The Annals of Statistics
1984, Vol. 12, No. 4, 1250–1262

DISTRIBUTION-FREE LOWER BOUNDS IN DENSITY ESTIMATION

BY LUC DEVROYE AND CLARK S. PENROD[1]

McGill University and The University of Texas at Austin

We consider the kernel estimate on the real line,

$$f_n(x) = (nh)^{-1} \sum_{i=1}^n K((X_i - x)/h),$$

where K is a bounded even density with compact support, and $X_1, \cdots, X_n$ are independent random variables with common density f. We treat the problem of placing a lower bound on the $L1$ error $J_n = E(\int | f_n - f |)$ which holds for all f. In particular, we show that there exist $A(K) \geq (9/125)^{1/5}$ depending only upon K, and $B^*(f) \geq 1$ depending only upon f such that

(i) for *all* f: $\inf_{h>0} n^{2/5} J_n \geq CA(K)B^*(f) + o(1) \geq 0.6076703\cdots + o(1)$ where $C = 1.028493\cdots$ is a universal constant;

(ii) for all f with compact support and two bounded continuous absolutely integrable derivatives, $\inf_{h>0} n^{2/5} J_n \leq C^*A(K)B^*(f) + o(1)$ where $C^* = 1.3768102\cdots$ is another universal constant. For this class of densities, we also obtain the exact asymptotic behavior of J_n.

1. Introduction. Assume that a density f on R^d is to be estimated from *data* $X_1, \cdots, X_n$ (independent random vectors with common density f). One expects that with a finite amount of data a given estimate has built-in limitations, even for the best densities f. In this paper we derive lower bounds for the L_1 performance of the *kernel estimate*

$$f_n(x) = (nh^d)^{-1} \sum_{i=1}^n K((X_i - x)/h) \tag{1}$$

(Parzen, 1962; Rosenblatt, 1956), where $h = h_n$ is a given sequence of positive numbers, and K is a given density (*kernel*). Throughout we assume that K satisfies:

$$K \geq 0, \quad \int K(x)\, dx = 1, \quad K(x) = K(-x), \quad \text{all } x, \tag{2}$$

K is bounded and has compact support. We are interested in lower bounds for

$$J_n = E\left(\int | f_n(x) - f(x) |\, dx\right)$$

for estimates defined by (1) and (2). The situation where K is not a density is

Received October 1982; revised January 1984.

[1] The research of both authors was supported by the Office of Naval Research under Contract N00014-81-K-0145.

AMS 1980 *subject classifications.* Primary 62G05; secondary 60F15.

Key words and phrases. Density estimation, lower bounds, nonparametric methods, kernel estimate, weak convergence.

not investigated here. The only interesting sequences h satisfy

$$\lim_{n\to\infty} h = 0, \quad \lim_{n\to\infty} nh = \infty \tag{3}$$

in view of the equivalence of (3) and the convergence to 0 of J_n for some f or for all f (see Devroye, 1983a).

Our choice of the L_1 error criterion arises out of several considerations, foremost among them being its scale independence. Rescaling the components of X_1 has no effect on the L_1 error as it does for L_p error $(E^{1/p}(\int |f_n(x) - f(x)|^p \, dx))$ when $p \neq 1$. Hence the L_p error, $p \neq 1$, actually has no absolute relation to the error committed. Thus, distribution-free lower bounds describing the limitations of an estimate cannot be obtained by L_p errors except when $p = 1$, which is the subject of this paper. Other factors influencing our choice of L_1 include the fact that by treating the L_1 error we avoid introducing such unnatural conditions as "f belongs to L_p". Another consideration is that the L_1 error is proportional to what is observed visually when one superimposes the graphs of f_n and f, although such graphs should not be used to corroborate the results of a simulation study since the errors can vary widely.

To set the stage for our main result, we mention first that for all sequences of density estimates $f_n = f_n(x, X_1, \cdots, X_n)$, and for all sequences $a_n \downarrow 0$ however slowly, there exists a density f that is infinitely many times continuously differentiable and another density f^* that is bounded and has compact support, such that $J_n \geq a_n$ infinitely often for both f and f^* (Devroye, 1983b). Results of this type are concerned with the most difficult members in rich enough families of densities. They do not give us information about the actual rate of convergence of most or all densities in these families. See also Boyd and Steele (1978) and Bretagnolle and Huber (1979) for similar results about the L_2 error and the L_p error, respectively.

