any $\lambda>0$. Using Lemma 1 and that the class of functions $\mathcal{G}=\{\gamma_h^d: 0<h\le 1/(2\|\beta\|_\infty)^{1/2}\}$ is bounded by $2^d\|\alpha\|_\infty^d$ and is clearly of VC type with $v=1$ since $|(\frac{\alpha(x)}{1+h_1^2\beta(x)})^d-(\frac{\alpha(x)}{1+h_2^2\beta(x)})^d|\le 2d6^d\|\alpha\|_\infty^d\|\beta\|_\infty^{1/2}|h_1-h_2|$, we see that the class of functions $\mathcal{K}$ defined as in (2.10) using this $\mathcal{G}$ and the kernel K and the functions γ_h from this proposition, is VC by Lemma 1 and, since it contains the classes

$$\mathcal{F}_h=\left\{K\left(\frac{t-\cdot}{h}\gamma_h(\cdot)\right)\gamma_h^d(\cdot):t\in\mathbb{R}^d\right\},\qquad h>0,\tag{2.15}$$

that these classes are all VC and admit the same constants A and v as $\mathcal{K}$. The continuity properties of K and γ_h imply that these classes are measurable. Since $0<\nu:=(2/3)\inf_x\alpha(x)\le\|\gamma_h\|_\infty\le 2\|\alpha\|_\infty<\infty$, we have, by the usual change of variables $u=(t-x)/h$,

$$\int_{\mathbb{R}^d}K^2\left(\frac{t-x}{h}\gamma(x)\right)\gamma^{2d}(x)f(x)\,dx\le 2^d(2T^{1/2}/\nu)^d\|K\|_\infty^2\|f\|_\infty\|\alpha\|_\infty^{2d}h^d.\tag{2.16}$$

So, we can take $\sigma_h^2:=C(K,\alpha)Ch^d$ with $C(K,\alpha)=2^d(2T^{1/2}/\nu)^d||K||_\infty^2||\alpha||_\infty^{2d}$, assuming $f\in\mathcal{P}_C$. Take now $h=h_n$ satisfying (2.8). If C_1 and C_2 are the constants in Talagrand's inequality (2.13) common to all the classes $\mathcal{F}_h$, it is then clear that there is $n\ge n_0$, n_0 large enough, so that there exists $\lambda>\sqrt{C_2C(K,\alpha)C}$ such that, for all $n\ge n_0$ (note that $\log(AU/\sigma_{h_n})\simeq c\log n$)

$$C_1\sqrt{n}\sigma_{h_n}\sqrt{\log\frac{AU}{\sigma_{h_n}}}<\lambda\sqrt{nh_n^d\log n}\ll n\sigma_{h_n}^2.\tag{2.17}$$

Then, Talagrand's inequality (2.13) applied in (2.14) gives

$$\begin{aligned}\sum_{n\ge n_0}\sup_{f\in\mathcal{P}_C}\Pr&\left\{\sqrt{\frac{nh_n^d}{\log n}}||\bar f_n-E\bar f_n||_\infty>\lambda\right\}\\ &\le C_2\sum_n\exp\left(-\frac{\lambda^2\log n}{C_2C(K,\alpha)C}\right)<\infty.\end{aligned}\tag{2.18}$$

□

Corollary 1 in Section 2.1 shows that the ideal estimator $f_{McK}(t;h_{2,n})$ from (1.5) has bias of the order of $h_{2,n}^4$ uniformly in $t\in\mathcal{D}_r$ and in $f\in\mathcal{P}_{C,4}$, and Proposition 2 (with $\beta\equiv 0$ in the definition of γ_h) gives that the uniform deviation from its mean, $\sup_{t\in\mathbb{R}^d}|f_{McK}(t;h_{2,n})-Ef_{McK}(t;h_{2,n})|$, has order $O_{\text{a.s.}}(\sqrt{\frac{\log n}{nh_{2,n}^d}})$ uniformly in $t\in\mathbb{R}^d$ and in $f\in\mathcal{P}_C$ (any $0<C<\infty$). Hence, bias and uniform deviation form the mean are of the same order for $h_{2,n}=((\log n)/n)^{1/(8+d)}$, and we have

$$\sup_{t\in\mathcal{D}_r}|f_{McK}(t;h_{2,n})-f(t)|=O_{\text{a.s.}}\big(((\log n)/n)^{4/(8+d)}\big) \quad \text{uniformly in } f\in\mathcal{P}_{C,4}. \tag{2.19}$$

Likewise, Corollary 2 in Section 2.1 and Proposition 2 give that, for $h_{2,n}=((\log n)/n)^{1/13}$

$$\sup_{t\in\mathcal{D}_r}|f_{JKH}(t;h_{2,n})-f(t)|=O_{\text{a.s.}}\big(((\log n)/n)^{6/13}\big) \quad \text{uniformly in } f\in\mathcal{P}_{C,6}. \tag{2.20}$$

Once proven that the union of the classes $\mathcal{F}_h$, $0<h<1/(2\|\beta\|_\infty)^{1/2}$, is VC bounded and measurable, and that inequality (2.16) holds, we could invoke the general theorem in Mason and Swanepoel [14] instead of Talagrand's inequality to prove the previous proposition, and we would even get uniformity in bandwidth. However, the balance between the bias and the centered stochastic component of the difference $f_{McK}(t;h_{2,n})-f(t)$ (or $f_{JKH}(t;h_{2,n})-f(t)$) in (2.19) (or (2.20)) prevents us from taking advantage of uniformity in bandwidth, and since the Mason-Swanepoel result is based on (2.13), using (2.13) rather than their theorem makes for a more direct proof in our case.

3. Completion of the proof of Theorem 1

In this section we develop the proof of Theorem 1. The pattern of proof is similar to that of the main results in Hall and Marron [9], corrected in Hall, Hu and Marron [10], and particularly in Giné and Sang [7], but details are different.

