The Annals of Statistics
1992, Vol. 20, No. 3, 1658–1667

SOME NONASYMPTOTIC BOUNDS FOR L_1 DENSITY ESTIMATION USING KERNELS

BY SOMNATH DATTA

University of Georgia

In this paper we obtain uniform upper bounds for the L_1 error of kernel estimators in estimating monotone densities and densities of bounded variation. The bounds are nonasymptotic and optimal in n, the sample size. For the bounded variation class, it is also optimal wrt an upper bound of the total variation. The proofs employ a one-sided kernel technique and are extremely simple.

1. Introduction. Let M_B be the class of all nonincreasing densities on $[0, 1]$ which are bounded by B. Note that the class is nonempty if and only if $B \geq 1$. Consider the estimation of a density $f \in M_B$ by an estimator f_n based on i.i.d. observations $X_1, X_2, \dots, X_n$ from the distribution with density f. This problem has received considerable attention in the past. The key references are Grenander (1956), who introduced the MLE for this problem, Prakasa Rao (1969), who provided a thorough analysis of the pointwise properties of the MLE, Groeneboom (1985), who obtained the exact convergence of the L_1 risk of the MLE, Devroye (1987), who had an entire chapter on the various methods used so far to estimate densities in M_B, and Birgé (1987a, b), who established the right lower bound for the L_1 minimax risk and obtained a minimax optimal estimator for this class. Also see Birgé (1989), where he proved, among other things, that the MLE is also minimax optimal for this class.

For each n, the L_1 minimax risk over the class M_B is given by

$$R_{n,B} = \inf_{f_n} \sup_{f \in M_B} E \int |f_n - f|. \tag{1.1}$$

The exact order of $R_{n,B}$ is known to be $\log^{1/3}(1 + B)n^{-1/3}$, if $(\log(1 + B))/n$ remains bounded. See Birgé (1987a) or Devroye (1987) for a proof. Birgé (1987b) proposed a modified histogram estimator, that is, a histogram with geometrically increasing interval sizes. He proved that this estimator is minimax optimal for this class, that is,

$$\sup_{f \in M_B} E \int |f_n(x) - f(x)| \, dx \leq c \log^{1/3}(1 + B) n^{-1/3}, \tag{1.2}$$

whenever the ratio $(\log(1 + B))/n$ is bounded. The value of the constant c can be determined from the upper bound of the above ratio.

Received October 1989; revised December 1991.

AMS 1980 *subject classifications*. Primary 62G07; secondary 62C20.

Key words and phrases. Monotone density, density of bounded variation, L_1 estimation, minimax risk, kernel estimator, nonasymptotic bound.

For estimating densities in M_B one can symmetrize a standard kernel estimator to obtain a density on $[0, \infty)$. The symmetrized kernel estimator is defined as

$$f_n(x) = \hat{f}_n(x) + \hat{f}_n(-x), \qquad x > 0, \tag{1.3}$$

where $\hat{f}_n$ is the standard kernel estimator with a symmetric kernel K based on $X_1, \dots, X_n$. The symmetrized kernel estimator satisfies

$$\limsup_{n \to \infty} n^{1/3} \sup_{f \in M_B} E \int |f_n(x) - f(x)|\, dx \le cB^{1/3} \tag{1.4}$$

for some constant c, if the bandwidth $h \propto B^{-2/3} n^{-1/3}$. See Devroye and Györfi (1985) or Devroye (1987) for a proof of (1.4). It shows that $\sup_{f \in M_B} E\int |f_n - f|$ has the right dependency on n but not on B. For this reason, Devroye called the symmetrized kernel estimator asymptotically minimax suboptimal. Note, once again, that the kernel used in the construction of $\hat{f}_n$ is assumed to be symmetric.