Our main result is a distribution-free lower bound valid for all densities f on R^1:

THEOREM 1. *For all densities f on R^1, the kernel estimate defined by* (1) *and* (2) *satisfies*

$$\liminf_{n\to\infty} \inf_h n^{2/5} J_n \geq CA(K)B^*(f) \geq CA(K) \geq C_1 > 0 \tag{4}$$

where

$C = 1.028493\cdots$ *is a universal constant*;

$$A(K) = \left(\int K^2\right)^{2/5}\left(\int x^2 K\right)^{1/5}$$ *is a factor only depending upon K*

(*note*: $A(K) \geq (9/125)^{1/5}$);

$$B^*(f) = \left[\frac{1}{2}\left(\int \sqrt{f}\right)^4 \sup_{a>0} \int |(f * \phi_a)''|\right]^{1/5}$$

where ϕ is a bounded density with compact support and two continuous derivatives, *and* $\phi_a = (1/a)\phi(x/a)$ *and* $*$ *is the convolution operator. The constant* C_1 *is* $0.6076703\cdots$. *In particular, when* $f \in \mathscr{F}$, *the class of all densities* f *of compact support, such that* f *and* f' *are absolutely continuous and* f'' *is bounded and continuous, then* $B^*(f) = B(f) = [(1/2)(\int \sqrt{f})^4 \int |f''|]^{1/5}$.

REMARK 1. (*Practical use of* (4)). Theorem 1 gives us information on the best possible performance of the kernel estimate for the nicest densities. It allows us to check at a glance how large n must be to achieve a certain L_1 error rate for very well-behaved densities. In many cases, (4) can be used to point out that "good" L_1 performance is not possible with a given value for n, since in first approximation $J_n \geq CA(K)n^{-2/5}$ for *all* f.

REMARK 2. (*Choice of* K). The factor $A(K)$ is minimized by the Epanechnikov kernel $K(x) = \frac{3}{4}(1 - x^2)$, $|x| \leq 1$ (see Bartlett, 1962, Epanechnikov, 1969 or Tapia and Thompson, 1978).

REMARK 3. (*Difficult densities*). The factor $B^*(f)$ indicates the difficulty posed by f for the kernel estimate. In the profound study of Bretagnolle and Huber (1979) it can also be found as a universal lower bound for the expected L_1 error of any density estimate if one is allowed to choose the "worst" density in a given class of densities. The rate $n^{-2/5}$ is thus not achievable whenever $B^*(f) = \infty$: this happens for one of two reasons: either $\int \sqrt{f} = \infty$ (which indicates the presence of a large tail, e.g. the Cauchy density falls into this category), or $\sup_{a>0} \int |(f*\phi_a)''| = \infty$ (which indicates the presence of at least one simple discontinuity, e.g. the uniform [0, 1] density and the exponential density fall into this category).

REMARK 4. (*Simulation of* f). In a computer simulation we could use f_n (instead of the unknown f) to generate an independent sample of size n_0. The efficiency of the replacement of f by f_n can be measured by

$$\sup_B \left| \int_B f_n(x)\,dx - \int_B f(x)\,dx \right| = \frac{1}{2}\int |f_n(x) - f(x)|\,dx.$$

Consider *accurate simulations* with sample size n_0 in which the average number of data points in each Borel set under f_n is within 1 of that under f. This is violated when $n_0J_n/2$ is of the order of magnitude of 1 or bigger. As a rule of thumb, we are "safe" in simulations of size

$$n_0 \leq 2/J_n.$$

But, for all f we have in first approximation $J_n \geq CA(K)/n^{2/5}$, which leads to the limitation

$$n_0 \leq 2n^{2/5}/(CA(K)) \simeq 3.29n^{2/5}$$

in the best possible case! The bound on n_0 is smaller when f is not well-behaved. Thus, to increase the simulation sample size n_0 by a factor of 100, it is necessary to increase the original sample size n by a factor of $100^{5/2} = 100{,}000$. For $n_0 = 1000$, we will need $n \geq 1{,}600{,}000$. This indicates the limitations of the use of nonparametric estimates f_n of the type discussed here for the purposes of computer simulation in all but a few situations, i.e. when n is gigantic. The situation for $d > 1$ is probably much worse.

THEOREM 2. (Exact asymptotic behavior of J_n). *For all f in $\mathscr{F}$, the kernel estimate defined by* (1), (2), (3) *satisfies*

$$J_n = J(n, h) + o(h^2 + 1/\sqrt{nh})$$

where

$$J(n, h) = \int \frac{\alpha\sqrt{f}}{\sqrt{nh}} \psi\left(\sqrt{nh^5}\,\frac{\beta | f''|}{2\alpha\sqrt{f}}\right), \quad \alpha = \sqrt{\textstyle\int K^2}, \quad \beta = \int x^2K(x)\,dx$$

and

$$\psi(u) = \sqrt{2/\pi}\left(u \int_0^u e^{-x^2/2}\,dx = e^{-u^2/2}\right), \quad u \geq 0.$$