Assumptions 2. The kernel K is assumed to satisfy all the conditions in Proposition 1 and Assumptions 1, and to have, besides, uniformly bounded second order partial derivatives. We also assume that the densities f are bounded and have at least four bounded and uniformly continuous derivatives, that is, $f\in\mathcal{P}_{C,4}$ for some $C<\infty$. The nondecreasing clipping function $p:\mathbb{R}\to\mathbb{R}$ is assumed to have two bounded derivatives, $p(s)\geq 1$ for all s and $p(s)=s$ for all $s\geq t_0\geq 1$. Here c and t_0 are fixed constants.We set $h_{1,n}=((\log n)/n)^{1/(4+d)}$ and $h_{2,n}=((\log n)/n)^{1/(8+d)}$, $n\in\mathbb{N}$.

Proving the theorem in $\mathbb{R}^d$ for any $d>0$ requires more precision than in dimension 1 (see Giné and Sang [7]), in particular we cannot undersmooth the preliminary estimator and we must proceed differently with several estimations.

By (2.19), in order to obtain a uniform convergence rate of $((\log n)/n)^{4/(8+d)}$ for the difference between the true estimator (1.6) and the density $f(t)$ we only need to show that the uniform convergence rate of the difference between the true estimator (1.6) and the ideal estimator (1.5) is at most of this order.

Recall $\alpha(t) := cp^{1/2}(c^{-2}t)$. Define $\delta(t) = \delta(t,n)$ by the equation

$$\begin{aligned}\delta(t) &= \frac{\alpha(\hat{f}(t;h_{1,n})) - \alpha(f(t))}{\alpha(f(t))} \\ &= \frac{p(c^{-2}\hat{f}(t;h_{1,n})) - p(c^{-2}f(t))}{p^{1/2}(c^{-2}f(t))[p^{1/2}(c^{-2}\hat{f}(t;h_{1,n})) + p^{1/2}(c^{-2}f(t))]},\end{aligned} \tag{3.1}$$

so that

$$\alpha(\hat{f}(t;h_{1,n})) = \alpha(f(t))(1+\delta(t)). \tag{3.2}$$

Since p is a Lipschitz function and $p \geq 1$,

$$|\delta(t)| \leq Bc^{-2}|\hat{f}(t;h_{1,n}) - f(t)| \tag{3.3}$$

for a constant B that depends only on p. Set

$$D(t;h_{1,n}) = \hat{f}(t;h_{1,n}) - E\hat{f}(t;h_{1,n}) \quad \text{and} \quad b(t;h_{1,n}) = E\hat{f}(t;h_{1,n}) - f(t)$$

and note that

$$\|D(\cdot;h_{1,n})\|_\infty = O_{a.s.}\left(\sqrt{\frac{\log h_{1,n}^{-1}}{nh_{1,n}^d}}\right) \quad \text{uniformly in } f \in \mathcal{P}_C \tag{3.4}$$

for all $0 < C < \infty$ by a result of Giné and Guillou [5], and that

$$\|b(\cdot;h_{1,n})\|_\infty = O_{\text{a.s.}}(h_{1,n}^2) \quad \text{uniformly in } f \in \mathcal{P}_{C,2} \tag{3.5}$$

by the classical bias computation for symmetric kernels. Then we have, by (3.3), (3.4) and (3.5),

$$\sup_{t\in\mathbb{R}^d} |\delta(t)| = O_{\text{a.s.}}(h_{1,n}^2) = o_{\text{a.s.}}(1) \quad \text{uniformly in } f \in \mathcal{P}_{C,2}. \tag{3.6}$$

We also have, for further use,

$$\delta(t) = \frac{\alpha'(f(t))[\hat{f}(t;h_{1,n}) - f(t)]}{\alpha(f(t))} + \frac{\alpha''(\eta)[\hat{f}(t;h_{1,n}) - f(t)]^2}{2\alpha(f(t))}, \tag{3.7}$$

where $\eta = \eta(t) \geq 0$ is between $\hat{f}(t;h_{1,n})$ and $f(t)$ (so, not only it depends on t but also on f and on the whole sample. Note that, since $p \geq 1$ and p' and p'' are uniformly bounded on $[0,\infty)$, we have $|\alpha''(\eta(t,h_{1,n}))| \leq c^{-3}A$ for some constant A that does not depend on n or t but only on p. It is convenient as well to record the following expansion of $\alpha^d(\hat{f})$ implied by (3.2) and (3.6):

$$\alpha^d(\hat{f}(t;h_{1,n})) = \alpha^d(f(t))(1 + d\delta(t)) + \delta_1(t) \tag{3.8}$$

with

$$\|\delta_1\|_\infty = O_{\text{a.s.}}(\|\delta\|_\infty^2) \quad \text{uniformly in } f \in \mathcal{P}_{C.2}, \tag{3.9}$$

hence, by (3.3) and (3.6),

$$\|\delta_1\|_\infty = O_{\text{a.s.}}(\|\hat{f}_n(\cdot;h_{1,n}) - f(\cdot)\|_\infty^2) \quad \text{uniformly in } f \in \mathcal{P}_{C,2}. \tag{3.10}$$

For the kernel K we have the expansion

$$
\begin{aligned}
&K\left(\frac{t-X_i}{h_{2,n}}\alpha(\hat f(X_i;h_{1,n}))\right)\\
&\quad = K\left(\frac{t-X_i}{h_{2,n}}\alpha(f(X_i))\right) \qquad (3.11)\\
&\qquad + \sum_{j=1}^d K_j'\left(\frac{t-X_i}{h_{2,n}}\alpha(f(X_i))\right)\frac{(t-X_i)_j}{h_{2,n}}\alpha(f(X_i))\delta(X_i)+\delta_2(t;X_i)
\end{aligned}
$$

with

$$\delta_2(t,x) := \sum_{j,\ell=1}^d K_{j,\ell}''(\xi)\frac{(t-x)_j(t-x)_\ell}{2h_{2,n}^2}\alpha^2(f(x))\delta^2(x),$$

where ξ is a (random) point in the line connecting the points $\frac{t-X_i}{h_{2,n}}\alpha(f(X_i))$ and $\frac{t-X_i}{h_{2,n}}\alpha(f(X_i))+\frac{t-X_i}{h_{2,n}}\alpha(f(X_i))\delta(X_i)$, as before. Since K has compact support, α is bounded from below by c (and above on bounded sets) and δ satisfies (3.6) and (3.3), we get that, for each n, on the set where $\|\hat f_n(\cdot;h_{1,n})-f(\cdot)\|_\infty^2 \le c^2/(2B)$ (so, $\|\delta\|_\infty \le 1/2$),

$$|\delta_2(t,x)| \le \frac{d^2\|K''\|_\infty}{2}(2T^{1/2}/c)^2\delta^2(x)I(\|t-x\|\le 2T^{1/2}c^{-1}h_{2,n}), \qquad (3.12)$$

in particular,

$$\sup_{t,x\in\mathbb{R}^d}|\delta_2(t,x)| = O_{\text{a.s.}}\left(\|\hat f_n(\cdot;h_{1,n})-f(\cdot)\|_\infty^2\right) \quad \text{uniformly in} \quad f\in\mathcal{P}_{C,2}. \qquad (3.13)$$

