The Annals of Statistics
1997, Vol. 25, No. 6, 2626–2637

NONASYMPTOTIC UNIVERSAL SMOOTHING FACTORS, KERNEL COMPLEXITY AND YATRACOS CLASSES

BY LUC DEVROYE[1] AND GÁBOR LUGOSI[2]

McGill University and Pompeu Fabra University

We introduce a method to select a smoothing factor for kernel density estimation such that, for *all densities* in all dimensions, the L_1 error of the corresponding kernel estimate is not larger than three times the error of the estimate with the optimal smoothing factor plus a constant times $\sqrt{\log n/n}$, where n is the sample size, and the constant depends only on the complexity of the kernel used in the estimate. The result is nonasymptotic, that is, the bound is valid for each n. The estimate uses ideas from the minimum distance estimation work of Yatracos. As the inequality is uniform with respect to all densities, the estimate is asymptotically minimax optimal (modulo a constant) over many function classes.

1. Introduction. We are given an i.i.d. sample $X_1, \ldots, X_n$ drawn from an unknown density f on $\mathbb{R}^d$. We consider the Akaike–Parzen–Rosenblatt density estimate

$$f_{nh}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i),$$

where $K: \mathbb{R}^d \to \mathbb{R}$ is a fixed kernel with $\int K = 1$, $K_h(x) = (1/h^d)K(x/h)$, and $h > 0$ is the smoothing factor [Akaike (1954); Parzen (1962); Rosenblatt (1956)]. Many data-dependent choices for h have been proposed in the literature. Most perform well for restricted classes of densities. An exception may be found in the recent work of Devroye and Lugosi (1996), where a data-dependent smoothing factor H is introduced for which

$$\sup_f \limsup_{n\to\infty} \frac{\mathbf{E}\int |f_{nH} - f|}{\inf_h \mathbf{E}\int |f_{nh} - f|} \le 3,$$

whenever the kernel K is nonnegative, Lipschitz and of a compact support. The estimate of that paper requires various parameter choices which in turn are used to define the procedure for finding H. In this paper, a "cleaner" related estimate is proposed, and explicit nonasymptotic performance guarantees are provided that are uniform over all f.

Received June 1996; revised June 1997.
[1]Supported by NSERC Grant A3456 and by FCAR Grant 90-ER-0291.
[2]Supported by OTKA Grant F 014174.
AMS 1991 *subject classification*. 62G05.
Key words and phrases. Density estimation, kernel estimate, convergence, smoothing factor, minimum distance estimate, asymptotic optimality.

2. The estimate. To define our estimate, we first introduce the class $\mathscr{R}_k$ of kernels of the form

$$K'(x) = \sum_{i=1}^{k} \alpha_i I_{A_i}(x),$$

where I_A denotes the indicator function of a set A, $k < \infty$, $\alpha_1, \dots, \alpha_k \in \mathbb{R}$ and $A_1, \dots, A_k$ are Borel sets in $\mathbb{R}^d$ with the following property: the intersection of an infinite ray $\{x\colon x = tx_0,\ t \ge 0\}$, anchored at the origin, with any A_i is an interval. This property is needed in the proof of Lemma 3 below. Examples of such A_i's include all convex sets and all star-shaped sets (a set A is star-shaped if $x \in A$ implies $\lambda x \in A$ for all $\lambda \in [0,1]$). The A_i's need not be disjoint. However, if the A_i's are disjoint rectangles, the sum looks a bit like a Riemann approximation of a function. Thus, kernels of the type given here are called *Riemann kernels* of parameter k. Denote the class of all such functions by $\mathscr{R}_k$. The most important examples include the uniform densities on ellipsoids, balls and hypercubes.

In our estimate, we first select k and $K' \in \mathscr{R}_k$ such that

$$\int |K - K'| \le \frac{1}{n}.$$

Note that this is always possible if K is Riemann integrable. The size k as a function of n will be discussed in Section 6.

A kernel estimate with kernel K' is piecewise constant and thus easy to work with in simulations.