Also,

$$J(n, h) \leq \sqrt{2/\pi}\,\frac{\alpha \int \sqrt{f}}{\sqrt{nh}} + \frac{\beta}{2} h^2 \int | f'' |$$

and thus

$$\limsup_{n\to\infty} \inf_{h>0} n^{2/5} E(J_n) \leq C^* A(K) B^*(f) \tag{5}$$

where

$$C^* = 5(8\pi)^{2/5} = 1.3768102\cdots.$$

The upper bound is not exceeded for the following choice of h when $f \in \mathscr{F}$, $B^*(f) < \infty$:

$$h = \left[\frac{\alpha}{2\beta}\,\frac{\int \sqrt{f}}{\sup_{a>0} \int | (f^*\phi_a)'' |}\,\sqrt{2/\pi}\right]^{2/5} n^{-1/5}.$$

REMARK 5. (*Sharpness of the bounds.*). Upper bound (5) is reasonably sharp since C^* is about 35% bigger than the constant C of (4). Rosenblatt (1979) obtained an inequality in the spirit of (5) with a slightly larger constant $C' = (5/2)^{9/5} = 1.435872\cdots$, under regularity conditions not nested with those of Theorem 2. Devroye and Györfi proved that (5) remains valid for *all* f with compact support (Devroye and Györfi, 1984).

REMARK 6. (*Choice of h*). The formula obtained for h in Theorem 2 can be used to construct a good data-dependent smoothing factor, at least when f is smooth enough and does not have too large tails. Notice that the optimal h

obtained via the L_2 theory is drastically different, e.g. it does not depend upon the "tail size" $\int \sqrt{f}$.

2. Proofs.

LEMMA 1. (See, e.g. de Guzman, 1981). *Let g be an absolutely integrable function on R^d, and let K be an arbitrary density on R^d. Then*

(i) $\lim_{h\downarrow 0} \int |g^*K_h - g| = 0$;
(ii) $\int |g^*K_h| \le \int |g|$;
(iii) *If K is bounded and has compact support, then $g^*K_h \to g$ for almost all x.*

LEMMA 2. *Let $X_1, \cdots, X_n$ be independent random variables with a common distribution. Let $E(X_1) = 0$, $E(X_1^2) = \sigma^2 > 0$, $\rho = E(|X_1|^3) < \infty$. Then,*

$$\sup_{a\in R} |E(|(\sigma\sqrt{n})^{-1} \textstyle\sum_{i=1}^n X_i - a|) - E(|N - a|)| \le c\rho\sigma^{-3}/\sqrt{n}$$

where c is a universal positive constant and N is a normal (0, 1) *random variable. In particular,*

$$E(|N - a|) = |a| P(|N| \le |a|) + \sqrt{2/\pi}\, e^{-a^2/2} = \psi(|a|).$$

NOTE. Since

$$\int_u^\infty e^{-x^2/2}\, dx \le \frac{1}{u} e^{-u^2/2},$$

we have $\psi(u) \ge u$. Also, by inspection, $\psi(u) \ge \sqrt{2/\pi}$. Furthermore,

$$\psi(u) \le u + \sqrt{2/\pi}, \quad \psi'(u) = \sqrt{2/\pi} \int_0^u e^{-x^2/2}\, dx \ge 0, \quad \text{and} \quad \psi''(u) \ge 0.$$

Thus, ψ is monotone, convex, and varies as $\sqrt{2/\pi}$ near $u = 0$ and as u when $u \to \infty$.

PROOF. Let F_n be the distribution function of $X = (\sigma\sqrt{n})^{-1} \sum_{i=1}^n X_i$, and let Φ be the distribution function of N. Clearly,

$$E(|X - a|) = \int_0^\infty P(|X - a| > t)\, dt = \int_0^\infty (1 - F_n(a + t) + F_n(a - t))\, dt,$$

and a similar equation is valid for N and Φ. The absolute value of the difference between both equations does not exceed

$$\int_0^\infty |\Phi(a + t) - F_n(a + t)|\, dt + \int_0^\infty |\Phi(a - t) - F_n(a - t)|\, dt$$
$$= \int_{-\infty}^\infty |\Phi(t) - F_n(t)|\, dt.$$

By well-known nonuniform estimates in the Berry–Esseen type central limit theorem (see Petrov, 1975, Theorem 14, page 125),

$$|\Phi(t) - F_n(t)| \le c\rho\sigma^{-3}/((1 + |t|^3)\sqrt{n})$$

for some universal constant c. Since $(1 + |t|^3)^{-1}$ is integrable, we obtain the desired result.

