ASYMPTOTIC PERFORMANCE BOUNDS FOR THE KERNEL ESTIMATE

Luc Devroye
School of Computer Science
McGill University

ABSTRACT. We consider an arbitrary sequence of kernel density estimates f_n with kernels K_n possibly depending upon n. Under a mild restriction on the sequence K_n, we obtain inequalities of the type

$$\mathbf{E}\left\{\int |f_n - f|\right\} \geq (1+o(1))\Psi(n,f)$$

where f is the density being estimated and $\Psi(n,f)$ is a function of n and f only. The function Ψ can be considered as an indicator of the difficulty of estimating f with any kernel estimate.

KEY WORDS AND PHRASES. Density estimation, L_1 error, inequalities, characteristic function, kernel estimate, performance bounds.

1980 AMS SUBJECT CLASSIFICATIONS. 60E15, 62G05.

Authors' address: School of Computer Science, McGill University, 3480 University Street, Montreal, Canada H3A 2K6. The authors' research was sponsored by NSERC Grant A3456 and FCAR Grant 90-ER-0291.

1. Introduction.

In this paper, we consider the kernel estimate

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_n(x - X_i)$$

where $\{K_n\}$ is a sequence of absolutely integrable functions (kernels) integrating to one, and $X_1, \ldots, X_n$ are iid random variables with common density f on the real line (Rosenblatt, 1956; Parzen 1962). The expected L_1 error $\mathbf{E}\left\{\int |f_n - f|\right\}$ is a function of n,f and K_n. Of these factors, the user can only choose K_n, the kernel. We note in passing that the most popular form for K_n is $K_n(x) \equiv (1/h)L(x/h)$ for some fixed function L integrating to one, and a scale factor $h > 0$ depending upon n only. In this case, we will call f_n the **standard kernel estimate**. We do not allow h to depend upon the data in this paper. In general, the shape of K_n can vary with n, and in this general setting the estimate is known as the **delta function estimate** (Walter and Blum, 1979).

The arguments for focusing on the L_1 error are expounded in Devroye and Györfi (1985) and Devroye (1987). Since this error is equal to twice the total variation distance between the probability measures induced by f and f_n, the L_1 error provides us with absolute numbers with a clear physical (and even graphical) interpretation. Lower bounds for the L_1 error allow us to draw direct conclusions about minimal sample sizes below which we are bound to have errors that are at least as big as a given value. Since the density f is not known beforehand, it seems useful to have bounds that do not depend upon f such as the following bound:

$$\inf_{K_n, f} \mathbf{E}\left\{\int |f_n - f|\right\} \geq \frac{1}{\sqrt{528n}}$$

(Devroye, 1986). This result implies that even if we are allowed to choose f and K_n, we can't possibly have an expected error that is smaller than $1/\sqrt{528n}$. In other words, the lower bound is the price we have to pay for the use of the kernel estimate. This result could be used to determine if n is large enough for someone to be able to use the kernel estimate. Another example of a result in this spirit concerns the standard kernel estimate with even bounded compact support nonnegative kernel L:

$$\inf_{L,f} \liminf_{n\to\infty} \inf_{h>0} n^{2/5} E(\int |f_n - f|) \geq 0.86$$

(Devroye and Penrod, 1984; Devroye and Györfi, 1985). If one wants to obtain a rate better than $n^{-2/5}$, it is absolutely necessary to drop one or more constraints on L, such as the nonnegativity. Both bounds may be grossly inadequate for some densities, since they provide information about the error we are bound to make for the "best" densities. Thus, it is also of interest to have lower bounds that do depend upon f. Such bounds could be used to discard the kernel estimate altogether for some densities. They also provide information about the limitations that come with the density f when a certain class of estimates (such as the standard kernel estimates) is used. They measure the "difficulty" associated with the estimation of f. In this paper, we consider just such lower bounds for

$$\inf_{K_n} \mathbf{E}\left\{\int |f_n - f|\right\}.$$

These bounds depend upon n and f only, and are in the spirit of a celebrated L_2 lower bound obtained by Watson and Leadbetter (1963) (see also Davis (1975, 1977)). As is well-known, L_2

errors vary with rescalings of the coordinate axis, and are thus not absolute numbers that can be used to compare performances of different estimates on different densities in a straightforward manner. In particular, the infimum over all f of the Watson-Leadbetter lower bound is zero. Nevertheless, from a lower bound, it should be possible (by examination) to design a specific estimate for which we come close to the lower bound for all (or at least many) densities f. In the L_2 setting, this undertaking was carried out successfully by Davis (1975, 1977), who argued that the kernel $L(x) = \sin(x)/\pi x$ with a carefully picked scale factor $h = h(n, f)$ is asymptotically optimal to within a constant factor for all densities. Some have discarded this kernel as unpractical, often complaining about its massive tails (see e.g. Tapia and Thompson (1978, p. 79)). The massive tails of L contribute very little to the L_2 error since squaring tails tends to obscure them. The L_1 theory is much more sensitive to the tails of both f and the kernel, but is also much more of a challenge since the techniques used by Watson and Leadbetter that were based upon Parseval's identity are no longer applicable. The issue of finding a universally nearly-optimal L_1 kernel is not addressed here, although we will briefly mention a kernel that is nearly optimal for all very smooth densities.

In the second half of the paper, we restrict the class of kernels to $K_n \in \mathcal{K}_s$, where $\mathcal{K}_s$ is a saturation class such as the class of all kernels of the form $(1/h)L(x/h)$ for symmetric L where all the moments of L up to and not including the s-th moment are zero. The class of estimates constructed in this manner is smaller than the class considered in the first half of the paper, and the lower bounds should thus be larger. We obtain lower bounds for

$$\inf_{K_n \in \mathcal{K}_s} \mathbf{E}\left\{ \int |f_n - f| \right\}$$

and

$$\inf_{f, K_n \in \mathcal{K}_s} \mathbf{E}\left\{ \int |f_n - f| \right\}$$

that decrease as a constant times $n^{-s/(2s+1)}$.

2. The main results.

We will relate the lower bound on the expected L_1 error for any sequence of kernel estimates to the shape of f. Since Bessel's equality used by Watson and Leadbetter in relating the best possible L_2 performance of any kernel estimate to the characteristic function of a density is not directly useful here, we have to resort to different methods. The strategy followed here is to bound the L_1 bias from below by a supremum norm for characteristic functions, and to bound the variational component in the L_1 error directly, i.e., without going through characteristic functions. As a consequence, the bounds depend upon functionals of f and its characteristic function.

The simplest treatment is the one in which we consider a sequence of consistent kernel estimates, and provide lower bounds that are $(1 + o(1))$ times an explicit function of n and f. We will restrict the sequence of kernels K_n as follows.