The second and last choice is that of a parameter $m \le n/2$ that will be used to split the data set into a small test set of size m and a large main sample of size $n - m$. Define the kernel estimates

$$f'_{n-m,h}(x) = \frac{1}{n-m} \sum_{i=1}^{n-m} K'_h(x - X_i)$$

for all $h > 0$. Let μ_m be the empirical measure defined by the rest of the data points: $X_{n-m+1}, \dots, X_n$, that is, for any Borel set $A \subseteq \mathbb{R}^d$,

$$\mu_m(A) = \frac{1}{m} \sum_{i=n-m+1}^{n} I_A(X_i).$$

Let H be that smoothing factor for which the quantity

$$\sup_{A \in \mathscr{A}} \left| \int_A f'_{n-m,h} - \mu_m(A) \right|$$

is minimal over $h \in (0, \infty)$, where $\mathscr{A}$ is a special (random) collection of sets to be defined below. If the minimum is not unique, we choose among the minimizing densities according to a prespecified rule; for example, we choose the smallest one. Observe that since $f'_{n-m,h}$ is piecewise constant and $K' \in \mathscr{R}_k$, a minimum always exists.

As $\mu_m(A)$ is close to $\int_A f$ for all A, one may expect that $\int_A f'_{n-m,h}$ is close to $\int_A f$ as well if $\mathscr{A}$ is not too large. If $\mathscr{A}$ is the class of all Borel sets, the criterion to be minimized is equal to 2 for all h and becomes useless. If $\mathscr{A}$ is too small, the closeness of $\int_A f'_{n-m,h}$ to $\int_A f$ does not imply the closeness of $f'_{n-m,h}$ to f. Thus, a compromise must be struck. Based on ideas from Yatracos (1985), for each $u, v > 0$, we define the set $A_{u,v}$ by

$$A_{u,v} = \left\{ x \in \mathbb{R}^d\colon \sum_{i=1}^{n-m} K'_u(x - X_i) \geq \sum_{i=1}^{n-m} K'_v(x - X_i) \right\} = \left\{ x\colon f'_{n-m,u}(x) \geq f'_{n-m,v}(x) \right\}.$$

We call the class of sets

$$\mathscr{A} = \{A_{u,v}\colon u > 0, v > 0\}$$

a *Yatracos class*. This class depends on $X_1, \dots, X_{n-m}$, and it becomes very rich, yet remains reasonably simple (even though it has an infinite number of members).

Finally, our estimate is

$$f_n \stackrel{\text{def}}{=} f_{n-m,H}.$$

Note that we have replaced K' by K again. The kernel K' is no longer needed. We may also use $f_n = f_{n,H}$ and refer to Devroye and Lugosi (1996) for analysis of this situation. For a practical implementation and experimental comparison, we refer to Devroye (1997).

3. Main result. Let K be a Riemann integrable kernel, and let n be a positive integer. The *kernel complexity of precision* $1/n$ of K is defined by

$$\kappa_n = \min\left\{ k\colon \text{there exists a } K' \in \mathscr{R}_k \text{ such that } \int |K - K'| \leq \frac{1}{n} \right\},$$

that is, κ_n is the smallest integer k such that there exists a Riemann kernel with parameter k whose L_1 distance from K is at most $1/n$. Clearly, if K is Riemann integrable, then $\kappa_n < \infty$ for all n. In fact, it will be shown in Section 6 that for most kernels used in practice, κ_n is usually of the order of n^α for some constant α.

THEOREM. Let K be a bounded kernel, and $m \leq n/2$. If κ_n is the kernel complexity of K of precision $1/n$, then there exists a Riemann kernel K' of parameter κ_n such that if K' is used in the estimate described in the previous section, then for all densities f,

$$\mathbf{E} \int |f_n - f| \leq 3\left(1 + \frac{2m}{n-m} + 8\sqrt{\frac{m}{n}}\right) \inf_h \mathbf{E} \int |f_{nh} - f| + 4\sqrt{\frac{\log(4e^8(m^2+1)(1 + 2\kappa_n m^2(n-m))^2)}{2m}} + \frac{4}{n}.$$

COROLLARY 1. *If we take* $m = \lfloor n/2 \rfloor$, *then*

$$\mathbf{E}\int |f_n - f| \le 43 \inf_h \mathbf{E}\int |f_{nh} - f| + c\sqrt{\frac{\log(n\kappa_n)}{n}},$$

where c *is a universal constant, independent of* f *and* K.