For the expression of $E(|N - a|)$, we note that for $a > 0$,

$$\begin{aligned} E(|N-a|) &= E(|N|) + E(|N-a| - |N|) \\ &= E(|N|) + aP(N<0) + E((a-2N)I_{[0<N<a]}) - aP(N>a) \\ &= E(|N|) + a - 2E(NI_{[0<N<a]}) - 2aP(N>a) \\ &= \sqrt{2/\pi} + a - aP(|N|>a) - 2\int_0^a \frac{te^{-t^2/2}}{\sqrt{2\pi}}\,dt \\ &= \sqrt{2/\pi} + aP(|N|<a) - \frac{2(1-e^{-a^2/2})}{\sqrt{2\pi}}, \end{aligned}$$

which was to be shown.

In the remainder of this section, T is an arbitrary interval, $[-r, r]$ is the support of K, K^* is an upper bound for K, and T^* is defined as $\{x: |x - y| \le hr$ for some $y \in T\}$. Thus, T^* depends upon h. Also, c is the constant of Lemma 2, $B_n(x) = E(f_n(x)) - f(x)$ is the *bias* at x, $V_n(x) = f_n(x) - E(f_n(x))$ is the *variation* at x, and $\sigma_n^2(x) = E(V_n^2(x))$ is the *variance* at x. We define C by $\inf_{u>0}\psi(u)/u^{1/5}$.

LEMMA 3.

$$|E(|f_n(x) - f(x)|) - \sigma_n(x)\psi(|B_n(x)|/\sigma_n(x))| \le cK^*/nh,$$

for all densities K satisfying (2).

PROOF OF LEMMA 3. Apply Lemma 2 to the random variables

$$Y_i = (1/h)K((X_i - x)/h) - E((1/h)K((X_i - x)/h)),$$

and use $a = B_n(x)/\sigma_n(x)$. We obtain an error term in Lemma 2 of the form

$$c\,\frac{E(|Y_1|^3)}{nE(Y_1^2)} \le \frac{cK^*}{nh}.$$

LEMMA 4. *Let T be a bounded interval. Then, for all $h > 0$ and all densities K,*

$$-\frac{\sqrt{h}}{\alpha} - \sqrt{\textstyle\int |f^*K_h^\dagger - f|\,\lambda(T)} \le \frac{\sqrt{nh}}{\alpha}\int_T \sigma_n - \int_T \sqrt{f} \le \sqrt{\textstyle\int |f^*K_h^\dagger - f|\,\lambda(T)}$$

where $K^\dagger = K^2/\int K^2$. Also, $\int_T |\sigma_n(\sqrt{nh}/\alpha) - \sqrt{f}| = o(1)$ as $h \downarrow 0$.

PROOF OF LEMMA 4. We first note that for bounded sets, we always have $\int_T \sqrt{f} < \infty$. We have $\sigma_n^2(x) = a(x) + b(x)$ where

$$a(x) = \alpha^2 f(x)/(nh)$$

and

$$b(x) = (f^*K_h^\dagger - f)\alpha^2/(nh) - (f^*K_h)^2/n.$$

Clearly, $a(x) \geq 0$. Thus, $\sqrt{a(x) + b(x)} \leq \sqrt{a(x)} + \sqrt{b_+(x)}$, and $\sqrt{a(x) + b(x)} \geq \sqrt{a(x)} - \sqrt{|\, b(x)\,|}$. Integrating over T and applying the Cauchy–Schwarz inequality gives

$$\begin{aligned}\int_T \sigma_n &\leq \frac{\alpha}{\sqrt{nh}}\left(\int_T \sqrt{f} + \int_T \sqrt{|\, f^*K_h^\dagger - f\,|}\right)\\ &\leq \frac{\alpha}{\sqrt{nh}}\left(\int_T \sqrt{f} + \sqrt{\int |\, f^*K_h^\dagger - f\,|\,\lambda(T)}\right),\end{aligned}$$

and

$$\int_T \sigma_n \geq \frac{\alpha}{\sqrt{nh}}\left(\int_T \sqrt{f} - \sqrt{\int |\, f^*K_h^\dagger - f\,|\,\lambda(T)} - \frac{\sqrt{h}}{\alpha}\int f^*K_h\right).$$

The first half of Lemma 4 follows easily from this. The last statement of Lemma 4 follows if $\int_T \sqrt{|\, b(x)\,|} = o(1)$. But this is a consequence of $h = o(1)$, $\lambda(T) < \infty$ and $\int |\, f^*K_h^\dagger - f\,| = o(1)$ (see Lemma 1).