Set

$$L_1(t)=\sum_{i=1}^d t_iK_i'(t) \quad \text{and} \quad L(t)=dK(t)+L_1(t), \quad t\in\mathbb{R}^d, \qquad (3.14)$$

and notice that by symmetry, integration by parts gives that L is a second order kernel (K_j' denotes the partial derivative of K in the direction of the i-th coordinate, and t_i denotes the i-th coordinate of $t\in\mathbb{R}^d$). The decompositions (3.2), (3.8) and (3.11) then give:

$$
\begin{aligned}
&nh_{2,n}^d\hat f(t;h_{1,n},h_{2,n})\\
&\quad = nh_{2,n}^d\bar f(t;h_{2,n})\\
&\quad + \sum_{i=1}^n L\left(\frac{t-X_i}{h_{2,n}}\alpha(f(X_i))\right)\alpha^d(f(X_i))\delta(X_i) \qquad (3.15)\\
&\quad + \sum_{i=1}^n\Bigg[K\left(\frac{t-X_i}{h_{2,n}}\alpha(f(X_i))\right)\delta_1(X_i)+\alpha^d(f(X_i))\delta_2(t,X_i) \qquad (3.16)\\
&\qquad\qquad + dL_1\left(\frac{t-X_i}{h_{2,n}}\alpha(f(X_i))\right)\alpha^d(f(X_i))\delta^2(X_i)\Bigg]\\
&\quad + \sum_{i=1}^n\left[L_1\left(\frac{t-X_i}{h_{2,n}}\alpha(f(X_i))\right)\delta(X_i)\delta_1(X_i)+d\alpha^d(f(X_i))\delta(X_i)\delta_2(t,X_i)\right] \qquad (3.17)\\
&\quad + \sum_{i=1}^n \delta_2(t,X_i)\delta_1(X_i). \qquad (3.18)
\end{aligned}
$$

The sums (3.16)–(3.18) are of lower and decreasing order, and will be dealt with first. Let us consider the first term from (3.16): we have

$$\begin{aligned}
&\Pr\left(\sup_{n\geq k}\frac{1}{nh_{2,n}^d h_{1,n}^4}\sum_{i=1}^n\left|K\left(\frac{t-X_i}{h_{2,n}}\alpha(f(X_i))\right)\delta_1(X_i)\right|>\tau\right)\\
&\quad\leq \Pr\left(\sup_{n\geq k}\frac{1}{nh_{2,n}^d}\left\|\sum_{i=1}^n\left|K\left(\frac{t-X_i}{h_{2,n}}\alpha(f(X_i))\right)\right|\right\|_\infty>\tau/b\right)\\
&\qquad+\Pr\left(\max_{n\geq k}\|h_{1,n}^{-4}\delta_1\|_\infty>b\right)
\end{aligned}\tag{3.19}$$

and the last term, for suitable $b<\infty$, converges to zero uniformly in $f\in\mathcal{P}_{C,2}$ as $k\to\infty$ by (3.4), (3.5) and (3.10). Now, by change of variables, for any $r>0$,

$$E\left|K^r\left(\frac{t-X_i}{h_{2,n}}\alpha(f(X_i))\right)\right|\leq C(K,r)\|f\|_\infty h_{2,n}^d$$

for some finite constant $C(K,r)$, so that, for a suitable constant m, the first term at the right hand side of (3.19) is bounded by

$$\sum_{n=k}^\infty \Pr\left(\frac{1}{nh_{2,n}^d}\left\|\sum_{i=1}^n(|K_{n,i}|-E\,|K_{n,1}|)\right\|_\infty>\tau/b-m\right),$$

where $K_{n,i}:=K(\frac{t-X_i}{h_{2,n}}\alpha(f(X_i)))$. Here we can use Talagrand's inequality (2.13) (if a class of functions is VC type, so is the class of its absolute values, by direct computation of L_2 distances), which, by the second moment estimate above ($r=2$) and boundedness of K, and for suitable τ (in particular making $\tau/b-m>0$), shows that this series is dominated, uniformly if $f\in\mathcal{P}_C$, by $\sum_{n\geq k}e^{-\eta nh_{2,n}^d}$ for some $\eta>0$, which tends to zero. (Note that we are using Talagrand's inequality for t in the upper limit of its domain (2.12), whereas typically one uses it for t in its lower limit.) Thus, we have proved that, uniformly in $f\in\mathcal{P}_{C,2}$,

$$\frac{1}{nh_{2,n}^d}\sup_{t\in\mathbb{R}^d}\sum_{i=1}^n|K_{n,i}\delta_1(X_i)|=O_{\text{a.s.}}(h_{1,n}^4)=o_{\text{a.s.}}((\log n)/n)^{4/(8+d)}.$$

Basically, what (2.13), used in the probability decomposition (3.19), does for us is to show that the order of the first term in (3.16) is at most the order of $\|\delta_1\|_\infty$ multiplied by the order of the sup of the expectations of the (absolute values of) the summands without $\delta_1(X_i)$. We can likewise follow this pattern of proof and get similar results for all the terms in (3.16)-(3.18). One has to use that the classes of functions $\left\{L_1\left(\frac{t-\cdot}{h}\alpha(f(\cdot)\right):t\in\mathbb{R}^d,h>0\right\}$ and $\left\{I(\|t-\cdot\|\leq 2T^{1/s}c^{-1}h):t\in\mathbb{R}^d,h>0\right\}$ are VC type by Lemma 1 (the class of indicator functions is needed in order to handle the three terms in (3.16)-(3.18) that contain δ_2). We then get

$$\begin{aligned}
&\frac{1}{nh_{2,n}^d}\sup_{t\in\mathbb{R}^d}|(3.16)|=O_{\text{a.s.}}(h_{1,n}^4),\qquad \frac{1}{nh_{2,n}^d}\sup_{t\in\mathbb{R}^d}|(3.17)|=O_{\text{a.s.}}(h_{1,n}^6)\quad\text{and}\\
&\frac{1}{nh_{2,n}^d}\sup_{t\in\mathbb{R}^d}|(3.18)|=O_{\text{a.s.}}(h_{1,n}^8)
\end{aligned}\tag{3.20}$$

uniformly in $f\in\mathcal{P}_{C,2}$.