COROLLARY 2. *Take* $m = \lfloor n/64 \rfloor$ *and assume* $n \ge 64$. *Then simple computations show the following*:

$$\begin{aligned}\mathbf{E}\int |f_n - f| &\le \frac{128}{21} \inf_h \mathbf{E}\int |f_{nh} - f| \\ &\quad + 32\sqrt{\frac{\log(128e^8(n/64)^6 n^2 \kappa_n^2)}{n}} + \frac{4}{n} \\ &\le \frac{128}{21} \inf_h \mathbf{E}\int |f_{nh} - f| \\ &\quad + 32\sqrt{\frac{22 + 8\log(n/64) + 2\log \kappa_n}{n}} + \frac{4}{n}.\end{aligned}$$

COROLLARY 3. *If* $m = o(n)$, $m/(n^{4/5}\log n) \to \infty$ *and* $\kappa_n = O(n^\alpha)$ *for some finite* α, *then*

$$\mathbf{E}\int |f_n - f| \le (3 + o(1)) \inf_h \mathbf{E}\int |f_{nh} - f| + o(n^{-2/5}).$$

As $\liminf_{n\to\infty} n^{2/5} \inf_h \mathbf{E}\int |f_{nh} - f| > 0$ *for any* f, $K \ge 0$ *and* d [*see Devroye and Györfi* (1985)], *we have*

$$\sup_f \limsup_{n\to\infty} \frac{\mathbf{E}\int |f_n - f|}{\inf_h \mathbf{E}\int |f_{nh} - f|} \le 3.$$

This universal asymptotic bound is shared with the related estimate of Devroye and Lugosi (1996).

COROLLARY 4. *Let* $s > 0$ *be even. If the kernel* K *is bounded, symmetric and has finite nonzero* s*th moment* (*for even* s) *and zero* i*th moments for* $0 < i < s$, *then regardless of the density and the choice of* h,

$$\liminf_{n\to\infty} n^{s/(2s+1)} \inf_h \mathbf{E}\int |f_{nh} - f| > 0$$

[*Devroye* (1988), *page* 1173]. *For such higher-order kernels, let* $m = o(n)$ *such that* $m/(n^{2s/(2s+1)}\log n) \to \infty$. *Then if* $\kappa_n = O(n^\alpha)$ *for some finite* α,

$$\mathbf{E}\int |f_n - f| \le (3 + o(1)) \inf_h \mathbf{E}\int |f_{nh} - f| + o(n^{-s/(2s+1)}),$$

and therefore

$$\sup_f \limsup_{n\to\infty} \frac{\mathbf{E}\int |f_n - f|}{\inf_h \mathbf{E}\int |f_{nh} - f|} \le 3.$$

Thus, the theorem covers all kernels of finite order.

COMPUTATIONAL NOTES. The user must pick m, K and K'. If K itself is a Riemann kernel, then one should pick $K' \equiv K$. As noted earlier, the piecewise constant nature of K' ensures that $f'_{n-m,h}$ is piecewise constant and thus easy to manage without having to worry about numerical errors. When K is not Riemann, the last section of this paper gives some guidance with respect to the choice of K'. Note that the kernels K and K' need not necessarily be positive. Finally, the corollaries of the previous section show that one should not take m smaller than about $n^{4/5}\log n$.

The estimate requires that $\int |K - K'| \le 1/n$. The value $1/n$ is chosen such that the error resulting from this approximation stays small (less than $4/n$). Since this value is much smaller than the other terms in the performance bound, one may be willing to use a less accurate approximation of K. For example, using a kernel K' with $\int |K - K'| = u$ lets us replace κ_n in the upper bound by $\kappa_{\lceil 1/u \rceil}$. Clearly, one would not want to choose u much larger than $m^{-1/2}$, since then the approximation error would dominate the error. Therefore, if $\kappa_n = O(n^\alpha)$ for some α, as in most interesting cases, no more than a constant factor in the lower-order term is at stake.

4. Proof of the Theorem.

LEMMA 1. *For each n, m and for all f,*

$$\int |f_n - f| \le 3 \inf_h \int |f_{n-m,h} - f| + 4 \sup_{A\in\mathscr{A}} \left| \int_A f - \mu_m(A) \right| + 4 \int |K - K'|.$$