In this paper we consider first the standard kernel estimator based on a totally asymmetric kernel. More precisely, let K be a kernel satisfying $K(x) = 0$ for $x > 0$, and let $\hat{f}_n$ be the standard kernel estimator based on K and $X_1, \dots, X_n$. We prove that for $h \propto B^{-2/3} n^{-1/3}$, $\hat{f}_n$ satisfies

$$\sup_{f \in M_B} E \int_0^1 |\hat{f}_n(x) - f(x)|\, dx \le cB^{1/3} n^{-1/3} \quad \text{for all } n, \tag{1.5}$$

for some constant c. Moreover, the actual bound on $\sup_{f \in M_B} E\int_0^1 |\hat{f}_n(x) - f(x)|\, dx$ in terms of n and h in this paper is very similar to the main terms of the bound in Devroye (1987) toward establishing (1.4). Consequently, the best possible constants are comparable. The best value of c in (1.4) given by Devroye (1987) is $(2^{1/3} + 2^{-2/3})(2k_1 k_2)^{1/3}$, whereas the best value of c in (1.5) given by the present paper is $(2^{1/3} + 2^{-2/3})(k_1 k_2)^{1/3}$, where $k_1 = \int |x| K\, dx$, $k_2 = \int K^2\, dx$. (This is justified because if K is a symmetric kernel and $K^* = 2K 1_{(-\infty, 0]}$ is the corresponding one-sided kernel then k_1 is the same for both but k_2 for K^* is twice the k_2 for K.) We also obtain a scale invariant upper bound for individual densities in a larger class, namely, monotone densities on $[0, \infty)$.

Next we obtain similar results for estimators based on a symmetric (two-sided) kernel. The uniform bound for the bias term over the monotone class using a symmetric kernel turns out to be slightly worse than that using the corresponding left-sided one. The excess bias due to a two-sided kernel increases with B.

In Section 3 we extend our method of analysis to another interesting class of densities, namely, densities on $[0, 1]$ which are of bounded variation with total variation bounded by some given $B > 0$. The minimax bound for the bounded variation class turns out to be optimal, that is, it coincides with the minimax bound in the class of all density estimators.

All the proofs are short and simple. None is based on the usual Taylor series arguments. Consequently, we do not need to impose any additional smoothness conditions, such as continuity or differentiability, on the density.

2. Bounds for the monotone class. Let, for $B > 0$, $M_{B,\mathbb{R}+} = \{f\colon f \text{ is a nonincreasing density on } [0,\infty) \text{ with } f(0) \le B\}$ and $X_1, X_2, \dots, X_n$ be i.i.d observations from the distribution with density $f \in M_{B,\mathbb{R}+}$. Let K be a kernel; that is, $K \ge 0$ and $\int K(x)\,dx = 1$. Let $h = h_n$ be a sequence of positive reals decreasing to 0. The standard kernel estimator of f based on K and $X_1, X_2, \dots, X_n$ is defined by

$$\hat{f}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right). \tag{2.1}$$

First we state and prove a lemma which will be used to bound the bias of $\hat{f}_n$.

LEMMA 1. *Let* $K \ge 0$, $\int K(x)\,dx = 1$, $k_1 = \int |x| K(x)\,dx < \infty$, $K1_{(0,\infty)} = 0$. *Let* g *be a nonnegative, nonincreasing function on* $[0,\infty)$, *such that* $\int_0^\infty g(x)\,dx < \infty$. *For* $h > 0$, *let* $g_h(x) = \int_{-\infty}^{0} K(u) g(x - uh)\,du$, $x \ge 0$. *Then* $g_h \le g$ *and* $\int_0^\infty (g(x) - g_h(x))\,dx \le g(0) k_1 h$.