A. There exists a constant M such that $\int |K_n| \leq M$ for all n. Also, $\int K_n = 1$ for all n, and $\int_{|x|>\delta} |K_n| \to 0$ as $n \to \infty$ for all $\delta > 0$. (Condition A implies that $\int |K_n * f - f| \to 0$ for all $f \in L_1$.)
B. $\int |K_n|^i < \infty$ for all n, and for $1 \leq i \leq 4$.
C. K_n, $K_n^2 / \int K_n^2$ and $K_n^4 / \int K_n^4$ are strong approximate identities (a sequence of kernels g_n is a strong approximate identity if for all $f \in L_1$, $g_n * f \to f$ at almost all x).

D. $\int K_n^4 / n \int^2 K_n^2 \to 0$ as $n \to \infty$.

For the standard kernel estimate with $K_n(x) = (1/h)L(x/h)$, conditions A-D are implied by

E. $\int L = 1$, $|L| \le g$ for some symmetric unimodal nonnegative function g with $\int g^i < \infty$ for $1 \le i \le 4$, and $h \to 0$ and $nh \to \infty$ as $n \to \infty$.

The sufficiency of condition E for A-D can easily be verified. Clearly, A-B hold. Condition D is satisfied when $nh \to \infty$. Finally, condition C holds if $h \to 0$ in view of some pointwise consistency results found for example in Butzer and Nessel (1971, pp. 132-135). The conditions $h \to 0, nh \to \infty$ are easily recognized as the necessary and sufficient conditions for L_1 convergence of f_n to f (see Devroye, 1983). The conditions on L are satisfied for virtually all kernels mentioned in the literature on density estimation.

A sequence $\{K_n\}$ satisfying A-D is said to be a **regular sequence** (for lack of a better term). The first result is captured in

THEOREM 1. *Let $\{K_n\}$ be a regular sequence of kernels, and let f_n be the corresponding kernel estimate. If $\int \sqrt{f} < \infty$, then*

$$\mathbf{E}\left\{\int |f - f_n|\right\} \ge (1 + o(1)) \inf_{0 \le u \le 1} \max\left(u, \frac{\int \sqrt{f}\sqrt{\Phi(2u)}}{8\sqrt{2\pi n}}\right),$$

where $\Phi(u) \stackrel{\text{def}}{=} \int_{t:|\varphi(t)| \ge u} dt$ and φ is the characteristic function for f. When $\int \sqrt{f} = \infty$,

$$\mathbf{E}\left\{\int |f - f_n|\right\} \ge (1 + o(1)) \inf_{0 \le u \le 1} \max\left(u, \left(\frac{T\Phi(2u)}{n}\right)^{1/2}\right),$$

for all constants T.

In section 3, we will discuss this theorem. Further theorems for restricted classes of kernels in the standard kernel estimate are given in section 5. It is shown there that restrictions on the shape of K generally impose limitations on the best possible rate of convergence, regardless of how smooth f is. The lower bounds of Theorem 1 and Theorems 2 and 3 of section 5 are valid for all densities on the real line, without restrictions.

3. Discussion of Theorem 1.

The asymptotic lower bound given in Theorem 1 shows the importance of two factors, the size of the tail of f (as measured by $\int \sqrt{f}$, which is proportional to how "spread out" f is), and the smoothness of f (as captured in the tail behaviour of the characteristic function φ). $\int \sqrt{f}$ appears in just about every inequality related to the rate of convergence of density estimates given in the L_1 study of Devroye and Györfi (1985), so its presence here is entirely natural. It is finite whenever $\int |x|^{1+\epsilon} f(x)dx < \infty$ for some $\epsilon > 0$. It is well-known that this factor is absent in L_2 asymptotics.

Watson and Leadbetter (1963) and Davis (1975, 1977) have related the L_2 behavior of kernel estimates to the characteristic function. Roughly speaking, the smaller the tail of the characteristic function, the smoother f, and the better the best achievable performance for f with some kernel K_n. Davis showed that the Watson-Leadbetter lower bounds can in fact be attained up to a multiplicative constant by using a standard kernel estimate with appropriate h and "sink" kernel $L(x) = \sin(x)/\pi x$. Below, we will give some examples of classes of densities and the accompanying lower bounds. The best achievable rate for a density f is a function of the two factors mentioned above.

It should be stressed that the shape of K_n is allowed to change with n. For the standard kernel estimate, with its fixed shape, the bounds may be rather loose.

Finite support characteristic functions. Assume that $T = \int_{|\varphi(t)|>0} dt$ is finite. Since the minimal u in the definition of the lower bound tends to zero, $\Phi(2u) \to T$. We conclude that

$$\mathbf{E}\left\{\int |f - f_n|\right\} \geq (1 + o(1))\frac{\int \sqrt{f}\sqrt{T}}{8\sqrt{2\pi n}}.$$

The densities in this class are very smooth; and indeed, this lower bound is the smallest among the lower bounds to be discussed here. By Lemma 7 (see next section) and the fact that $\int |\varphi| \leq T$, we see immediately that $\int \sqrt{f}\sqrt{T} \geq \sqrt{2\pi}$, which yields the universal bound

$$\mathbf{E}\left\{\int |f - f_n|\right\} \geq \frac{1 + o(1)}{8\sqrt{n}},$$

valid for all f in the given class. □

Characteristic functions with infinite support. When $T = \int_{|\varphi(t)|>0} dt = \infty$, we can formally replace T in the previous remark by arbitrary large constants. Hence, we conclude that

$$\liminf_{n\to\infty} \sqrt{n}\mathbf{E}\left\{\int |f - f_n|\right\} = \infty.$$

For the standard kernel estimate satisfying condition E, a slightly more general result was obtained in Devroye and Györfi (1985, Theorem 5.16, part 3). □

A universal asymptotic lower bound. The previous two remarks taken together imply that for **all** densities f,

$$\liminf_{n\to\infty} \sqrt{n}\mathbf{E}\left\{\int |f - f_n|\right\} \geq \frac{1}{8}. \ \square$$

Densities with a large tail. Consider a standard kernel estimate satisfying condition E. Then Lemma 6 (see next section) implies that

$$\liminf_{n\to\infty} \sqrt{nh}\mathbf{E}\left\{\int |f - f_n|\right\} = \infty$$

whenever $\int \sqrt{f} = \infty$. This all but coincides with part 2 of Theorem 5.16 of Devroye and Györfi (1985). □