PROOF OF LEMMA 1. Fix an $\varepsilon > 0$, and let $\bar{f}$ be an estimate $f'_{n-m,h}$ (based on the kernel K') such that, for all $h > 0$,

$$\int |\bar{f} - f| \le \int |f'_{n-m,h} - f| + \varepsilon.$$

Then

$$\begin{aligned}
\int |f'_{n-m,H} - f| &\le \int |\bar{f} - f| + \int |f'_{n-m,H} - \bar{f}| \\
&= \int |\bar{f} - f| + 2 \sup_{A\in\mathscr{A}} \left| \int_A f'_{n-m,H} - \int_A \bar{f} \right| \quad \text{(by Scheffé's theorem)}, \\
&\le \int |\bar{f} - f| + 2 \sup_{A\in\mathscr{A}} \left| \int_A f'_{n-m,H} - \mu_m(A) \right| + 2 \sup_{A\in\mathscr{A}} \left| \mu_m(A) - \int_A \bar{f} \right| \\
&\le \int |\bar{f} - f| + 4 \sup_{A\in\mathscr{A}} \left| \mu_m(A) - \int_A \bar{f} \right| \quad \text{(by the definition of } H\text{)} \\
&\le \int |\bar{f} - f| + 4 \sup_{A\in\mathscr{A}} \left| \int_A f - \int_A \bar{f} \right| + 4 \sup_{A\in\mathscr{A}} \left| \mu_m(A) - \int_A f \right| \\
&\qquad \text{(by the triangle inequality)}
\end{aligned}$$

$$\leq 3\int|\bar{f}-f|+4\sup_{A\in\mathscr{A}}\left|\mu_m(A)-\int_A f\right| \quad \text{(by Scheffé's theorem)}$$

$$\leq 3\inf_h \int|f'_{n-m,h}-f|+\varepsilon+4\sup_{A\in\mathscr{A}}\left|\mu_m(A)-\int_A f\right|.$$

But since ε is arbitrary, we have

$$\int|f'_{n-m,H}-f|\leq 3\inf_h\int|f'_{n-m,h}-f|+4\sup_{A\in\mathscr{A}}\left|\mu_m(A)-\int_A f\right|.$$

On the other hand, since, for each h, $\int|f_{n-m,h}-f'_{n-m,h}|\leq\int|K-K'|$, for the L_1 error of our estimate $f_n=f_{n-m,H}$, we have

$$\int|f_n-f|\leq\int|f'_{n-m,H}-f|+\int|K-K'|$$

$$\leq 3\inf_h\int|f'_{n-m,h}-f|+4\sup_{A\in\mathscr{A}}\left|\mu_m(A)-\int_A f\right|+\int|K-K'|$$

(by the argument above)

$$\leq 3\inf_h\int|f_{n-m,h}-f|+4\sup_{A\in\mathscr{A}}\left|\mu_m(A)-\int_A f\right|+4\int|K-K'|,$$

which proves Lemma 1. □

The first term on the right-hand side of the inequality of Lemma 1 may be bounded by the following result.

LEMMA 2 [Devroye and Lugosi (1996)]. *Let K be a bounded kernel. If $m>0$ is a positive integer such that $2m\leq n$, then*

$$1\leq\frac{\inf_h\mathbf{E}\int|f_{n-m,h}-f|}{\inf_h\mathbf{E}\int|f_{n,h}-f|}\leq 1+\frac{2m}{n-m}+8\sqrt{\frac{m}{n}}.$$

Therefore,

$$\inf_h\mathbf{E}\int|f_{n-m,h}-f|\leq\inf_h\mathbf{E}\int|f_{n,h}-f|\left(1+\frac{2m}{n-m}+8\sqrt{\frac{m}{n}}\right).$$

To obtain suitable upper bounds for $\sup_{A\in\mathscr{A}}|\int_A f-\mu_m(A)|$, we use an inequality by Vapnik and Chervonenkis (1971) for uniform deviations of the empirical measure μ_m over the Yatracos class of sets $\mathscr{A}$.

Let $y_1,\ldots,y_m\in\mathbb{R}^d$ be fixed points. Define the *shatter coefficient*

$$s(\mathscr{A},m)=\sup_{y_1,\ldots,y_m\in\mathbb{R}^d}\left|\{y_1,\ldots,y_m\}\cap A\colon A\in\mathscr{A}\right|.$$

The purpose of the next lemma is to obtain a simple upper bound for $s(\mathscr{A},m)$ if K' is a Riemann kernel. It is convenient to let the *rank of $\mathscr{A}$* be $r(\mathscr{A})=n-m$, the size of the sample used in the definition of $\mathscr{A}$.