PROOF. Clearly, $g_h(x) \le g(x)$ for all $x \ge 0$, because K is a density and g is nonincreasing. To prove the second assertion, assume, without loss of generality, that $\int g = 1$. Then

$$\begin{aligned}\int_0^\infty (g(x) - g_h(x))\,dx &= 1 - \int_0^\infty \int_{-\infty}^{0} K(u) g(x - uh)\,du\,dx\\ &= 1 - \int_{-\infty}^{0} K(u) \int_{-uh}^{\infty} g(v)\,dv\,du,\end{aligned}$$

using that K vanishes outside $(-\infty, 0]$, the Fubini theorem and a change of variable,

$$= \int_{-\infty}^{0} K(u) \int_0^{-uh} g(v)\,dv\,du \le -g(0) h \int_{-\infty}^{0} u K(u)\,du = g(0) h k_1,$$

since both K and g integrate out to 1, and $g \le g(0)$. $\square$

Let E denote the expectation on $X_1, X_2, \dots, X_n$. Assume that the kernel K is left sided, that is, $K1_{(0,\infty)} = 0$, for the next theorem and the two corollaries following it.

THEOREM 1. *Let* $\hat{f}_n$ *be given by* (2.1). *Then for all* $f \in M_{B,\mathbb{R}+}$,

$$E\int_0^\infty \left|\hat{f}_n(x) - f(x)\right| dx \le \left(\int_0^\infty f^{1/2}(x)\,dx\right)\left(\frac{k_2}{nh}\right)^{1/2} + k_1 B h,$$

where $k_1 = \int |x| K(x)\,dx$, $k_2 = \int K^2(x)\,dx$.

PROOF. As always, we split $E\int|\hat{f}_n - f|$ into two parts:

$$E\int_0^\infty |\hat{f}_n(x) - f(x)|\,dx \le E\int_0^\infty |\hat{f}_n(x) - E\hat{f}_n(x)|\,dx + \int_0^\infty |E\hat{f}_n(x) - f(x)|\,dx$$

$$\overset{\text{definition}}{=} \text{VARIATION} + \text{BIAS}.$$

Now, for fixed x,

$$\begin{aligned} \operatorname{var}(\hat{f}_n(x)) &\le \frac{1}{nh^2}\int_0^\infty K^2\left(\frac{x-y}{h}\right) f(y)\,dy \\ &= \frac{1}{nh}\int_{-\infty}^0 K^2(u) f(x-uh)\,du \end{aligned} \tag{2.2}$$

by the substitution $u = (x-y)/h$ and the fact that K vanishes outside $(-\infty, 0]$. Therefore

$$\operatorname{var}(\hat{f}_n(x)) \le \frac{k_2 f(x)}{nh} \tag{2.3}$$

since f is nonincreasing. Hence, by the Fubini theorem and the Cauchy–Schwarz inequality,

$$\text{VARIATION} \le \int_0^\infty \left(\operatorname{var}(\hat{f}_n(x))\right)^{1/2} dx \le \left(\int_0^\infty f^{1/2}(x)\,dx\right)\left(\frac{k_2}{nh}\right)^{1/2}.$$

Next note that $E\hat{f}_n = f_h$. Therefore, by Lemma 1,

$$\text{BIAS} = \int_0^\infty (f(x) - f_h(x))\,dx \le k_1 B h$$

since $f(0) \le B$. This completes the proof of the theorem. □

COROLLARY 1. *For any $f \in M_{B,\mathbb{R}^+}$, the bound in Theorem 1 is minimized for $h = (k_2(\int f^{1/2})^2/(4nk_1^2B^2))^{1/3}$ and for this choice of h,*

$$E\int_0^\infty |\hat{f}_n(x) - f(x)|\,dx \le (2^{1/3} + 2^{-2/3})(k_1k_2)^{1/3} B^{1/3}\left(\int f^{1/2}\right)^{2/3} n^{-1/3}.$$

Note that the above bound is scale invariant. This bound is useful if $\int f^{1/2} < \infty$. A sufficient condition for this is the finiteness of the second moment of f [see Devroye (1987), Lemma 7.2].

We will continue to use 2 in place of $(2^{1/3} + 2^{-2/3})$ for simplicity. Since $\int f^{1/2} \le 1$ for all $f \in M_B$, we get the following uniform bounds over M_B.