LEMMA 5. *Let f be a density in $\mathscr{F}$, let K satisfy* (2), *and let* $\lim_{n\to\infty} h = 0$. *Then the quantity*

$$q_n(x) = |\, B_n(x) - (\beta/2)h^2 f''(x)\,|$$

satisfies $q_n(x) = o(h^2)$ for all x, and $\int q_n(x)\,dx = o(h^2)$.

PROOF. Taylor's expansion with remainder gives

$$\begin{aligned}f(y) = f(x) &+ (y - x)f'(x) + \frac{(y-x)^2}{2} f''(x)\\ &+ \left[\int_x^y (y - z)f''(z)\,dz - \frac{(y-x)^2}{2} f''(x)\right].\end{aligned}$$

Thus,

$$\begin{aligned}&E(f_n(x)) - f(x)\\ &\quad= \int \frac{1}{h} K\left(\frac{y-x}{h}\right)(f(y) - f(x))\,dy\\ &\quad= \frac{h^2}{2} f''(x) \int y^2 K(y)\,dy \qquad (6)\\ &\quad\quad+ \int \frac{1}{h} K\left(\frac{y-x}{h}\right)\left[\int_x^y (y - z)f''(z)\,dz - \frac{(y-x)^2}{2} f''(x)\right] dy.\end{aligned}$$

Clearly, $q_n(x)$ is equal to the absolute value of the last term of (6). But $q_n(x)/h^2$ does not exceed

$$\int \frac{1}{h} K\left(\frac{y-x}{h}\right)\left(\frac{y-x}{h}\right)^2 \sup_{|y-x|\le ch} |f''(y) - f''(x)|\, dy = \beta o(1).$$

Thus, $\int q_n(x)\,dx = o(h^2)$ if we can find an integrable function dominating $q_n(x)/h^2$. But clearly, $q_n(x)/h^2 \le \beta \sup_x |f''(x)|$ on T^*, and $q_n(x)/h^2 = 0$ on T^{*c}, and $\lambda(T^*) = O(\lambda(T)) < \infty$, which concludes the proof of Lemma 5. □

NOTE. If f does not have compact support, then $\int_T q_n(x)\,dx = o(h^2)$ for all bounded T.

LEMMA 6. *For all $f \in \mathscr{F}$, $B(f) = B^*(f)$, and $B(f) \ge 1$. Also, for all f, $B^*(f) \ge 1$.*

PROOF. We start by noting that $(f^*\phi_h)'' = f''^*\phi_h$, and that $\int |f''^*\phi_h| \le \int |f''|$ (which shows that $B(f) \ge B^*(f)$). Also, $B(f) \le B^*(f)$ because the integrability of f'' implies that $\int |f''^*\phi_h| \to \int |f''|$ as $h \downarrow 0$ (see Lemma 1).

To prove that $B(f) \ge 1$ for $f \in \mathscr{F}$, we note that $f'(y) - f'(x) = \int_x^y f''(z)\,dz$. Thus, using $(\ )_+$ and $(\ )_-$ for the positive and negative parts of a function, we have

$$\int_{-\infty}^{+\infty} (f''(y))_-\,dy \le f'(x) \le \int_{-\infty}^{+\infty} (f''(y))_+\,dy, \quad \text{all} \quad x,$$

and

$$\int_{-\infty}^{+\infty} (f''(y))_+\,dy + \int_{-\infty}^{+\infty} (f''(y))_-\,dy = 0,$$

so that we may conclude that

$$\sup|f'(x)| \le \frac{1}{2}\int_{-\infty}^{+\infty} |f''(y)|\,dy.$$

But we also have $1 = \int f \le \sqrt{\sup f} \int \sqrt{f}$. Combining these inequalities shows that $B(f)^5 \ge \sup|f'(x)|/\sup f^2(x)$. Clearly, by a geometrical argument,

$$1 = \int f(x)\,dx \ge \int (\sup f(x) - |y| \sup|f'(x)|)_+\,dy = \frac{\sup f^2(x)}{\sup|f'(x)|},$$

and the proof of $B(f) \ge 1$ is complete.

To prove that $B^*(f) \ge 1$ for all f, we can assume without loss of generality that $\int \sqrt{f} < \infty$. Then, observe the following:

A. $$\int \sqrt{f} = \int \sqrt{f}^*\phi_a; \quad \int (\sqrt{f}^*\phi_a) \cdot \sup(\sqrt{f}^*\phi_a) \ge \int (\sqrt{f}^*\phi_a)^2;$$

and, by Fatou's Lemma and Lemma 1,

$$\liminf_{a\downarrow 0} \int (\sqrt{f}*\phi_a)^2 \geq \int \liminf_{a\downarrow 0}(\sqrt{f}*\phi_a)^2 = \int f = 1.$$