LEMMA 3. *Let $K' = \sum_{i=1}^{k} \alpha_i I_{A_i}$ be a Riemann kernel of parameter k. Then*

$$s(\mathscr{A}, m) \le (m+1)(1+2kmr(\mathscr{A}))^2.$$

PROOF. Set $r = r(\mathscr{A})$. Define the vector

$$z_u = \left(\sum_{i=1}^{r} K'\left(\frac{y_1 - X_i}{u}\right), \dots, \sum_{i=1}^{r} K'\left(\frac{y_m - X_i}{u}\right) \right) \in \mathbb{R}^m.$$

As $u \uparrow \infty$, each component of z_u changes every time $(y_j - X_i)/u$ enters or leaves a set A_l, $1 \le l \le k$ for some X_i, $1 \le i \le r$. Note that, for fixed $(y_j - X_i)$, the evolution is along an infinite ray anchored at the origin. By our assumption on the possible form of the sets A_l, the number of different values a component can take in its history (as $u \uparrow \infty$) is clearly bounded by $2kr$. As there are m components, the cardinality of the set of different values of z_u is bounded as

$$\left|\{z_u \colon u > 0\}\right| \le 1 + 2kmr.$$

Thus,

$$\left|\{(z_u, z_v) \colon u, v > 0\}\right| \le (1+2kmr)^2.$$

Let $\mathscr{W} = \{(w, w') \colon (w, w') = (z_u, z_v) \text{ for some } u, v > 0\}$. For fixed $(w, w') \in \mathscr{W}$, let $U_{(w,w')}$ denote the collection of all (u, v) such that $(z_u, z_v) = (w, w')$. For $(u, v) \in U_{(w,w')}$, we have

$$y_i \in A_{u,v} \quad \text{if and only if} \quad w_i \ge \left(\frac{u}{v}\right)^d w'_i,$$

where w, w' have components w_i, w'_i, respectively, $1 \le i \le m$. Thus,

$$\begin{aligned} &\left|\{\{y_1, \dots, y_m\} \cap A_{u,v} \colon (u, v) \in U_{(w,w')}\}\right| \\ &\quad \le \left|\{(I_{w_1 \ge cw'_1}, \dots, I_{w_m \ge cw'_m}) \colon c \ge 0\}\right| \le m+1. \end{aligned}$$

But then

$$\begin{aligned} \left|\{\{y_1, \dots, y_m\} \cap A_{u,v} \colon (u, v) > 0\}\right| &\le (m+1)\left|U_{(w,w')}\right| \\ &\le (m+1)(1+2kmr)^2. \end{aligned}$$

□

A variant of the Vapnik–Chervonenkis inequality [Vapnik and Chervonenkis (1971); see Devroye (1982)] states that, for $\varepsilon > 0$,

$$\begin{aligned} &\mathbf{P}\left\{ \sup_{A \in \mathscr{A}} \left| \mu_m(A) - \int_A f \right| > \varepsilon \,\middle|\, X_1, \dots, X_{n-m} \right\} \\ &\quad \le 4e^8 s(\mathscr{A}, m^2) e^{-2m\varepsilon^2} \le 4e^8 (m^2+1)(1+2km^2 r(\mathscr{A}))^2 e^{-2m\varepsilon^2}, \end{aligned}$$

where we used Lemma 3. This implies by standard bounding that

$$\mathbf{E}\left\{ \sup_{A \in \mathscr{A}} \left| \mu_m(A) - \int_A f \right| \,\middle|\, X_1, \dots, X_{n-m} \right\} \le \sqrt{\frac{\log(4e^8 s(\mathscr{A}, m^2))}{2m}}$$

[see Devroye, Györfi and Lugosi (1996), page 208]. As $r(\mathscr{A}) = n - m$ and $s(\mathscr{A}, m)$ is uniformly bounded over all (random) collections $\mathscr{A}$, the proof of the theorem is complete. □

5. Kernel complexity. In this section we obtain bounds for κ_n, the kernel complexity of precision $1/n$ appearing in the theorem, for several examples of kernels. Note that the theorem has the form

$$\mathbf{E}\int |f_n - f| \le 3\left(1 + \frac{2m}{n-m} + 8\sqrt{\frac{m}{n}}\right)\inf_h \mathbf{E}\int |f_{nh} - f| + c\sqrt{\frac{\log n}{m}}$$

for some constant c which is independent of f, whenever $\kappa_n = O(n^\alpha)$ for some $\alpha < \infty$. Such kernels are *polynomially Riemann approximable*. All kernels that we have found in papers are in this class.