COROLLARY 2. *For all n,*

$$\sup_{f \in M_B} E\int_0^1 |\hat{f}_n(x) - f(x)|\,dx \le \left(\frac{k_2}{nh}\right)^{1/2} + k_1 B h \quad \textit{for all } h > 0.$$

Hence

$$\inf_h \sup_{f \in M_B} E\int_0^1 |\hat{f}_n(x) - f(x)|\,dx \le 2(k_1 k_2)^{1/3} B^{1/3} n^{-1/3}.$$

Next we analyze the performance of a kernel estimator based on a symmetric (two-sided) kernel. The proof is a little more involved and the technique of bounding the variation term works for densities with compact support only. Note that the bound on the bias term is tight [attained for $B > 1$ by $f =$ uniform $(0, B)$ and $K = 1_{[-1,1]}/2$] and 1.5 times larger than that for the corresponding left-sided kernel.

THEOREM 2. *Suppose that the kernel K used in $\hat{f}_n$ is symmetric. Then for all n and h,*

$$\sup_{f \in M_B} E\int_0^1 |\hat{f}_n(x) - f(x)|\,dx \le \left(\frac{k_2}{nh}\right)^{1/2} + \frac{3}{2} k_1 B h.$$

Consequently,

$$\inf_h \sup_{f \in M_B} E\int_0^1 |\hat{f}_n(x) - f(x)|\,dx \le 2(1.5 k_1 k_2)^{1/3} B^{1/3} n^{-1/3}.$$

We need the following lemma to bound the variation term. The result is valid for any density f on $\mathbb{R}$ but useful only if f has support on [0, 1].

LEMMA 2. *The kernel estimator given by* (2.1) *satisfies*

$$E\int_0^1 |\hat{f}_n(x) - E\hat{f}_n(x)|\,dx \le \left(\frac{k_2}{nh}\right)^{1/2}$$

for any density f on $\mathbb{R}$.

PROOF. As in (2.2),

$$\operatorname{var}(\hat{f}_n(x)) \le \frac{1}{nh^2} \int K^2\left(\frac{x-y}{h}\right) f(y)\,dy$$

for any $x \in \mathbb{R}$. Integrating the above pointwise bound over all of $\mathbb{R}$ and using Fubini's theorem, one gets

$$\int \operatorname{var}(\hat{f}_n(x))\,dx \le \frac{k_2}{nh}. \tag{2.4}$$

Another source of (2.4) is Bretagnolle and Huber (1979); see (4.14) of that paper. Now note that the Lebesgue measure restricted to [0, 1] is a probability. Therefore, by two applications of the Cauchy–Schwarz inequality and (2.4), we have

$$E\int_0^1 |\hat{f}_n(x) - E\hat{f}_n(x)|\,dx \le \left(\int_0^1 \operatorname{var}(\hat{f}_n(x))\,dx\right)^{1/2}$$
$$\le \left(\int \operatorname{var}(\hat{f}_n(x))\,dx\right)^{1/2} \le \left(\frac{k_2}{nh}\right)^{1/2}. \qquad \square$$

PROOF OF THEOREM 2. The bound on the variation term is obtained in Lemma 2.

Next, to bound the bias term, realize that

$$E\hat{f}_n(x) = \frac{1}{2}E\hat{f}_{n1}(x) + \frac{1}{2h}\int K_1\left(\frac{y-x}{h}\right)f(y)\,dy,$$

where $\hat{f}_{n1}$ is the density estimator based on $K_1 = 2K1_{(-\infty,0]}$. Therefore

$$\text{(2.5)}\quad \text{BIAS} \le \frac{1}{2}\text{BIAS}(\hat{f}_{n1}) + \frac{1}{2}\int_0^1 \left|\frac{1}{h}\int K_1\left(\frac{y-x}{h}\right)f(y)\,dy - f(x)\right|dx.$$