The estimation of (3.15) is much more difficult. We decompose it into several pieces using first the expansion (3.7) of δ, and then the decomposition of $\hat{f} - f$ into variance D and bias b:

$$\frac{1}{nh_{2,n}^d}\sum_{i=1}^n L_{n,i}\alpha^d(f(X_i))\delta(X_i)$$

$$= \frac{1}{dnh_{2,n}^d}\sum_{i=1}^n L_{n,i}(\alpha^d)'(f(X_i))D(X_i;h_{1,n}) \tag{3.21}$$

$$+ \frac{1}{dnh_{2,n}^d}\sum_{i=1}^n L_{n,i}(\alpha^d)'(f(X_i))b(X_i;h_{1,n}) \tag{3.22}$$

$$+ \frac{1}{2nh_{2,n}^d}\sum_{i=1}^n L_{n,i}\alpha^{d-1}(fX_i))(\alpha^d)''(\eta(X_i))[\hat{f}(X_i;h_{1,n}) - f(X_i)]^2, \tag{3.23}$$

where $L_{n,i} := L(\frac{t-X_i}{h_{2,n}}\alpha(f(X_i)))$ here. Notice that the term (3.23) is very similar to the terms in (3.16), and it has clearly the same order (recall (3.3)), that is

$$\sup_{t\in\mathbb{R}^d} |(3.23)| = O_{\text{a.s.}}(h_{1,n}^4) \text{ uniformly in } \mathcal{P}_{C,2}. \tag{3.24}$$

We devote two subsections to the estimation of the remaining two terms and anticipate that the main term is (3.21).

3.1. Estimation of the bias term (3.22)

Consider the classes of functions

$$\mathcal{Q}_n := \left\{ Q(x) = L\left(\frac{t-x}{h_{2,n}}\alpha(f(x))\right)(\alpha^d)'(f(x))b(x;h_{1,n}) : t \in \mathbb{R}^d \right\}. \tag{3.25}$$

Recall that $L(t) = \sum_{i=1}^d t_i K_i' + dK(t)$ and that $K(t) = \Phi(\|t\|^2)$, Φ twice boundedly differentiable and with bounded support. Hence, $L(t) = 2\|t\|^2\Phi'(\|t\|^2) + d\Phi(\|t\|^2)$. Since the function $u\Phi'(u) + d\Phi(u)$ is of bounded variation and bounded, the kernel L satisfies the hypotheses of the kernel K in Lemma 1 (with $s = 2$). So, these classes conform, for each n, to Lemma 1 with $\mathcal{G}$ the class consisting of the single function $(\alpha^d)'(f(x))b(x,h_{1,n})$, which, by (3.5), is uniformly bounded by $M(c,p,C)h_{1,n}^2$ if $f \in \mathcal{P}_{C,2}$, for some constant M depending only on c, p and C. We conclude by that lemma that they are VC each with envelope $M(c,p,C,K)h_{1,n}^2$ for some other constant depending on the stated objects, and all with the same characteristic constants A and v. Since the continuity hypotheses make these classes measurable, this will allow us to apply Talagrand's inequality. If we set

$$Q_i(t) = L\left(\frac{t-X_i}{h_{2,n}}\alpha(f(X_i))\right)(\alpha^d)'(f(X_i))b(X_i;h_{1,n})$$

it then follows, by the bound (3.5) on b, by boundedness and bounded support of L, by boundedness of p' and $p \geq 1$, that, for all t and all $f \in \mathcal{P}_{C,2}$,

$$\sup_{t\in\mathbb{R}^d} EQ_i^2(t) \leq \|b(\cdot;h_{1,n})\|_\infty^2 \|(\alpha^d)'\|_\infty^2 \|f\|_\infty h_{2,n}^d \sup_{t\in\mathbb{R}^d}\int_{\mathbb{R}^d} L^2\big(u\alpha\big(f(t-uh_{2,n})\big)\big)\,du$$
$$\leq M(C,c,p,K)h_{1,n}^4 h_{2,n}^d,$$

and similarly,

$$\sup_{t\in\mathbb{R}^d} |Q_i(t)| \lesssim \bar{M}(c,p,C,K)h_{1,n}^2.$$

We then have

$$\sup_{t\in\mathbb{R}^d}\left|\frac{1}{nh_{2,n}^d}\sum_{i=1}^n Q_i(t)\right| \le \sup_{t\in\mathbb{R}^d}\left|\frac{1}{nh_{2,n}^d}\sum_{i=1}^n [Q_i(t)-EQ_i(t)]\right| + \sup_{t\in\mathbb{R}^d}\frac{1}{h_{2,n}^d}|EQ_1(t)|$$

and inequality (2.13), with $\sigma^2 = M(C,c,p,K)h_{1,n}^4 h_{2,n}^d$ and $F = 2M(c,p,C,K)h_{1,n}^2$ gives that, for some D, $L>1$,

$$\sum_n \sup_{f\in\mathcal{P}_C}\Pr_f\left\{\sup_{t\in\mathbb{R}^d}\left|\sum_{i=1}^n[Q_i(t)-EQ_i(t)]\right| \ge D\sqrt{nh_{1,n}^4 h_{2,n}^d \log n}\right\}$$
$$\le C_2\sum_n \exp(-L\log n) < \infty.$$