Super-polynomial characteristic functions. In this remark, we consider characteristic functions for which $|\varphi(t)| \geq \alpha/|t|^\beta$ for some $\alpha, \beta > 0$ and all t. This class includes the gamma family. We note that $\Phi(u) \geq 2(\alpha/u)^{1/\beta}$. Thus,

$$\mathbf{E}\left\{\int |f - f_n|\right\} \geq (1 + o(1)) \inf_{0\leq u\leq 1} \max\left(u, \frac{C}{\left(nu^{\frac{1}{\beta}}\right)^{1/2}}\right),$$

where $C \stackrel{\text{def}}{=} \int \sqrt{f}\left(((\alpha/2)^{1/\beta})/(64\pi)\right)^{1/2}$. The u minimizing the maximum is obtained by equating the two terms in the maximum:

$$u = \left(\frac{C}{\sqrt{n}}\right)^{\frac{\beta}{\beta+1}}.$$

We conclude that

$$\mathbf{E}\left\{\int |f - f_n|\right\} \geq (1+o(1))C^{\frac{\beta}{\beta+1}}n^{-\frac{\beta}{2\beta+2}} = (1+o(1))\frac{(\int\sqrt{f})^{\frac{\beta}{\beta+1}}\alpha^{\frac{1}{2\beta+2}}}{2^{\frac{1}{2\beta+2}}(64\pi n)^{\frac{\beta}{2\beta+2}}}.$$

The lower bound decreases at a polynomial rate in n, with power $-\beta/(2\beta+2)$ strictly between 0 and $-1/2$. Thus, by picking β small enough, any slow polynomial rate can be achieved. It should be noted that the presence of one discontinuity of the ordinary kind implies that $\limsup |t\varphi(t)| > 0$, from which it can be concluded that $\Phi(u) \geq c/u$ for some constant $c > 0$. By the argument used above with $\beta \equiv 1$, we see that at best we have a lower bound decreasing as $n^{-1/3}$. $\square$

Super-exponential characteristic functions. Many densities, including the normal density, and indeed all stable densities, are so smooth that their tails drop off at an exponential rate. Assume for example that for some positive constants α, β, γ, $|\varphi(t)| \geq \alpha exp(-\gamma|t|^\beta)$ for all t. Using the same technique of the previous section, we observe that

$$\mathbf{E}\left\{\int |f - f_n|\right\} \geq (1+o(1))\frac{\int\sqrt{f}\log^{\frac{1}{2\beta}}(n)}{(2\gamma)^{\frac{1}{2\beta}}8\sqrt{\pi n}}.$$

This decreases as $\log^{1/(2\beta)}(n)/\sqrt{n}$. For the normal density, we obtain the lower bound $\log^{1/4}(n)/\sqrt{n}$. $\square$

Absolutely integrable characteristic functions. The absolute integrability of φ implies that f is absolutely continuous. If $\int |\varphi| = \infty$, we can say that f is not very smooth. It is not difficult to see that for all $\epsilon > 0$, $\Phi(u) \geq c/(u\log^{1+\epsilon}(1/u))$ for some positive constant c and all u small enough. This can be used in Theorem 1 to conclude that

$$\liminf_{n\to\infty}\left(n\log^{1+\epsilon} n\right)^{\frac{1}{3}}\mathbf{E}\left\{\int |f - f_n|\right\} > 0$$

where $\epsilon > 0$ is arbitrary. Since $n^{-1/3}$ can hardly be considered a "good" rate of convergence, we may conclude that kernel estimates are ill suited for estimating densities whose characteristic function is not absolutely integrable. $\square$

4. Some technical lemmas.

At a crucial junction, we need the following lower bound, adapted from Devroye and Györfi (1985, Lemma 27, pp. 136-137), and based upon inequalities of Haagerup (1978) and Szarek (1976):

Lemma 1. *Let $X_1,\ldots,X_n$ be iid zero mean random variables with finite first absolute moment. Then*

$$\mathbf{E}\left\{|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i|\right\} \geq \frac{1}{2}\mathbf{E}\left\{\left(\frac{1}{2n}\sum_{i=1}^{n}(X_i-X_i')^2\right)^{1/2}\right\} \geq \frac{1}{\sqrt{8}}\mathbf{E}\left\{|X_1|\right\},$$

where $X_1',\ldots,X_n'$ are iid random variables, distributed as $X_1,\ldots,X_n$, but independent of this sequence.

Lemma 2. *Let $X_1,\ldots,X_n$ be iid zero mean random variables with finite fourth absolute moment. Then*

$$\mathbf{E}\left\{\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i\right|\right\} \geq \frac{1}{2}\left(\mathbf{E}\left\{X_1^2\right\}\right)^{1/2} - \left(\frac{1}{4n}\right)^{1/4}\mathbf{E}^{1/4}\left\{X_1^4\right\}.$$

Proof. From Lemma 1, we recall

$$\begin{aligned}
\mathbf{E}\left\{|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i|\right\} &\geq \frac{1}{2}\mathbf{E}\left\{\left(\frac{1}{2n}\sum_{i=1}^{n}(X_i-X_i')^2\right)^{1/2}\right\}\\
&\geq \frac{1}{2}\left(\mathbf{E}\left\{X_1^2\right\}\right)^{1/2} - \frac{1}{2}\mathbf{E}\left\{\left(\frac{1}{2n}\left|\sum_{i=1}^{n}(X_i-X_i')^2-\mathbf{E}\left\{(X_i-X_i')^2\right\}\right|\right)^{1/2}\right\}\\
&\geq \frac{1}{2}\left(\mathbf{E}\left\{X_1^2\right\}\right)^{1/2} - \frac{1}{2}\mathbf{E}^{1/4}\left\{\frac{1}{(2n)^2}\left(\sum_{i=1}^{n}(X_i-X_i')^2-\mathbf{E}\left\{(X_i-X_i')^2\right\}\right)^2\right\}\\
&\geq \frac{1}{2}\left(\mathbf{E}\left\{X_1^2\right\}\right)^{1/2} - \frac{1}{2}\mathbf{E}^{1/4}\left\{\frac{1}{4n}(X_1-X_1')^4\right\}\\
&\geq \frac{1}{2}\left(\mathbf{E}\left\{X_1^2\right\}\right)^{1/2} - \frac{1}{2}\mathbf{E}^{1/4}\left\{\frac{16}{4n}X_1^4\right\}\\
&= \frac{1}{2}\left(\mathbf{E}\left\{X_1^2\right\}\right)^{1/2} - \mathbf{E}^{1/4}\left\{\frac{X_1^4}{4n}\right\}. \square
\end{aligned}$$

Lemma 3. *Let f_n be the kernel estimate with kernel K_n. Then, for all f,*

$$\begin{aligned}
\mathbf{E}\left\{|f_n(x)-K_n*f(x)|\right\} \geq \frac{1}{\sqrt{4n}}\left(K_n^2*f(x)\right)^{1/2} - \frac{1}{\sqrt{4n}}|K_n*f(x)|\\
-(\frac{1}{n})^{\frac{3}{4}}\left((8K_n^4*f(x))^{1/4}+8^{1/4}|K_n*f(x)|\right)
\end{aligned}$$

where $*$ is the convolution operator (i.e., $g*f(x)\stackrel{\text{def}}{=}\int g(x-y)f(y)dy$).