UNIFORM KERNELS. If $K(x) = I_A(x)$ for a star-shaped set A, then obviously $\kappa_n = 1$ for all $n > 1$.

ISOSCELES TRIANGULAR DENSITY. If $K(x) = (1-|x|)_+$, then elementary calculation shows that, for all n, $\kappa_n \le n + 1$.

SYMMETRIC UNIMODAL KERNELS. As a first main example, consider symmetric unimodal densities (i.e., $K \ge 0$ and $\int K = 1$) on the real line. Let β be the last positive value for which $\int_\beta^\infty K \le 1/(4n)$. Partition $[0, \beta]$ and $[-\beta, 0]$ into $N = \lceil 4nK(0)\beta \rceil$ equal intervals. On each interval, let K' be constant with value equal to the average of K over that interval. Let $\gamma = \int_\beta^\infty K/K(\beta)$, and set $K'(x) = K(\beta)$ on $[\beta, \beta+\gamma]$ and $[-\beta-\gamma, -\beta]$. Note that $\int K' = 1$, $\int |K - K'| \le 1/n$ and that K' is Riemann with parameter $k = 2N + 2 \le 8nK(0)\beta + 10$. Thus, $\kappa_n \le 8nK(0)\beta + 10$.

EXAMPLE 1 (Bounded compact support densities). If $K(x) \le aI_{[-b,b]}(x)$ and K is symmetric, nonnegative and unimodal (such as the Epanechnikov–Bartlett kernel), then $\kappa_n \le 8nab + 10$.

EXAMPLE 2 (The normal density). When $K(x) = (\sqrt{2\pi})^{-1}e^{-x^2/2}$, we have $K(0) = (\sqrt{2\pi})^{-1}$. Since, for $\beta \ge 1$,

$$\int_\beta^\infty \frac{1}{\sqrt{2\pi}}\exp(-x^2/2)\,dx \le \frac{1}{\sqrt{2\pi}}\frac{1}{\beta}\exp(-\beta^2/2) \le \frac{1}{\sqrt{2\pi}}\exp(-\beta^2/2),$$

we may take $\beta = \sqrt{2\log(4n/\sqrt{2\pi})}$. Thus, for all $n > 1$,

$$\kappa_n \le \frac{8n\sqrt{\log n}}{\sqrt{\pi}} + 10.$$

EXAMPLE 3 (The Cauchy density). Take $K(x) = 1/(\pi(1 + x^2))$. Note that $K(0) = 1/\pi$, and that $\beta = \pi/(4n)$ will do. Therefore,

$$\kappa_n \le \frac{32n^2}{\pi^2} + 10.$$

EXAMPLE 4 (Densities with polynomial tails). Note that if K is a symmetric unimodal density, and $|K(x)| \le c/(1 + |x|^{\gamma+1})$ for some $c < \infty$, $\gamma > 0$, then $\kappa_n = O(n^{1+1/\gamma})$. In fact, for most cases of interest, $\kappa_n = O(n^\alpha)$ for some finite constant $\alpha > 0$. This remains so even for d dimensions.

KERNELS OF BOUNDED VARIATION. If K is symmetric and a difference of two monotone functions, that is, $K = K_1 - K_2$, $K_1 \downarrow 0$, $K_2 \downarrow 0$ on $[0, \infty)$, then each K_1, K_2 may be approximated as above. Thus, in particular, if K is of *bounded variation*, and $|K(x)| \le c/(1+|x|^{\gamma+1})$ for some $c < \infty$, $\gamma > 0$, then we may approximate with $\kappa_n = O(n^{1+1/\gamma})$. Nearly every one-dimensional kernel falls in this class.