Since $\sqrt{nh_{1,n}^4 h_{2,n}^d\log n}/(nh_{2,n}^d) \ll ((\log n)/n)^{4/(8+d)}$, the term (3.22) will be at most of order $((\log n)/n)^{4/(8+d)}$ only if the expectation term $\sup_{t\in\mathbb{R}^d}\frac{1}{h_{2,n}^d}|EQ_1(t)|$ is of this order or smaller (uniformly in $t\in\mathbb{R}^d$ and $f\in\mathcal{P}_{C,2}$). The obvious bound for $E|Q_1(t)|$, that one obtains just like the bound above for $EQ_1^2(t)$, is of the order of $h_{1,n}^2 h_{2,n}^d$, which then gives an order of $h_{1,n}^2$ for the term (3.22). This is not good enough, although it would be if we undersmoothed the preliminary estimator a little by taking $h_{1,n} = ((\log n)/n)^{2/(8+d)}$ instead of $h_{1,n} = ((\log n)/n)^{1/(4+d)}$: this works, and in fact we made this choice in Giné and Sang [7] on a related problem and $d=1$, however, some extra work along the lines suggested by Hall and Marron [9] will allow us to prove the right rate for (3.22) with the optimal $h_{1,n}$, as follows. In the setting of the proof of Proposition 1, but in dimension d, the inverse function theorem yields the existence and differentiability of $V_t(u)$, the inverse function of $U_t(v) = v\alpha(f(t-v))$ in a neighborhood of zero independent of t (this can be readily seen, or one can see it in McKay [16]). This, together with the facts that L has bounded support, α is bounded away from zero and $h_{2,n}\to 0$, justifies the change of variables $hz = (t-s)\alpha(f(t-(t-s)))$ in the expression of $EQ_1(t)$, to get (omitting the subindex n in the bandwidth),

$$\begin{aligned}
&\frac{1}{h_2^d}EQ_1(t)\\
&\quad= \frac{1}{h_2^d}\int_{\mathbb{R}^d} L\left(\frac{t-s}{h_2}\alpha(f(s))\right)(\alpha^d)'(f(s))f(s)\\
&\qquad\qquad\times\int_{\mathbb{R}^d}\frac{1}{h_1^d}K\left(\frac{s-u}{h_1}\right)(f(u)-f(s))\,du\,ds\\
&\quad= -\int_{\mathbb{R}^d}\left((\alpha^d)'(f(t-V_t(h_2z)))f(t-V_t(h_2z))\left(\frac{\partial V_t}{\partial v}\right)_{v=h_2z}\right.\\
&\qquad\left.\times\int_{\mathbb{R}^d}K(y)\,(f(t-V_t(h_2z)-yh_1)-f(t-V_t(h_2z)))\,dy\right)L(z)\,dz\\
&\quad:= \int_{\mathbb{R}^d}F(h_2z)G(h_2z)L(z)\,dz.
\end{aligned}$$

Then, using that $\int L(z)\,dz = \int z_i L(z)\,dz = 0$ and expanding, we obtain

$$\frac{1}{h_2^d}EQ_1(t) = -\frac{h_2^2}{2}\int_{\mathbb{R}^d}\left(\sum_{i,j=1}^d (F''_{i,j}G + F'_iG'_j + F'_jG'_i + FG''_{i,j})(\theta(h_2z))z_iz_j\right)L(z)\,dz.$$

Now, F and its partial derivatives are bounded, L is bounded and has bounded support, and, if $f \in \mathcal{P}_{C,4}$, then, expanding $g(t-V_t(h_2z)-yh_1)-g(t-V_t(h_2z))$, for $g=f, f'_i, f''_{i,j}$, and using the symmetry of K, we get that $\|G\|_\infty$, $\|G'_i\|_\infty$, $\|G''_{i,j}\|_\infty$ are all $O(h_1^2)$ uniformly in $f \in \mathcal{P}_{C,4}$. We conclude

$$\sup_{t\in\mathbb{R}^d} \frac{1}{h_{2,n}^d}|EQ_1(t)| = O(h_{1,n}^2 h_{2,n}^2) \quad \text{uniformly in } f \in \mathcal{P}_{C,4}, \tag{3.26}$$

and this in turn gives, together with the above application of Talagrand's inequality,

$$\begin{aligned}\sup_{t\in\mathbb{R}^d} &\left|\frac{1}{nh_{2,n}^2}\sum_{i=1}^n L\left(\frac{t-X_i}{h_{2,n}}\alpha(f(X_i))\right)(\alpha^d)'(f(X_i))b(X_i;h_{1,n})\right| \\ &= o_{\text{a.s.}}(n^{-4/(8+d)})\end{aligned} \tag{3.27}$$

uniformly in $f \in \mathcal{P}_{C,4}$.

3.2. Estimation of the variance term (3.21)

This term requires U-processes. Given a function H of two variables, and i.i.d. variables X and Y such that $H(X,Y)$ is integrable, recall the U-statistic notation $U_n(H) = \frac{1}{n(n-1)}\sum_{1\le i\ne j\le n} H(X_i,X_j)$, where the variables X_i are i.i.d. copies of X. Also, recall the second order Hoeffding projection of $H(X,Y)$, $\pi_2(H)(X,Y) = H(X,Y) - E_X H(X,Y) - E_Y H(X,Y) + EH$. If we set

$$H_t(X,Y) := L\left(\frac{t-X}{h_{2,n}^d}\alpha(f(X))\right)(\alpha^d)'(f(X))K\left(\frac{X-Y}{h_{1,n}}\right), \tag{3.28}$$

then (3.21) decomposes into a diagonal term and a U-statistic term, as follows:

$$\begin{aligned}&\frac{n^2h_{1,n}^d h_{2,n}^d}{n(n-1)}\frac{1}{nh_{2,n}^d}\sum_{i=1}^n L\left(\frac{t-X_i}{h_{2,n}}\alpha(f(X_i))\right)(\alpha^d)'(f(X_i))D(X_i;h_{1,n}) \\ &\quad = \frac{1}{n(n-1)}\sum_{i=1}^n \big(H_t(X_i,X_i) - E_Y H_t(X_i,Y)\big) + U_n\left(\pi_2(H_t(\cdot,\cdot))\right) \\ &\qquad + \frac{1}{n}\sum_{i=1}^n \big(E_X H_t(X,X_i) - EH_t\big).\end{aligned} \tag{3.29}$$

These are two empirical process terms and a canonical U-statistic term. The last term will turn out to be the only significant one.