PROOF. If we apply Lemmas 1 and 2 with $X_i := K_n(x - X_i) - K_n * f$, then we obtain

$$\mathbf{E}\left\{|f_n(x) - K_n * f(x)|\right\} \geq \frac{1}{\sqrt{8n}}\mathbf{E}\left\{|K_n(x - X_1) - K_n * f(x)|\right\}.$$

Also,

$$\begin{aligned}
&\mathbf{E}\left\{|f_n(x) - K_n * f(x)|\right\} \\
&\geq \frac{1}{\sqrt{4n}}\left(\mathbf{E}\left\{(K_n(x - X_1) - K_n * f(x))^2)\right\}^{1/2}\right) \\
&\quad - (\frac{1}{4^{1/3}n})^{3/4}\mathbf{E}^{1/4}\left\{(K_n(x - X_1) - K_n * f(x))^4\right\} \\
&\geq \frac{1}{\sqrt{4n}}\mathbf{E}\left\{(K_n^2(x - X_1))\right\}^{1/2} - \frac{1}{\sqrt{4n}}\left((K_n * f(x))^2)\right)^{1/2} \\
&\quad - (\frac{1}{4^{1/3}n})^{\frac{3}{4}}\mathbf{E}^{1/4}\left\{8K_n^4(x - X_1) + 8(K_n * f(x))^4\right\} \\
&\geq \frac{1}{\sqrt{4n}}\left(K_n^2 * f(x)\right)^{1/2} - \frac{1}{\sqrt{4n}}|K_n * f(x)| - \left(\frac{1}{n}\right)^{\frac{3}{4}}\left((8K_n^4 * f(x))^{1/4} + 8^{1/4}|K_n * f(x)|\right). \square
\end{aligned}$$

LEMMA 4. *If $\{K_n\}$ is a regular sequence of kernels, then $\int K_n^2 \to \infty$.*

PROOF. The statement follows from the fact that $\int K_n^2 f \geq \int^2 |K_n| f$ for all densities f by the Cauchy-Schwarz inequality. Take for f the uniform density on $[-\delta, \delta]$, for $\delta > 0$ arbitrarily small, and observe that

$$\int K_n^2 \geq \frac{1}{2\delta}\int_{-\delta}^{\delta} |K_n|.$$

By condition A, the right-hand-side has a limit infimum that is at least one divided by 2δ. This concludes the proof of Lemma 4. $\square$

LEMMA 5. *Let f_n be the kernel estimate with a regular sequence of kernels $\{K_n\}$. Then, for all f,*

$$\liminf_{n\to\infty}\left(\frac{4n}{\int K_n^2}\right)^{1/2}\mathbf{E}\left\{\int |f_n(x) - K_n * f(x)| dx\right\} \geq \int \sqrt{f}.$$

PROOF. From Lemma 3, writing L_n for $K_n^2/\int K_n^2$, we have

$$\left(\frac{4n}{\int K_n^2}\right)^{1/2}\mathbf{E}\left\{|f_n(x)-K_n*f(x)|\right\}$$
$$\geq (L_n*f(x))^{1/2}-|K_n*f(x)|/\left(\int K_n^2\right)^{1/2}-\frac{2n^{-1/4}}{\left(\int K_n^2\right)^{1/2}}\left((8K_n^4*f(x))^{1/4}+8^{1/4}|K_n*f(x)|\right).$$

The first term on the rhs tends to $\sqrt{f}$ for almost all x. The second term's numerator tends to f for almost all x, while its denominator tends to ∞ by Lemma 4. The absolute value of the third and last term is $(1+o(1))$ times

$$\frac{2n^{-1/4}}{\left(\int K_n^2\right)^{1/2}}\left((8f(x)\int K_n^4)^{1/4}+8^{1/4}f(x)\right).$$

By condition D and Lemma 4, this is $o(1)$ for almost all x. We can now apply Fatou's lemma:

$$\liminf_{n\to\infty}\int\left(\frac{4n}{\int K_n^2}\right)^{1/2}\mathbf{E}\int(|f_n(x)-K_n*f(x)|)\,dx$$
$$\geq\int\liminf_{n\to\infty}\left(\left(\frac{4n}{\int K_n^2}\right)^{1/2}\mathbf{E}\left\{|f_n(x)-K_n*f(x)|\right\}\right)dx$$
$$\geq\int\sqrt{f}.\ \square$$

LEMMA 6. *Let f_n be a kernel estimate with kernel K_n, and let f be an arbitrary density. Then*

$$\mathbf{E}\left\{\int|f-f_n|\right\}\geq\int|f-K_n*f|\geq\sup_t|\varphi(t)||1-\psi_n(t)|,$$

where φ is the characteristic function of f, and ψ_n is the Fourier transform for K_n (i.e., $\psi_n(t)\stackrel{\text{def}}{=}\int e^{itx}K_n(x)dx$). Also,

$$\mathbf{E}\left\{\int|f-f_n|\right\}\geq\frac{1}{2}\mathbf{E}\left\{\int|f_n-K_n*f|\right\}.$$

PROOF. See, e.g., Devroye and Györfi, 1985, p. 139. $\square$

A starting point for lower bounds.

Combining Lemmas 5 and 6, we see that for the kernel estimate with a regular sequence of kernels $\{K_n\}$,

$$\mathbf{E}\left\{\int|f-f_n|\right\}\geq(1+o(1))\max\left(\int|f-K_n*f|,\frac{\int\sqrt{f}\sqrt{\int K_n^2}}{2\sqrt{4n}}\right).$$

Unfortunately, the bias term in the lower bound cannot be bounded using a Taylor series expansion for f as is standard practice because some kernels are such that all their positive moments are zero. To get more information and to eliminate K_n from the lower bound, we will use the characteristic function route suggested by the first inequality of Lemma 6. Alternately, we could also have used the following measure of the "unsmoothness" of f:

$$U(f,c) \stackrel{\text{def}}{=} \inf_{K:\int K=1,\int K^2=c} \int |f-K*f|,$$

where the behavior of $U(f,c)$ for $c \downarrow 0$ is important. However, characterizing unsmoothness in terms of the tail behavior of the characteristic function of f seems to lead to results that are easier to interpret. We should warn the reader here that we now run the risk of obtaining lower bounds that are not attainable in some cases, as the inequality used in Lemma 6 is not necessarily always tight.