B. Because $f*\phi_a$ has two bounded continuous derivatives, has a Lipschitz first derivative and an integrable second derivative,

$$\sup|(f*\phi_a)'| \leq \int |(f*\phi_a)''|/2.$$

C. $1 \geq \sup(f*\phi_a)^2/\sup|(f*\phi_a)'|$.

Combining A, B and C gives for all $a > 0$,

$$B^*(f) \geq \left[\int (\sqrt{f}*\phi_a)^2\right]^4 \frac{\sup(f*\phi_a)^2}{\sup(\sqrt{f}*\phi_a)^4}.$$

By Jensen's inequality, $(\sqrt{f}*\phi_a)^2 \leq f*\phi_a$. Thus,

$$B^*(f) \geq \liminf_{a\downarrow 0}\left[\int (\sqrt{f}*\phi_a)^2\right] = 1. \quad \square$$

PROOF OF THEOREM 1. We have

$$\inf_h E(J_n) \geq \min(\inf_{h\sqrt{n}\leq 1} E(J_n), \inf_{h\sqrt{n}\geq 1} E(J_n)). \tag{7}$$

Consider a sequence h such that $E(J_n) \sim \inf_h E(J_n)$. It is clear that $E(J_n) \to 0$ for all f, because (3) is sufficient for $E(J_n) \to 0$ (Devroye, 1983a). But because $E(J_n) \geq \int |f*K_h - f|$, we must have $h \to 0$ (Devroye, 1983a). We will now treat each infimum in (7) separately.

First, if h is such that $h \geq 1/\sqrt{n}$, all n, and $E(J_n) \sim \inf_{h\sqrt{n}\geq 1} E(J_n)$, then by what we mentioned above, $h \to 0$. Also, $nh \to \infty$, and in fact, $nh/n^{2/5} \to \infty$. Now, let T be a bounded interval, and $a > 0$ be an arbitrary constant. We have for such h the following lower bound for $E(J_n)$:

$$\int_T E(|f_n - f|) \geq \int_T \sigma_n \cdot \psi\left(\frac{\int_T |B_n|}{\int_T \sigma_n}\right) - \frac{cK^*}{nh}\lambda(T)$$

(By Lemma 3, the convexity of ψ and Jensen's inequality)

$$\geq C\left(\int_T \sigma_n\right)^{4/5}\left(\int_T |B_n|\right)^{1/5} - o(n^{-2/5}) \quad \text{(Definition of } C\text{)}$$

$$\sim Cn^{-2/5}\left(\alpha \int_T \sqrt{f}\right)^{4/5}\left(\frac{\beta}{2}\int_T |(f*\phi_a)''|\right)^{1/5} \quad \text{(see below)} \tag{8}$$

$$= n^{-2/5}CA(K)\left[\frac{1}{2}\left(\int_T \sqrt{f}\right)^4 \int_T |(f*\phi_a)''|\right]^{1/5}$$

(Definition of $A(K)$).

Here we used the definition of $C = \inf_{u>0}\psi(u)/u^{1/5}$, and the fact that

$$\int_T |B_n| = \int_T |f^*K_h - f| \ge \int |f^*\phi_a{}^*K_h - f^*\phi_a| \quad \text{(all } a > 0\text{, Lemma 1)}$$

$$\ge \frac{h^2}{2}\beta \int_T |(f^*\phi_a)''|(1 + o(1)) \quad \text{(Lemma 5).}$$

Next, let h be a sequence such that $h \le 1/\sqrt{n}$ for all n, and $E(J_n) \sim \inf_{h\sqrt{n}\le 1} E(J_n)$. From Devroye (1983a), we have $nh \to \infty$. Also,

$$E(J_n) \ge \frac{1}{2} E\left(\int |f_n - f^*K_h|\right).$$

By Fatou's lemma,

$$\lim\inf_{n\to\infty} n^{2/5} E(J_n) \ge \int \frac{1}{2} \lim\inf_{n\to\infty} n^{2/5} E(|f_n - f^*K_h|),$$

and the right-hand side of this is ∞ when for almost all x with $f(x) > 0$, $\lim\inf_{n\to\infty} n^{2/5}E(|f_n - f^*K_h|) = \infty$. To show this, we will use the Berry–Esseen central limit theorem used in Lemma 2. Let $\tilde\sigma_n^2(x) = \operatorname{Var}(K_h(X_1 - x))$ and let M be an arbitrarily large positive number. Let Z be a normal (0, 1) random variable. Then,

$$\begin{aligned} &n^{2/5}E(|f_n - f^*K_h|) \\ &\ge MP\left(|f_n - f^*K_h| \ge \frac{M}{n^{2/5}}\right) = MP\left(|f_n - f^*K_h| \frac{\sqrt{n}}{\tilde\sigma_n} \ge \frac{m\sqrt{n}}{\tilde\sigma_n n^{2/5}}\right) \\ &\ge M\left(P\left(|Z| \ge \frac{Mn^{1/10}}{\tilde\sigma_n}\right) - 2c\tilde\sigma_n^{-3} n^{-1/2} E(|K_h(X_1 - x) - E(K_h(X_1 - x))|^3)\right). \end{aligned}$$