For the first empirical process in (3.29), set $\bar{Q}_i(t) = H_t(X_i,X_i) - E_Y H_t(X_i,Y)$ and observe that, very much as in the simple bounds for moments of Q_i in the previous subsection,

$$\sup_{f\in\mathcal{P}_C}\sup_{t\in\mathbb{R}^d} E|\bar{Q}_1(t)| \le L_1 h_{2,n}^d, \quad \sup_{f\in\mathcal{P}_C}\sup_{t\in\mathbb{R}^d} E\bar{Q}_1^2(t) \le L_2 h_{2,n}^d, \quad \sup_{f\in\mathcal{P}_C}\sup_{t\in\mathbb{R}^d} |\bar{Q}_1(t)| \le L_3$$

for some finite constants $L_i = L_i(C,c,p,K)$. So,

$$\begin{aligned}\sup_{t\in\mathbb{R}^d}\frac{1}{n^2h_{1,n}^d h_{2,n}^d}\left|\sum_{i=1}^n \bar{Q}_i(t)\right| &\le \frac{1}{n^2h_{1,n}^d h_{2,n}^d}\sup_{t\in\mathbb{R}^d}\left|\sum_{i=1}^n(\bar{Q}_i(t) - E\bar{Q}_1(t))\right| \\ &\quad + \frac{L_1}{nh_{1,n}^d}.\end{aligned} \tag{3.30}$$

The sup part corresponds to the empirical process over the class of functions of x

$$\bar{\mathcal{Q}}_n = \left\{ L\left(\frac{t-x}{h_{2,n}}\alpha(f(x))\right)(\alpha^d)'(f(x))\left(K(0) - EK\left(\frac{x-X}{h_{1,n}}\right)\right) : t \in \mathbb{R}^d \right\},$$

which, by Lemma 1 is of VC type with respect to a constant envelope and admits characteristic constants A and v independent of n and f, just as in the previous subsection for the classes $\mathcal{Q}_n$ defined by (3.25). Then, as in this previous instance, Talagrand's inequality (2.13) gives that there exist $D_1, D_2 > 1$ such that

$$\sum_n \sup_f \Pr_f \left\{ \sup_{t\in\mathbb{R}^d} \left| \sum_{i=1}^n (\bar{Q}_i(t) - E\bar{Q}_1(t)) \right| > D_1 \sqrt{nh_{2,n}^d \log n} \right\} \le C_2 \sum_n n^{-D_2} < \infty,$$

which, since $\sqrt{nh_{2,n}^d \log n}/(n^2 h_{1,n}^d h_{2,n}^d) \ll ((\log n)/n)^{4/(8+d)}$ and since also $nh_{1,n}^d \gg (n/\log n)^{4/(8+d)}$, together with (3.30) yields

$$\sup_{t\in\mathbb{R}^d} \frac{1}{n^2 h_{1,n}^d h_{2,n}^d} \left| \sum_{i=1}^n (H_t(X_i, X_i) - E_Y H_t(X_i, Y)) \right| = o_{a.s.}\left(((\log n)/n)^{4/(8+d)}\right) \tag{3.31}$$

uniformly in $f \in \mathcal{P}_C$. The canonical U-statistic term in (3.29) is best handled by means of an exponential inequality of Major [12]: Let $\mathcal{F}$ be a uniformly bounded countable VC class of functions and let $\|F\|_\infty^2 \ge \sigma^2 \ge \|\mathrm{Var}(f(X_1, X_2))\|_{\mathcal{F}}$; then, there exist $0 < C_i < \infty$, $1 \le i \le 3$, depending on v and A such that, for all t satisfying

$$C_1 n\sigma \log \frac{2\|F\|_\infty}{\sigma} \le t \le \frac{n^2\sigma^3}{\|F\|_\infty^2}$$

we have

$$\Pr\left\{ \left\| \sum_{1\le i\ne j\le n} \pi_2^P f(X_i, X_j) \right\|_{\mathcal{F}} > t \right\} \le C_2 \exp\left(-C_3 \frac{t}{n\sigma}\right). \tag{3.32}$$

Major states the theorem for $\{\pi_2^P f\}$ of VC type, but it is easy to see that if $\mathcal{F}$ is VC type for F then $\{\pi_2^P f : f \in \mathcal{F}\}$ is VC type for the envelope $4F$. Our classes $\mathcal{F}$ will be the classes $\{H_t : t \in \mathbb{R}^d\}$. Note that they depend on n via $h_{i,n}$, $i = 1, 2$, but we do not display this dependence because they are VC type for a fixed constant envelope, admitting characteristic constants A and v independent of n: this follows from Lemma 1 with L instead of K, and with $\mathcal{G}$ consisting of the single bounded function $(\alpha^d)'(f(X))K(\frac{X-Y}{h_{1,n}})$ (see Section 3.1, proof that the classes defined in (3.25) are VC). Since, as is easy to check, for $f \in \mathcal{P}_C$,

$$EH_t^2(X, Y) \le M(c, p, C, K) h_{1,n}^d h_{2,n}^d,$$

we can take $\sigma^2 = M(c, p, C, K) h_{1,n}^d h_{2,n}^d$ and conclude that there exist $D_1, D_2 > 1$ such that

$$\sum_n \sup_{f:\|f\|_\infty \le C} \Pr_f \left\{ \sup_{t\in\mathbb{R}^d} |U_n(\pi_2(H_t))| > D_1 \sqrt{h_{1,n}^d h_{2,n}^d} (\log n)/n \right\} \le C_2 \sum_n n^{-D_2} < \infty.$$

Since $(\log n)/[n\sqrt{h_{1,n}^d h_{2,n}^d}] \ll ((\log n)/n)^{4/(8+d)}$ we obtain

$$\sup_{t\in\mathbb{R}^d} \frac{1}{h_{1,n}^d h_{2,n}^d} |U_n(\pi_2(H_t))| = o_{\text{a.s.}}(n^{-4/(8+d)}) \quad \text{uniformly in} \quad f \in \mathcal{P}_C. \tag{3.33}$$