PRODUCT KERNELS. If $K = K_1 \times \cdots \times K_d$ is a product of d univariate kernels, and if we approximate K_i with K'_i with parameter $\kappa_{nd}^{(i)}$ for all i (where $\kappa_{nd}^{(i)}$ is the kernel complexity of K_i of precision nd) and form $K' = K'_1 \times \cdots \times K'_d$, then K' is a weighted sum of indicators of product sets, and it is Riemann with parameter not exceeding $\prod_{i=1}^d \kappa_{nd}^{(i)}$. Furthermore,

$$\begin{aligned}\int |K - K'| &\le \int |K_1 \times \cdots \times K_{d-1} \times K_d - K_1 \times \cdots \times K_{d-1} \times K'_d| \\ &\qquad + \cdots + \int |K_1 \times K'_2 \cdots \times K'_d - K'_1 \times K'_2 \cdots \times K'_d| \\ &\le d\left(\frac{1}{nd}\right) \\ &= \frac{1}{n}.\end{aligned}$$

Thus, it suffices to replace κ_n throughout by $\prod_{i=1}^d \kappa_{nd}^{(i)}$, and only worry about univariate kernel approximations.

KERNELS THAT ARE FUNCTIONS OF $\|x\|$. Assume that $K(x) = M(\|x\|)$, where M is a bounded nonnegative monotone decreasing function on $[0, \infty)$. Then we may approximate M by a stepwise constant function M', and use the Riemann kernel $K'(x) = M'(\|x\|)$ in the estimate as an approximation of K. Clearly,

$$\int |K(x) - K'(x)|\, dx = \int_0^\infty c_d u^{d-1} |M(u) - M'(u)|\, du,$$

where c_d is d times the volume of the unit ball in $\mathbb{R}^d$. We may define M' as follows. Let β be the largest positive number for which $\int_\beta^\infty c_d u^{d-1} M(u)\, du \le$

$1/(2n)$. Partition $[0, \beta]$ into $N = \lceil 2nc_d M(0)\beta^d \rceil$ equal intervals. On each interval, let M' be equal to the average of M over that interval. Let $\gamma = \int_\beta^\infty c_d u^{d-1} M(u)\,du / M(\beta)$, and set $M'(u) = M(\beta)$ on $u \in [\beta, \beta + \gamma]$ and let $M'(u) = 0$ for $u > \gamma$. Clearly, $\int K' = 1$, and K' is Riemann with parameter $k = N + 1 \le 2nc_d K(0)\beta^d + 2$. Moreover,

$$\begin{aligned}\int |K(x) - K'(x)|\,dx &= \int_0^\beta c_d u^{d-1} |M(u) - M'(u)|\,du \\ &\quad + \int_\beta^\infty c_d u^{d-1} |M(u) - M'(u)|\,du \\ &\le \frac{1}{2n} + c_d \beta^{d-1} \int_0^\beta |M(u) - M'(u)|\,du \\ &\le \frac{1}{2n} + c_d \beta^{d-1} \frac{M(0)\beta}{N} \\ &\le \frac{1}{n}.\end{aligned}$$

Thus,

$$\kappa_n \le 2nc_d M(0)\beta^d + 2.$$

THE MULTIVARIATE STANDARD NORMAL KERNEL. We may apply the bound of the previous paragraph to the multivariate normal density. First note that it suffices to take $\beta = 2\sqrt{2 \log n}$. From this, we deduce that the kernel complexity is

$$\kappa_n = O(n \log^{d/2} n).$$

6. Minimax optimality and adaptation. In a minimax setting, a subclass $\mathscr{F}$ of densities of interest is given, and the minimax risk is commonly defined by

$$R_n(\mathscr{F}) \stackrel{\text{def}}{=} \inf_{f_n} \sup_{f \in \mathscr{F}} \mathbf{E} \int |f_n - f|,$$

where the infimum is over all density estimates. For many smoothness classes it is known that, if f_{nh} is the kernel estimate with an appropriate kernel K, then

$$\sup_{f \in \mathscr{F}} \inf_h \mathbf{E} \int |f_{nh} - f| \le CR_n(\mathscr{F})$$

for some universal constant $C > 1$ [see, e.g., Devroye (1987)]. In fact, the proof of such a result usually reveals a formula for h as a function of $f \in \mathscr{F}$. However, we do not know f, and so we are stuck. If we use the present data-dependent bandwidth H, then with $m = o(n)$ and $\kappa_n = O(n^a)$ for some finite a, we have

$$\sup_{f \in \mathscr{F}} \mathbf{E} \int |f_{nH} - f| \le (3C + o(1))R_n(\mathscr{F}) + O\big(\sqrt{\log n / m}\big).$$

In many cases, the last term is negligible. Thus, our results may be used for existence proofs of minimax optimal estimators; if one can find a formula $h = h(f, n)$ for the bandwidth that gives a certain rate, then that same rate will be achieved with H.