The first term is no more than $(Bk_1h)/2$ by the result for a left-sided kernel. Next note that

$$\text{(2.6)}\quad \begin{aligned} f_h^*(x) &\overset{\text{definition}}{=} \frac{1}{h}\int K_1\left(\frac{y-x}{h}\right)f(y)\,dy \\ &= \int K_1(u)f(x+uh)\,du \ge f(x)\int_{-x/h}^0 K_1(u)\,du \end{aligned}$$

for all $x \ge 0$. Also, $\int_{-x/h}^0 K_1(u)\,du \le 1$ since K_1 is a density. Therefore

$$\text{(2.7)}\quad \begin{aligned} \int_0^1 |f_h^*(x) - f(x)|\,dx &\le \int_0^1 \left(f_h^*(x) - f(x)\int_{-x/h}^0 K_1(u)\,du\right)dx \\ &\quad + \int_0^1 f(x)\left(1 - \int_{-x/h}^0 K_1(u)\,du\right)dx. \end{aligned}$$

By similar arguments as used before (e.g., in the proof of Lemma 1) both terms on the right-hand side of (2.7) can be shown to be bounded by k_1Bh. The uniform bound now follows by combining this with (2.5) and Lemma 2.

The minimax bound can be obtained by minimizing the uniform bound wrt h. $\square$

We end this section with a few remarks.

REMARK 1. Symmetry of a kernel is a traditional requirement in kernel density estimation. Recently Cline (1988) gave some theoretical justification for this in the context of L_2 estimation.

Generally speaking, a one-sided (asymmetric) kernel gives rise to higher variation than the corresponding symmetric kernel. (It is also reflected by the fact that k_2 for a one-sided kernel is twice the k_2 for the corresponding symmetric kernel.) On the other hand, for nonincreasing densities on $[0, \infty)$, the (integrated) bias of an estimator based on a symmetric (two-sided) kernel will be worse than that based on a left-sided kernel because of the nature of the density near the origin. When n is large and h is not too small, the bias term is the dominating factor and hence the L_1 error using a left-sided kernel will be smaller. The opposite happens when h is sufficiently small. Also note that for a given h (and n), the uniform L_1 bound using a symmetric kernel will exceed that for a left-sided kernel if B is large.

REMARK 2. What happens for optimal (or near optimal) values of h in terms of uniform performance can be seen by comparing the minimax bounds given by Corollary 2 and Theorem 2. In spite of the presence of the factor 1.5, the minimax bound using a symmetric kernel happens to be sharper than the minimax bound using the corresponding left-sided one [or the symmetrization trick of Devroye and Györfi (1985)] because k_2 for a one-sided kernel is twice the k_2 for the associated symmetric kernel.

REMARK 3. For our results so far (including the above remarks) we have taken the range of integration to be the support of the density to be estimated, that is, either $[0, 1]$ or $[0, \infty)$. If instead, one considers the usual L_1 risk, then one needs to integrate the pointwise error over the entire real line. This will only cause an additional amount $e_n = 1 - E\int_0^1 \hat{f}_n(x)\,dx$, which is no more than $k_1 Bh$, for $f \in M_B$ (and also for $f \in M_{B,\mathbb{R}+}$ if K is left sided). More generally, $e_n \le (\sup f) k_1 h$, for any density f on $[0, 1]$. These can be verified by simple Fubini arguments.

REMARK 4. For estimating densities on $[0, 1]$, it may sometimes be desirable to have a density estimator with the same support. One may then consider the restricted (and normalized) kernel estimator $\hat{f}_n^* = \hat{f}_n 1_{[0,1]} / (\int_0^1 \hat{f}_n(x)\,dx)$. This estimator will have L_1 error bound of the same order (as $\hat{f}_n$) because

$$(2.8) \quad \int_0^1 \left| \hat{f}_n^*(x) - f(x) \right| dx \le \int_0^1 \left| \hat{f}_n(x) - f(x) \right| dx + \left(1 - \int_0^1 \hat{f}_n(x)\,dx \right)$$

for all densities f on $[0, 1]$. Alternatively, one may simply use

$$(2.9) \qquad \int_0^1 \left| \hat{f}_n^*(x) - f(x) \right| dx \le 2 \int_0^1 \left| \hat{f}_n(x) - f(x) \right| dx$$

for all densities f on $[0, 1]$. The inequality (2.8) follows from Lemma 4 in Chapter 7 of Devroye and Györfi (1985) or Theorem 1.5 in Devroye (1987). The second inequality follows from the proof of Theorem 1.5 in Devroye (1987).