By inspection of the proof of Lemma 4 and by Lemma 1, it is easy to see that $\tilde\sigma_n^2(x) \sim \alpha^2 f(x)/h$ for almost all x, as $h \to 0$. Also, by the c_r-inequality and Lemma 1,

$$\begin{aligned} &E(|K_h(X_1 - x) - E(K_h(X_1 - x))|^3) \\ &\le 4E\left(h^{-3}K^3\left(\frac{X_1 - x}{h}\right)\right) + 4(E(K_h(X_1 - x)))^3 = 4h^{-2}f^*(K^3)_h + 4(f^*K_h)^3 \\ &\sim 4h^{-2}f(x)\int K^3 + 4f(x)^3 \sim 4h^{-2}f(x)\int K^3, \quad \text{almost all } x. \end{aligned}$$

Because $n^{1/10}/\tilde\sigma_n(x) \sim n^{1/10}\sqrt{h}/\alpha\sqrt{f(x)} \le 1/(n^{3/20}\alpha\sqrt{f(x)}) \to 0$ for almost all x with $f(x) > 0$, we have

$$n^{2/5}E(|f_n - f^*K_h|) \ge M\left(1 - (2c + o(1))\frac{4\int K^3\sqrt{f(x)}}{\sqrt{nh}\alpha^3}\right)$$

$$= M(1 - o(1)), \quad \text{almost all } x \text{ with } f(x) > 0.$$

Since M was arbitrary, we have shown that

$$\lim\,\inf_{n\to\infty}\inf_{h\sqrt{n}\le 1} n^{2/5}E(J_n) = \infty$$

and this, together with (8), the definition of $B^*(f)$ and the monotone convergence theorem implies

$$\begin{aligned}\lim\,\inf_{n\to\infty}&\inf_h n^{2/5}E(J_n)\\ &\ge \sup_{a>0,\,\text{bounded}\,T} CA(K)\left(\int_T \sqrt{f}\right)^{4/5}\left(\int_T |(f^*\phi_a)''|\right)^{1/5} \Big/ 2^{1/5}\\ &= CA(K)B^*(f).\end{aligned}$$

This concludes the proof of Theorem 1. □

We need one last technical lemma before we can attack Theorem 2.

LEMMA 7. *For nonnegative numbers u, v, w, z, we have*

$$\left| u\psi\left(\frac{v}{u}\right) - w\psi\left(\frac{z}{w}\right)\right| \le |v - z| + \sqrt{2/\pi}\,|u - w|.$$

PROOF OF LEMMA 7. We verify first that $0 \le \psi'(u) \le 1$, all $u \ge 0$, and that for all $v \ge 0$, $|(u\psi(v/u))'| \le \sqrt{2/\pi}$. Thus,

$$\begin{aligned}\left| u\psi\left(\frac{v}{u}\right) - w\psi\left(\frac{z}{w}\right)\right| &\le \left| u\psi\left(\frac{v}{u}\right) - u\psi\left(\frac{z}{u}\right)\right| + \left| u\psi\left(\frac{z}{u}\right) - w\psi\left(\frac{z}{w}\right)\right|\\ &\le |v - z| + \sqrt{2/\pi}\,|u - w|. \quad \square\end{aligned}$$

PROOF OF THEOREM 2. Theorem 2 has several components. First, we assume that $f \in \mathscr{F}$ and that f has compact support contained in a bounded interval T. Take T so large that for every x in the support of f, the interval $[x - a, x + a]$ is contained in T, where a is a number sufficiently large so that $K_h(u) = 0$ for all n and all $|u| > a$.

We will begin with the inequality of Lemma 7 applied in the following manner:

$$u := \sigma_n(x);\quad v := |B_n(x)|;\quad w := \frac{\alpha\sqrt{f(x)}}{\sqrt{nh}};\quad z := \frac{\beta}{2}h^2|f''(x)|.$$

Now,

$$\int_T |v - z| = o(h^2) \quad \text{(Lemma 5)}$$

and

$$\int_T |u - w| = o((nh)^{-1/2}) \quad \text{(Lemma 4).}$$

Thus, combining this into the inequality of Lemma 3 gives

$$\left| \int_T E(|f_n - f|) - J(n, h) \right|$$
$$\le \left| \int_T E(|f_n - f|) - \int \sigma_n \psi\left(\frac{|B_n|}{\sigma_n}\right) \right| + \left| \int \sigma_n \psi\left(\frac{|B_n|}{\sigma_n}\right) - J(n, h) \right|$$
$$\le \frac{cK^*}{nh} \lambda(T) + o(h^2) + o((nh)^{-1/2}),$$

where we used the fact that $J(n, h) = \int_T w\psi(z/w)$.