Having dealt with the first two terms in the last two lines of (3.29), we will now handle the third and last, namely,

$$T(t;h_{1,n},h_{2,n})=\frac{1}{nh_{1,n}^d h_{2,n}^d}\sum_{i=1}^n (E_X H_t(X,X_i)-EH_t) \tag{3.34}$$

or, setting, for ease of notation,

$$\begin{aligned} g(t,x)&=E_X H_t(X,x)\\ &=E_X\left[L\left(\frac{t-X}{h_{2,n}}\alpha(f(X))\right)K\left(\frac{X-x}{h_{1,n}}\right)(\alpha^d)'(f(X))\right], \end{aligned} \tag{3.35}$$

$$T(t;h_{1,n},h_{2,n})=\frac{1}{nh_{1,n}^d h_{2,n}^d}\sum_{i=1}^n (g(t,X_i)-Eg(t,X)).$$

Let $\mathcal{G}$ be the class of functions $\{g(t,\cdot):t\in\mathbb{R}^d\}$. We check that this class is of VC type and apply Talagrand's inequality once more. We have, for any $s,t\in\mathbb{R}^d$, and Borel probability measure Q,

$$\begin{aligned} &E_Q\big(g(t,x)-g(s,x)\big)^2\\ &\quad\le\int E_X\left((\alpha^d)'(f(X))K\left(\frac{X-x}{h_{1,n}}\right)\right)^2\\ &\qquad\times E_X\left(L\left(\frac{t-X}{h_{2,n}}\alpha(f(X))\right)-L\left(\frac{s-X}{h_{2,n}}\alpha(f(X))\right)\right)^2 dQ(x)\\ &\quad\le\|(\alpha^d)'\|_\infty^2\|f\|_\infty h_{1,n}^d\|K\|_2^2\\ &\qquad\times\int\left(L\left(\frac{t-y}{h_{2,n}}\alpha(f(y))\right)-L\left(\frac{s-y}{h_{2,n}}\alpha(f(y))\right)\right)^2 f(y)\,dy\\ &\quad=\|(\alpha^d)'\|_\infty^2\|f\|_\infty h_{1,n}^d\|K\|_2^2 E_f(\ell_t-\ell_s)^2, \end{aligned}$$

where ℓ_s and ℓ_t are functions from the class $\mathcal{L}:=\{L(\frac{t-\cdot}{h}\alpha(f(\cdot))):t\in\mathbb{R}^d,h>0\}$ which is VC for a constant envelope by Lemma 1 (as L satisfies the hypotheses of K in that lemma -see Section 3.1). This lemma then proves that for all Q and for all $f\in\mathcal{P}_C$,

$$N(\mathcal{G},L_2(Q),\varepsilon)\le\left(\frac{R(c,p,d,K,C)h_{1,n}^{d/2}}{\varepsilon}\right)^{8d+20} \tag{3.36}$$

for $0<\varepsilon<R(c,p,d,K,C)h_{1,n}^{d/2}$ where $R=R(c,p,d,K,C)$ depends only on the stipulated parameters, in particular, $\mathcal{G}$ is VC for the constant envelope $Rh_{1,n}^{d/2}$ (that depends on n), with characteristic constants $A=1$ and $v=8d+20$ independent of n and $f\in\mathcal{P}_C$.

In order to apply Talagarand's (2.13) inequality, we need to estimate $Eg^2(t,X)$. With the change of variables $x=t-h_{1,n}w-h_{1,n}z$, $y=t-h_{2,n}z$, $u=t-h_{1,n}w-h_{2,n}z-h_{1,n}s$, of determinant $h_{1,n}^{2d}h_{2,n}^d$, we obtain

$$\begin{aligned} &Eg^2(t,X_1)\\ &\quad=\int_{\mathbb{R}^d}\left\{\int_{\mathbb{R}^d}f(x)K\left(\frac{x-u}{h_{1,n}}\right)L\left(\frac{t-x}{h_{2,n}}\alpha(f(x))\right)(\alpha^d)'(f(x))\,dx\right.\\ &\qquad\left.\times\int_{\mathbb{R}^d}f(y)K\left(\frac{y-u}{h_{1,n}}\right)L\left(\frac{t-y}{h_{2,n}}\alpha(f(y))\right)(\alpha^d)'(f(y))\,dy\right\}f(u)\,du \end{aligned} \tag{3.37}$$

$$\leq \|f\|_\infty^3 h_{1,n}^{2d} h_{2,n}^d \|(\alpha^d)'\|_\infty^2 \times \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} L\Big(\Big(\frac{h_{1,n}}{h_{2,n}}w+z\Big)\alpha(f(t-h_{1,n}w-h_{2,n}z))\Big) \times L\big(z\alpha\big(f(t-h_{2,n}z)\big)\big)K(s)K(s+w)\,ds\,dw\,dz$$
$$\leq B_K\|f\|_\infty^3 h_{1,n}^{2d}h_{2,n}^d\|(\alpha^d)'\|_\infty^2 c^{-1},$$

where the last inequality follows from the bounded support of L and K and $\alpha(t)\geq c$. Then, since the envelope F of the VC class $\mathcal{G}$ can be taken to be $B_K h_{1,n}^{d/2}$ and σ^2 to be $B_K C^3 h_{1,n}^{2d}h_{2,n}^d c^{-1}$ (by (3.36) and (3.37)), for constants B_K that depend only on K, we get by (2.12) and (2.13) that there exist constants $D_1, D_2>1$ depending only on K, C, d, p and c such that

$$\sup_{f\in\mathcal{P}_C}\Pr_f\left\{\left\|\sum_{i=1}^n(g(\cdot,X_i)-Eg(\cdot,X))\right\|_\infty \geq D_1\sqrt{nh_{1,n}^{2d}h_{2,n}^d\log n}\right\}\leq C_2 n^{-D_2},$$

and note that

$$\sqrt{nh_{1,n}^{2d}h_{2,n}^d\log n}/(nh_{1,n}^d h_{2,n}^d)=((\log n)/n)^{4/(8+d)}$$

we get that

$$\sup_{t\in\mathbb{R}^d}|T(t;h_{1,n}h_{2,n})| = O_{\text{a.s.}}\big([(\log n)/n]^{4/(8+d)}\big)\ \text{ uniformly in } f\in\mathcal{P}_C. \tag{3.38}$$

Combining the estimates (3.31), (3.33) and (3.38) with (3.29) yields

$$\sup_{t\in\mathbb{R}^d}\frac{1}{nh_{2,n}^d}\left|\sum_{i=1}^n L\left(\frac{t-X_i}{h_{2,n}}\alpha(f(X_i))\right)(\alpha^d)'(f(X_i))D(X_i;h_{1,n})\right| = O_{\text{a.s.}}\big([(\log n)/n]^{4/(8+d)}\big) \tag{3.39}$$

uniformly in $f\in\mathcal{P}_C$.