A more interesting problem occurs when we define $\mathscr{F}$ up to a parameter, such as the class of all Lipschitz densities on $[0, 1]$ with unknown Lipschitz constant α. For fixed α, the class is denoted by $\mathscr{F}_\alpha$. Assume that we know that, for each α,

$$\sup_{f\in\mathscr{F}_\alpha} \inf_h \mathbf{E}\int |f_{nh} - f| \le C_\alpha R_n(\mathscr{F}_\alpha). \tag{1}$$

When α is not given beforehand, the challenge is to find a data-dependent H such that

$$\sup_\alpha \frac{\sup_{f\in\mathscr{F}_\alpha} \mathbf{E}\int |f_{nH} - f|}{R_n(\mathscr{F}_\alpha)} \le C'$$

for some suitable constant C'. In that case, we may say that H adapts itself nicely to the union of the classes $\mathscr{F}_\alpha$. Such a point of view is not without merit. Assume that H is picked by the method of this paper. Then, assuming that m grows linearly with n, and that $\kappa_n = O(n^a)$ for some finite $a > 0$, we see that there exist universal constants D and E such that

$$\begin{aligned}\sup_\alpha \frac{\sup_{f\in\mathscr{F}_\alpha} \mathbf{E}\int |f_{nH} - f|}{R_n(\mathscr{F}_\alpha)} &\le \sup_\alpha \frac{\sup_{f\in\mathscr{F}_\alpha} D\inf_h \mathbf{E}\int |f_{nh} - f| + E\sqrt{(\log n)/n}}{R_n(\mathscr{F}_\alpha)}\\ &\le \sup_\alpha \frac{DC_\alpha R_n(\mathscr{F}_\alpha) + E\sqrt{(\log n)/n}}{R_n(\mathscr{F}_\alpha)}\\ &= D\sup_\alpha C_\alpha + \frac{E\sqrt{(\log n)/n}}{\inf_\alpha R_n(\mathscr{F}_\alpha)}.\end{aligned}$$

In the majority of the interesting cases, this is $D\sup_\alpha C_\alpha + o(1)$. Indeed, then, one may use H and be assured of good adaptive capabilities whenever (1) holds and the constants C_α are uniformly bounded. Typically, (1) is easy to verify, so that one need not be concerned with the details of the random bandwidth H. Furthermore, the universal nature of the above result says something very powerful about the kernel estimate and about the bandwidths described in this paper.

Acknowledgments. The authors would like to thank both referees and an Associate Editor for their help.

REFERENCES

AKAIKE, H. (1954). An approximation to the density function. *Ann. Inst. Statist. Math.* **6** 127–132.

DEVROYE, L. (1982). Bounds for the uniform deviation of empirical measures. *J. Multivariate Anal.* **12** 72–79.

DEVROYE, L. (1987). *A Course in Density Estimation*. Birkhäuser, Boston.

DEVROYE, L. (1988). Asymptotic performance bounds for the kernel estimate. *Ann. Statist.* **16** 1162–1179.

DEVROYE, L. (1997). Universal smoothing factor selection in density estimation: theory and practice. *Test*. To appear.

DEVROYE, L. and GYÖRFI, L. (1985). *Nonparametric Density Estimation. The L_1 View.* Wiley, New York.

DEVROYE, L., GYÖRFI, L. and LUGOSI, G. (1996). *A Probabilistic Theory of Pattern Recognition*. Springer, New York.

DEVROYE, L. and LUGOSI, G. (1996). A universally acceptable smoothing factor for kernel density estimation. *Ann. Statist.* **24** 2499–2512.

PARZEN, E. (1962). On the estimation of a probability density function and the mode. *Ann. Math. Statist.* **33** 1065–1076.

ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837.

VAPNIK, V. N. and CHERVONENKIS, A. YA. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.* **16** 264–280.

YATRACOS, Y. G. (1985). Rates of convergence of minimum distance estimators and Kolmogorov's entropy. *Ann. Statist.* **13** 768–774.

SCHOOL OF COMPUTER SCIENCE
MCGILL UNIVERSITY
MONTREAL
CANADA H3A 2A7
E-MAIL: luc@cs.mcgill.ca

DEPARTMENT OF ECONOMICS
POMPEU FABRA UNIVERSITY
RAMON TRIAS FARGAS, 25-27
08005 BARCELONA
SPAIN