The inequality involving $J(n, h)$ follows from $\psi(u) \le u + \sqrt{2/\pi}$:

$$J(n, h) = \int_T w\psi\left(\frac{z}{w}\right) \le \int_Z z + \sqrt{2/\pi} \int_T w.$$

Let us turn now to all densities f having compact support and let us denote the quantity $\sup_{a>0} \int |(f*\phi_a)''|$ appearing in the definition of $B^*(f)$ by L. Again, from Lemma 3 and the inequality $\psi(u) \le u + \sqrt{2/\pi}$, we obtain

$$\int_T E(|f_n - f|) \le \int_T \left(\sqrt{2/\pi}\, \sigma_n + |B_n|\right) + \frac{cK^*}{nh} \lambda(T),$$

and by Lemmas 4 and 5, this is further bounded from above by

$$\sqrt{2/\pi}\, \frac{\alpha}{\sqrt{nh}} \int \sqrt{f} + \sqrt{2/\pi}\, \frac{\alpha}{\sqrt{nh}} \sqrt{\int |f*K_h^\dagger - f|\, \lambda(T)} + \frac{\beta}{2} h^2 L + \frac{cK^*}{nh} \lambda(T)$$

where $K^\dagger$ is the density defined in Lemma 4. The second term is $o((nh)^{-1/2})$ where $h = o(1)$ (Lemma 1). The last term is $o((nh)^{-1/2})$ when $nh \to \infty$. This proves the first upper bound for general f. If we take the value of h given in the statement of the Theorem (i.e., the value that minimizes the main term in the upper bound), then

$$\sqrt{2/\pi}\, \frac{\alpha}{\sqrt{nh}} \int \sqrt{f} + \frac{\beta}{2} h^2 L = \frac{C^* A(K) B^*(f)}{n^{2/5}},$$

and this concludes the proof of Theorem 2. □

REFERENCES

BARTLETT, M. S. (1963). Statistical estimation of density functions. *Sankhya A* **25** 245–254.

BOYD, D. W. and STEELE, J. M. (1978). Lower bounds for nonparametric density estimation rates. *Ann. Statist.* **6** 932–934.

BRETAGNOLLE, J. and HUBER, C. (1979). Estimation des densités: risque minimax. *Z. Wahrsch. verw. Gebiete* **47** 119–137.

DE GUZMAN, M. (1981). *Real Variable Methods in Fourier Analysis.* North Holland, Amsterdam.

DEHEUVELS, P. (1977). Estimation non paramétrique de la densité par histogrammes généralisés. *Revue de Statistique Appliquée* **25** 5–42.

DEVROYE, L. (1983a). The equivalence of weak, strong and complete convergence in $L1$ for kernel density estimates. *Ann. Statist.* **11** 896–904.

DEVROYE, L. (1983b). On arbitrarily slow rates of global convergence in density estimation. *Z. Wahrsch. verw. Gebiete* **62** 475–483.

DEVROYE, L. and GYÖRFI, L. (1984). *Nonparametric Density Estimation: The L_1 View.* Wiley, New York.

EPANECHNIKOV, V. A. (1969). Nonparametric estimates of a multivariate probability density. *Theory Probab. Appl.* **14** 153–158.

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

PETROV, V. V. (1975). *Sums of Independent Random Variables.* Springer-Verlag, Berlin.

ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–835.

ROSENBLATT, M. (1971). Curve estimates. *Ann. Math. Statist.* **42** 1815–1842.

ROSENBLATT, M. (1979). Global measures of deviation for kernel and nearest neighbor density estimates. In Smoothing techniques for curve estimation. Th. Gasser and M. Rosenblatt Eds. *Lecture Notes in Math.* **757** 181–190. Springer-Verlag, Berlin.

TAPIA, R. A. and THOMPSON, J. R. (1978). *Nonparametric Probability Density Estimation.* The Johns Hopkins University Press, Baltimore.

SCHOOL OF COMPUTER SCIENCE
MCGILL UNIVERSITY
805 SHERBROOKE ST. W.
MONTREAL, CANADA H3A 2K6

APPLIED RESEARCH LABORATORIES
THE UNIVERSITY OF TEXAS AT AUSTIN
P.O. BOX 8029
AUSTIN, TEXAS 78712