PROOF. Assume first that $\int\sqrt{f}<\infty$. Combining Lemma 5 and Lemma 6, we see that

$$\mathbf{E}\left\{\int |f-f_n|\right\} \geq \max\left(\sup_t |\varphi(t)||1-\psi_n(t)|, \frac{\int\sqrt{f}\sqrt{\int K_n^2}(1+o(1))}{2\sqrt{4n}}\right)$$

where ψ_n is the Fourier transform of K_n. Assume next that

$$\sup_t |\varphi(t)||1-\psi_n(t)| = u.$$

Then, applying Bessel's equality (also known as Parseval's identity), we have

$$\int K_n^2 = \frac{1}{2\pi}\int \psi_n^2 = \frac{1}{2\pi}\int (1-(1-\psi_n(t))^2 dt \geq \frac{1}{2\pi}\int \left(1-\frac{u}{|\varphi(t)|}\right)_+^2 dt \geq \frac{1}{8\pi}\int_{t:|\varphi|(t)\geq 2u} dt = \frac{\Phi(2u)}{8\pi}.$$

Thus, we certainly have

$$\mathbf{E}\left\{\int |f-f_n|\right\} \geq (1+o(1)) \inf_{0\leq u\leq 1} \max\left(u, \frac{\int\sqrt{f}\sqrt{\Phi(2u)}}{8\sqrt{2\pi n}}\right),$$

which proves Theorem 1 when $\int\sqrt{f}<\infty$. When $\int\sqrt{f}=\infty$, we can formally replace $(1+o(1))\int\sqrt{f}$ in the bounds shown above by $(1+o(1))T$ where T is arbitrarily large. This concludes the proof of the Lemma. □

LEMMA 7. *For any density f,*

$$\int\sqrt{f} \geq \frac{1}{\sqrt{\sup f}} \geq \left(\frac{2\pi}{\int|\varphi|}\right)^{1/2},$$

where φ is the characteristic function for φ.

PROOF. The first inequality follows from the fact that

$$1 = \int f \leq \sqrt{\sup f}\int\sqrt{f}.$$

The second inequality is a simple corollary of the standard formula for inverting a characteristic function,

$$\sup f \leq \frac{1}{2\pi}\int|\varphi|. \ \square$$

5. Saturation: restricted classes of kernels.

Assume next that the kernel in the standard kernel estimate is restricted in certain ways. It is known that some restrictions lead to limitations on the best possible rate of convergence, regardless of how smooth f is. This should be contrasted with the lower bounds of the previous section, in which the best possible rate is only determined by the smoothness of f, as long as $\int \sqrt{f} < \infty$. A point in case is the result of Devroye and Penrod (1984) mentioned at the outset of this paper, which states that for the standard kernel estimate with arbitrary nonnegative kernel L and arbitrary sequence of smoothing factors h, $n^2/5\mathbf{E}\left\{\int |f-f_n|\right\} \geq 0.86 + o(1)$.

The latter result will be generalized here to include other (possibly negative-valued) kernels. In addition, the proofs given here are shorter than those of Devroye and Penrod (1984). The constants in the various inequalities are slightly worse, however.

We need one technical lemma (Lemma 10), delayed until the end of the section. We also need to introduce the class of kernels $\mathcal{K}_s$ (where s is an even positive integer) consisting of all even kernels L for which $\int L = 1$, $\int x^i L(x)dx = 0$ for $1 \leq i < s$, $\nu_s \stackrel{\text{def}}{=} \int |x|^s |L(x)|dx < \infty$, and $\mu_s \stackrel{\text{def}}{=} |\int x^s L(x)dx| \neq 0$.

THEOREM 2. *Let f be a density with characteristic function φ. Let f_n be the standard kernel estimate and let its kernel L belong to $\mathcal{K}_s$ for some even s and satisfy condition E. If L has Fourier transform ψ, then*

$$\liminf_{n\to\infty} n^{\frac{s}{2s+1}} \inf_{h>0} \mathbf{E}\left\{\int |f-f_n|\right\} \geq \left(\frac{\int \sqrt{f}\sqrt{\int \psi^2}}{2\sqrt{8\pi}}\right)^{\frac{2s}{2s+1}} \left(\frac{\mu_s}{s!}\sup_t |t|^s|\varphi(t)|\right)^{\frac{1}{2s+1}}.$$

This inequality remains valid even if $\int \sqrt{f} = \infty$ or $\sup_t |t|^s|\varphi(t)| = \infty$.

PROOF. Let h^* be a sequence such that $\mathbf{E}\left\{\int |f-f_n|\right\} \sim \inf_h \mathbf{E}\left\{\int |f-f_n|\right\}$. If $\liminf h^* > 0$, then $\liminf \mathbf{E}\left\{\int |f-f_n|\right\} > 0$ in view of Lemma 6, and the fact that ψ is not identical one in some open neighborhood of the origin (for otherwise μ_s would be zero). Thus, assume that $h^* \to 0$. If $\limsup nh^* < \infty$, then $\liminf \mathbf{E}\left\{\int |f-f_n|\right\} > 0$ for arbitrary f and absolutely integrable kernel L (see Lemma S1 of Devroye (1987), which is an extension towards nonnegative-valued kernels of Devroye and Penrod (1984)). Thus, we can assume that $h^* \to 0$ and $nh^* \to \infty$. Hence, the estimate satsfies condition E. For simplicity, we now write h instead of h^*. From Lemmas 5 and 6, we recall that

$$\mathbf{E}\left\{\int |f-f_n|\right\} \geq \max\left(\sup_t |\varphi(t)||1-\psi(th)|, \frac{\int \sqrt{f}\sqrt{\int \psi^2}(1+o(1))}{2\sqrt{8\pi nh}}\right)$$

when the standard kernel estimate satisfies condition E, and $\int \sqrt{f} < \infty$. By Lemma 10 (see end of this section), the last expression is bounded from below by

$$\max\left((1+o(1))\frac{h^s\mu_s}{s!}\sup_t |t|^s|\varphi(t)|, \frac{\int \sqrt{f}\sqrt{\int \psi^2}(1+o(1))}{2\sqrt{8\pi nh}}\right)$$

$$\geq (1+o(1)) \inf_{h>0} \max \left(\frac{h^s \mu_s}{s!} \sup_t |t|^s |\varphi(t)|, \frac{\int \sqrt{f} \sqrt{\int \psi^2}}{2\sqrt{8\pi n h}} \right)$$

when $L \in \mathcal{K}_s$ and $\sup_t |t|^s|\varphi(t)| < \infty$. The maximum can be minimized by equating both terms. This yields the formula

$$h^{s+\frac{1}{2}} = \frac{s! \int \sqrt{f} \sqrt{\int \psi^2}}{2\mu_s \sqrt{8\pi n}} \sup_t |t|^s |\varphi(t)|.$$