Plugging in the estimates (3.20), (3.24), (3.27) and (3.39) into the decompositions (3.15)–(3.18) and (3.21)–(3.23) of $\hat f(t;h_{1,n},h_{2,n})-\bar f(t;h_{2,n})$, yields:

Proposition 3. *Under Assumptions 2, for any $C<\infty$ the difference between the actual and the ideal estimators of a density f satisfies*

$$\sup_{t\in\mathbb{R}^d}|\hat f(t;h_{1,n},h_{2,n})-\bar f(t;h_{2,n})| = O_{\text{a.s.}}\left(\left(\frac{\log n}{n}\right)^{4/(8+d)}\right)\ \text{uniformly in } f\in\mathcal{P}_{C,4}.$$

Moreover, uniformly in $f\in\mathcal{P}_{C,4}$,

$$\sup_{t\in\mathbb{R}^d}|\hat f(t;h_{1,n},h_{2,n})-\bar f(t;h_{2,n})-T(t,h_{1,n}h_{2,n})| = o_{\text{a.s.}}\left(\left(\frac{\log n}{n}\right)^{4/(8+d)}\right).$$

3.3. End of the proof of Theorem 1

Proposition 3 together with the results in Section 2 for the bias (Corollary 1) and the variance (Proposition 2) of the ideal estimator complete the proof of the asymptotic estimate (1.10) in Theorem 1. To prove (1.11), we note that, by (3.4) and (3.5),

$\|\hat{f}(t;h_{1,n}) - f\|_\infty = O_{a.s.}(((\log n)/n)^{2/(4+d)})$ uniformly in $f \in \mathcal{P}_{C,2}$, that is, there exists $\lambda < \infty$ such that

$$\lim_{k\to\infty} \sup_{f\in\mathcal{P}_{C,2}} \Pr_f \left\{ \sup_{n\geq k} \left(\frac{n}{\log n}\right)^{2/(4+d)} \|\hat{f}(t_i;h_{1,n}) - f\|_\infty > \lambda \right\} = 0. \tag{3.40}$$

Since $\|\hat{f}(\omega) - f\|_\infty \leq \lambda((\log n)/n)^{2/(4+d)}$ implies $\hat{\mathcal{D}}_r^n(\omega) \subset \mathcal{D}_r$ as soon as $r > \lambda((\log n)/n)^{2/(4+d)}$, (1.11) follows immediately from (1.10) and (3.40). This concludes the proof of Theorem 1.

References

[1] Abramson, I. (1982). On bandwidth variation in kernel estimates—A square-root law, *Ann. Statist.* **10** 1217–1223.

[2] Deheuvels, P. (2000). Uniform limit laws for kernel density estimators on possible unbounded intervals. In: N. Limnios, M. Nikulin (Eds.), Recent Advances in Reliability Theory: Methodology, Practice and Inference. Birkhauser, Boston, pp. 477–492.

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

[4] Giné, E. and Guillou, A. (2001). On consistency of kernel density estimators for randomly censored data: rates holding uniformly over adaptive intervals, *Ann. I. H. Poincaré-Probab. et Stat.* **37** 503–522.

[5] Giné, E. and Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. I. H. Poincaré-Probab. et Stat.* **38** 907–921.

[6] Giné, E. and Mason, D. M. (2007). Laws of the iterated logarithm for the local U-statistic process. *J. Theoret. Probab.* **20** 457–485.

[7] Giné, E. and Sang, H. (2010). Uniform asymptotics for kernel density estimators with variable bandwidths. *J. Nonparametric Statistics* **22** 773–795.

[8] Hall, P. (1990). On the bias of variable bandwidth kernel estimators. *Biometrika* **77** 529–535.

[9] Hall, P. and Marron, J. S. (1988). Variable window width kernel estimates of probability densities. *Probab. Theory Related Fields* **80** 37–49. Erratum: *Probab. Theory Related Fields* **91** 133.

[10] Hall, P., Hu, T. and Marron, J. S. (1995). Improved variable window kernel estimates of probability densities. *Ann. Statist.* **23** 1–10.

[11] Jones, M. C., McKay, I. J. and Hu, T.-C. (1994). Variable location and scale kernel density estimation. *Ann. Inst. Statist. Math.* **46** 521–535.

[12] Major, P. (2006). An estimate on the supremum of a nice class of stochastic integrals and U-statistics. *Probab. Theory Related Fields* **134** 489–537.

[13] Mason, D. M. (2012). Proving consistency of non-standard kernel estimators. Preprint.

[14] Mason, D. M. and Swanepoel, J. W. H. (2010 in On-line-first). A general result on the uniform in bandwidth consistency of kernel-type function estimators. TEST, to appear.

[15] McKay, I. J. (1993). A note on bias reduction in variable kernel density estimates. *Canad. J. Statist.* **21** 367–375.

[16] McKay, I. J. (1993). Variable kernel methods in density estimation. Ph.D. dissertation, Queen's University.

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

[18] Novak, S. Yu. (1999). A generalized kernel density estimator. (Russian) *Teor. Veroyatnost. i Primenen.* **44** 634–645; translation in Theory Probab. Appl. **44** (2000) 570–583.

[19] Samiuddin, M. and El-Sayyad, G. M. (1990). On nonparametric kernel density estimates. *Biometrika* **77** 865–874.

[20] Talagrand, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126** 505–563.

[21] Terrell, G. R. and Scott, D. (1992). Variable kernel density estimation. *Ann. Statist.* **20** 1236–1265.