Resubstitution yields

$$\mathbf{E}\left\{ \int |f - f_n| \right\}$$

$$\geq (1+o(1)) \left(\frac{\int \sqrt{f} \sqrt{\int \psi^2}}{2\sqrt{8\pi n}} \right)^{\frac{2s}{2s+1}} \left(\frac{\mu_s}{s!} \sup_t |t|^s |\varphi(t)| \right)^{\frac{1}{2s+1}}.$$

This proves Theorem 2 when $\int \sqrt{f} < \infty$ and $\sup_t |t|^s|\varphi(t)| < \infty$. If either of these factors is infinite, it is easy to verify that Theorem 2 is also valid. □

Theorem 2 describes the phenomenon of **saturation**: the best possible rate of convergence for any density f is $n^{-s/(2s+1)}$ for $L \in \mathcal{K}_s$, and is thus limited by s, a parameter depending upon K only. The fact that negative-valued kernels (i.e. kernels with $s > 2$) can, in some cases, yield rates of convergence faster than $n^{-2/5}$ has been known for a long time; see e.g. Bartlett (1963).

The lower bound is of the form $A(f)B(L)$ where $A(f)$ depends upon f only, and $B(L)$ depends upon the kernel only. We will consider each factor in turn.

LEMMA 8. *Define*

$$B(L) \stackrel{\text{def}}{=} \left(\frac{\sqrt{\int \psi^2}}{2\sqrt{8\pi}} \right)^{\frac{2s}{2s+1}} \left(\frac{\mu_s}{s!} \right)^{\frac{1}{2s+1}}.$$

Then

$$B(L) \geq \left(\frac{1}{2\sqrt{8\pi}} \right)^{\frac{2s}{2s+1}} \left(\frac{\mu_s}{2\nu_s} \right)^{\frac{1}{2s+1}}.$$

PROOF. By Taylor's series expansion for ψ,

$$|\psi(t) - 1| \leq \frac{|t|^s}{s!} \sup_u |\psi^(s)(u)| = \frac{|t|^s}{s!} \sup_u |\int x^s L(x) e^{iux} dx|$$
$$\leq \frac{|t|^s}{s!} \int x^s |L(x)| dx \stackrel{\text{def}}{=} \frac{|t|^s}{s!} \nu_s.$$

But then, $|\psi(t)| \geq 1/2$ when $|t|^s \leq s!/(2\nu_s)$. Hence,

$$\int \psi^2 \geq \frac{1}{2} \left(\frac{s!}{2\nu_s} \right)^{\frac{1}{s}}.$$

Thus,

$$B(L) \geq \left(\frac{1}{2\sqrt{8\pi}}\right)^{\frac{2s}{2s+1}} \left(\frac{\mu_s}{2\nu_s}\right)^{\frac{1}{2s+1}}. \square$$

It should be recognized that the lower bound for $B(L)$ can take any small value; this of course is due to the fact that when $\mu_s = 0$, we have in many cases, $L \in \mathcal{K}_{s+2}$, and the lower bound should in fact be zero, since we are bumped up to the next higher s. In the special case that $s = 2$ and $L \geq 0$, we see that $\nu_2 = \mu_2$, and our lower bound becomes

$$B(L) \geq \frac{1}{2(8\pi)^{2/5}}.$$

Let us now turn to the factor

$$A(f) \stackrel{\text{def}}{=} \left(\int \sqrt{f}\right)^{\frac{2s}{2s+1}} \left(\sup_t |t|^s |\varphi(t)|\right)^{\frac{1}{2s+1}}.$$

This quantity is an appropriate measure of the "difficulty" posed by f for kernels $L \in \mathcal{K}_s$. Again, it is composed of a tail factor, $\int \sqrt{f}$, and a smoothness factor depending upon the behavior of φ. It can be verified that $A(f)$ is scale and translation invariant. Its value depends upon the shape of the density only. It is infinite for one of two reasons only, either because the tail is too large, or the density is not smooth enough. In any case, we have

Lemma 9. *For any density f,*

$$A(f) \geq \left(\frac{\pi(s-1)}{s}\right)^{\frac{s}{2s+1}}.$$

Proof. We can assume without loss of generality that both factors in the definition of $A(f)$ are finite. We begin by noting that $\int \sqrt{f} \geq \sqrt{2\pi}/\left(\int |\varphi|\right)^{1/2}$ where φ is the characteristic function for f. See, e.g., Lemma 7. If $\sup_t |t|^s|\varphi(t)| = a$ for some constant a, then $|\varphi| \leq \min(1, a/|t|^s)$, and thus,

$$\int |\varphi| \leq \int \min\left(1, \frac{a}{|t|^s}\right) dt = \frac{2s}{s-1} a^{\frac{1}{s}}.$$

Therefore,

$$A(f) \geq \left(\frac{2\pi}{\int |\varphi|}\right)^{\frac{s}{2s+1}} \left(\sup_t |t|^s|\varphi(t)|\right)^{\frac{1}{2s+1}} \geq \left(\frac{2\pi(s-1)}{2sa^{1/s}}\right)^{\frac{s}{2s+1}} a^{\frac{1}{2s+1}} = \left(\frac{\pi(s-1)}{s}\right)^{\frac{s}{2s+1}}. \square$$

Theorem 3. *Consider any density f, and all standard kernel estimates covered by Theorem 2. Then*

$$\liminf_{n\to\infty} n^{\frac{s}{2s+1}} \mathbf{E}\left\{\int |f - f_n|\right\} \geq \left(\frac{s-1}{32s}\right)^{\frac{s}{2s+1}} \left(\frac{\mu_s}{2\nu_s}\right)^{\frac{1}{2s+1}}.$$

Proof. Theorem 3 follows immediately by combining the bounds of Lemmas 9 and 10 and Theorem 2. □

For nonnegative kernels and $s = 2$, we obtain, in particular,

$$\liminf_{n\to\infty} n^{2/5} \mathbf{E}\left\{\int |f - f_n|\right\} \geq 2^{-13/5} = 0.164938488\ldots.$$

This bound is not as good as the bound of approximately 0.86 obtained by Devroye and Penrod (1984), due to the fact that many shortcuts were taken here to simplify the proofs and to obtain bounds that are valid for all kernels. Since for nonnegative kernels, $\mathbf{E}\left\{\int |f - f_n|\right\} \to 0$ implies $h \to 0$ and $nh \to \infty$ (Devroye, 1983), we see that for any even nonnegative kernel L satisfying condition E, and having a finite second moment,

$$\liminf_{n\to\infty} n^{2/5} \inf_h > 0 \mathbf{E}\left\{\int |f - f_n|\right\} \geq A(f)B(L) \geq 2^{\frac{-13}{5}} = 0.164938488\ldots.$$

Lemma 10. *Let f be any density (with characteristic function φ), and let L be any kernel $L \in \mathcal{K}_s$ (with Fourier transform ψ). Assume that $h \downarrow 0$. If $\sup_t |t|^s|\varphi(t)| < \infty$, then*

$$\sup_t |1 - \psi(th)||\varphi(t)| \geq (1 + o(1)) \frac{\mu_s h^s}{s!} \sup_t |t|^s |\varphi(t)|.$$

Otherwise,

$$\liminf_{h\downarrow 0} \frac{\sup_t |1 - \psi(th)||\varphi(t)|}{h^s} = \infty.$$

Proof. Since $|L|$ has finite absolute s-th moment, $\psi^(s)$ exists everywhere and is continuous (Kawata, 1972, pp. 411). The proof is in two steps. Assume first that $|t|^s|\varphi(t)| \to 0$ as $|t| \to \infty$. Then, for $t > 0$,

$$\psi(th) = 1 + \frac{(th)^s \psi^(s)(0)}{s!} + \frac{(th)^s(\psi^(s)(\xi) - \psi^(s)(0))}{s!}$$

for some $0 \leq \xi \leq th$ (by Taylor's series expansion and the conditions imposed on L). Clearly,

$$\begin{aligned} &\sup_t |1 - \psi(th)||\varphi(t)| \\ &\geq \sup_t \frac{|th|^s \mu_s |\varphi(t)|}{s!} - \sup_t \frac{(th)^s R(th)|\varphi(t)|}{s!}. \end{aligned}$$

where $R(u) \stackrel{\text{def}}{=} \sup_{|\xi|\leq|u|} |\psi^(s)(\xi) - \psi^(s)(0)|$. Since $\int |x|^s|L(x)|dx < \infty$, $\psi^(s)$, and thus $R(u)$, remains bounded. Also, $R(u) \to 0$ as $u \to 0$. Thus, for every $\epsilon > 0$, we can find $\delta > 0$ such that

$$\begin{aligned} \sup_t \frac{(th)^s R(th)|\varphi(t)|}{s!} &\leq \epsilon \sup_{|t|\leq\delta/h} \frac{(th)^s|\varphi(t)|}{s!} + \sup_u R(u) \sup_{|t|>\delta/h} \frac{(th)^s|\varphi(t)|}{s!} \\ &\leq \epsilon h^s \sup_t \frac{t^s|\varphi(t)|}{s!} + o(h^s). \end{aligned}$$

Thus,

$$\sup_t |1 - \psi(th)||\varphi(t)| \geq (1 + o(1)) \sup_t \frac{|th|^s \mu_s |\varphi(t)|}{s!}.$$

For general φ, we introduce a mollifier with characteristic function $\chi(tu)$, where u is a scale factor, and χ has the property that $t^s\chi(t) \to 0$ as $|t| \to \infty$. What we proved above certainly remains valid for $|\varphi(t)||\chi(tu)|$ (with u fixed):

$$\sup_t |1-\psi(th)||\varphi(t)| \geq \sup_t |1-\psi(th)||\varphi(t)||\chi(tu)|$$

$$\geq (1+o(1)) \sup_t \frac{|th|^s \mu_s |\varphi(t)||\chi(tu)|}{s!}.$$

The right-hand-side divided by h^s is at least

$$(1+o(1))(1-\epsilon) \sup_{|tu|\leq\delta} \frac{|th|^s \mu_s |\varphi(t)|}{s!}$$

where $\epsilon > 0$ is arbitrary, $\delta > 0$ is a function of χ and ϵ, and u can be chosen at will. If we let u tend to zero, it is clear that the supremum approaches

$$\sup_t \frac{|th|^s \mu_s |\varphi(t)|}{s!},$$

even if this is infinite. This concludes the proof of the Lemma. □

6. Are the bounds attainable?

We will say that a lower bound is **attainable** if we can exhibit a kernel estimate with expected L_1 error bounded above by $c + o(1)$ times the lower bound for some constant c. It is the purpose of this section to indicate that the bounds of Theorems 2 and 3 are attainable, and that the same is true for those of Theorem 1 whenever f is very smooth.

THE SATURATION BOUNDS OF THEOREMS 2 AND 3. In Devroye and Györfi (1985, p.208), kernel estimates are constructed that have expected L_1 bounds that vary as $n^{-s/(2s+1)}$ (s even) provided that f has compact support (hence, $\int \sqrt{f} < \infty$), that f has $s-1$ absolutely continuous derivatives, and that $f^(s)$ is continuous. In addition, it is assumed that $\int |f^(s)| < \infty$. But under these conditions, the characteristic function φ of f satisfies the inequality

$$\sup_t |t|^s |\varphi(t)| \leq \int |f^(s)| \leq inf.$$

The condition that f have compact support can be relaxed to $\int |x|^{1+\epsilon} f(x)dx < \infty$ for some $\epsilon > 0$ (Devroye, 1987). In these cases, the lower bound of Theorem 2 is attained up to a multiplicative constant. For the important case of nonnegative kernels, the bounds of Theorem 2 and 3 cannot be improved upon, except possibly by a multiplicative constant.

THE BOUNDS OF THEOREM 1. We will argue that the bound of Theorem 1 is attainable for all very smooth densities f. However, to avoid a lengthy technical treatment, we will make our case with the help of two examples, all densities whose characteristic functions have compact support, and all densities with a characteristic function of the form $\varphi(t) = \exp(-(1+o(1))|t|^\beta)$ as $|t| \to \infty$. For performance that is not restricted by the form of the kernel as in the case of saturation, one should consider the standard kernel estimate with kernels whose characteristic function is flat in an open neighborhood of the origin. With such kernels (coined superkernels in Devroye (1987); see also Devroye and Györfi (1985, section 5.11)), one generally obtains the rates predicted by Theorem 1. Consider first densities f with a bounded support characteristic function. The use of the "sink" kernel $\sin(x)/(\pi x)$ leads to $1/n$ error rates for the mean square integrated error

(Davis (1975, 1977); Ibragimov and Khasminskii (1982)). When we choose a kernel L whose characteristic function is zero in a neighborhood of the origin, and for which $\int(1+x^2)L^2 < \infty$, and when we keep h fixed at a certain positive value, then $\mathbf{E}\left\{\int|f-f_n|\right\} = O(1/\sqrt{n})$ for all f with finite second moment (Devroye, 1987). Consider next densities with characteristic function of the form

$$\varphi(t) = e^{-(1+o(1))|t|^\beta}$$

as $|t| \to \infty$, where $\beta > 0$ is a constant. These densities are very smooth. For example, when $\beta \geq 1$, the densities are analytic (Kawata, 1972, pp. 439-440). Assume furthermore that $\int |x|^{1+\epsilon} f(x)dx < \infty$ for some $\epsilon > 0$. (This condition implies that $\int \sqrt{f} < \infty$, and is only slightly stronger than $\int \sqrt{f} < \infty$.) In view of the example worked out in section 3, it suffices to show that $\mathbf{E}\left\{\int|f-f_n|\right\}$ is not greater than a constant times $\log^{1/(2\beta)} n/\sqrt{n}$. We again choose the standard kernel estimate with smoothing factor h varying as a constant (to be picked further on) times $\log^{-1/\beta} n$, and kernel L defined via its characteristic function (actually, Fourier transform)

$$\psi(t) = \begin{cases} 1 & |t| \leq 1, \\ \frac{1}{1+(|t|-1)^4} & |t| > 1. \end{cases}$$

It is easy to verify that $|L|$ is bounded (by $\frac{1}{2\pi}\int|\psi|$), and $x^2|L(x)|$ is bounded (by $\frac{1}{2\pi}\int|\psi^(2)|$). Thus, $\int_z^\infty |L(x)|dx \leq c/z$ for some constant c and all $z > 0$, a fact that will be needed further on. Let us define $L_h(x) = (1/h)L(x/h)$. We will have achieved our goal if we can show that $\mathbf{E}\left\{\int|f_n - f * L_h|\right\} \leq c_1/\sqrt{nh}$ and that $\int|f - f * L_h| \leq c_2 e^{-c_3/h^\beta}$ for some positive constants c_i. The moment condition on f implies that

$$\mathbf{E}\left\{\int |f_n - f * L_h|\right\} \sim \frac{\int\sqrt{f}\sqrt{\int L^2}}{\sqrt{nh}}$$

as $n \to \infty$ (see Devroye, 1987). Since φ is absolutely integrable, we can recover f by standard inversion. In particular,

$$\begin{aligned} \sup_x |f(x) - f * L_h(x)| &\leq \frac{1}{2\pi}\int |\varphi(t)||1-\psi(th)|dt \\ &\leq \frac{1}{2\pi}\int_{|t|>1/h} |\varphi(t)|dt \\ &\leq e^{-(1+o(1))h^{-\beta}} \end{aligned}$$

as $h \downarrow 0$ (where we absorbed all the constants conveniently in the $o(1)$ term). Let us also define $\delta \stackrel{\text{def}}{=} e^{1/((1+\epsilon)h^\beta)}$. Then,

$$\begin{aligned} \int |f - f * L_h| &\leq 2\delta e^{-(1+o(1))h^{-\beta}} + \int_{|x|>\delta} (f + f * |L_h|) \\ &\leq 2e^{-\frac{\epsilon+o(1)}{1+\epsilon}h^{-\beta}} + \delta^{-(1+\epsilon)}\int |x|^{1+\epsilon} f + \int |L_h| \left(\int_{|x|>\frac{\delta}{2}} f + \int_{|x|>\frac{\delta}{2}} \frac{|L_h|}{\int |L_h|}\right) \\ &\leq e^{-\frac{\epsilon+o(1)}{1+\epsilon}h^{-\beta}} + e^{-1/h^\beta}\int |x|^{1+\epsilon} f + \int |L| \int_{|x|>\frac{\delta}{2}} f + \int_{|x|>\frac{\delta}{2}} |L_h|. \end{aligned}$$

The third term is smaller than a constant times e^{-1/h^β}, and the fourth term is not greater than $\int_{|x|>\delta/(2h)} |L| \leq 2hc/\delta$ for some constant c. This concludes the proof of our claim.

7. References.

M. S. Bartlett, "Statistical estimation of density functions," *Sankhya Series A*, vol. 25, pp. 245–254, 1963.

P. L. Butzer and R. J. Nessel, *Fourier Analysis and Approximation*, Birkhäuser Verlag, Basel and Stuttgart, 1971.

K. B. Davis, "Mean square error properties of density estimates," *Annals of Statistics*, vol. 5, pp. 1025–1030, 1975.

K. B. Davis, "Mean integrated square error properties of density," *Annals of Statistics*, vol. 5, pp. 530–535, 1977.

L. Devroye, "The equivalence of weak, strong and complete convergence in L1 for kernel density estimates," *Annals of Statistics*, vol. 11, pp. 896–904, 1983.

L. Devroye, *A Course in Density Estimation*, Birkhauser Verlag, Basel, 1987.

L. Devroye, "An L1 asymptotically optimal kernel estimate," Technical Report, McGill University, 1987.

L. Devroye, "A universal lower bound for the kernel estimate," *Statistics and Probability Letters*, vol. 8, pp. 419–423, 1989.

L. Devroye and L. Györfi, *Nonparametric Density Estimation. The Lsub1 View*, John Wiley, New York, N.Y., 1985.

L. Devroye and C. S. Penrod, "Distribution-free lower bounds in density estimation," *Annals of Statistics*, vol. 12, pp. 1250–1262, 1984.

U. Haagerup, "Les meilleures constantes de l'inegalité de Khintchine," *Comptes Rendus de l'Académie des Sciences de Paris*, vol. A286, pp. 259–262, 1978.

I. A. Ibragimov and R. Z. Khasminskii, "Estimation of distribution density belonging," *Theory of Probability and its Applications*, vol. 27, pp. 551–562, 1982.

T. Kawata, *Fourier Analysis in Probability Theory*, Academic Press, New York, 1972.

E. Parzen, "On the estimation of a probability density function and the mode," *Annals of Mathematical Statistics*, vol. 33, pp. 1065–1076, 1962.

M. Rosenblatt, "Remarks on some nonparametric estimates of a density function," *Annals of Mathematical Statistics*, vol. 27, pp. 832–837, 1956.

S. J. Szarek, "On the best constants in the Khintchine inequality," *Studia Mathematica*, vol. 63, pp. 197–208, 1976.

R. A. Tapia and J. R. Thompson, *Nonparametric Probability Density Estimation*, Johns Hopkins University Press, Baltimore, 1978.

G. Walter and J. Blum, "Probability density estimation using delta sequences," *Annals of Statistics*, vol. 7, pp. 328–340, 1979.

G. S. Watson and M. R. Leadbetter, "On the estimation of the probability density," *Annals of Mathematical Statistics*, vol. 34, pp. 480–491, 1963.