REMARK 5. Since the symmetrized kernel estimator given by (1.3) satisfies

$$\int |f_n - f| \le \int |\hat{f}_n - f|$$

for all densities f on $[0, 1]$, our Theorem 2 proves (1.4) [which is Theorem 8.5 of Devroye (1987)] via Remark 3. In fact, it holds nonasymptotically.

3. Bounds for the bounded variation class. Next we study the performance of kernel estimators for estimating densities of bounded variation. The minimax bound in this case turns out to be optimal in the class of all density estimators. For the bounded variation class, the same uniform bound holds for all kernels. Therefore a symmetric kernel will be preferable in this case for optimal performance (in order to minimize k_2).

Let V_B stand for the class of all densities on $[0, 1]$ which are of bounded variation on $[0, 1]$ with the total variation not exceeding $B > 0$. Clearly, $M_B \subset V_B$. The following result establishes bounds on the bias term $\int_0^1 |E\hat{f}_n(x) - f(x)|\, dx$, where $\hat{f}_n$ is the standard kernel estimator of f defined by (2.1) and $f \in V_B$.

LEMMA 3. *For any $f \in V_B$ and $h > 0$,*

$$\int_0^1 \left| E\hat{f}_n(x) - f(x) \right| dx \le (B + f(1)) k_1 h \quad \text{if } K \text{ vanishes outside } (-\infty, 0],$$

$$\le (B + f(0)) k_1 h \quad \text{if } K \text{ vanishes outside } [0, \infty).$$

Thus, for any kernel K,

$$\sup_{f \in V_B} \int_0^1 \left| E\hat{f}_n(x) - f(x) \right| dx \le (2B + 1) k_1 h.$$

PROOF. First consider the case when K vanishes outside $(-\infty, 0]$. Let $n(x)$ and $p(x)$ denote the negative and positive variations, respectively, of f on $[0, x]$ for $0 \le x \le 1$. Then f can be expressed as [see Apostol (1974), page 138]

$$f = w - r + f(1), \tag{3.1}$$

where $w(x) = N - n(x)$, $r(x) = P - p(x)$, and $N = n(1)$, $P = p(1)$ denote the total negative and positive variations, respectively. Extend w and r to all of $[0, \infty)$ by defining them as 0 on $(1, \infty)$. Also, let 1^* be the function $1^*(x) = 1$ if $0 \le x \le 1$, and $1^*(x) = 0$ if $x > 1$. Then w, r and 1^* are nonnegative, nonincreasing functions on $[0, \infty)$ satisfying $w \le N$, $r \le P$ and $1^* \le 1$. Moreover, $N + P = V$ is the total variation of f which is no more than B if $f \in V_B$.

For any $x \in [0, 1]$, (3.1) yields

$$E\hat{f}_n(x) = w_h(x) - r_h(x) + f(1) 1_h^*(x).$$

By Lemma 1, $w_h \le w$, $r_h \le r$ and $1_h^* \le 1^*$. Hence

$$\begin{aligned} \text{BIAS} &\le \int_0^\infty (w(x) - w_h(x))\, dx + \int_0^\infty (r(x) - r_h(x))\, dx \\ &\quad + f(1) \int_0^\infty (1^* - 1_h^*(x))\, dx \\ &\le (N + P + f(1)) k_1 h = (V + f(1)) k_1 h \le (B + f(1)) k_1 h \end{aligned}$$

by Lemma 1.

For the case, when $K = 0$ outside $[0, \infty)$, represent f as

$$f = p - n + f(0). \tag{3.2}$$

Using the facts that $0 \le p \le P$, $0 \le n \le N$ and p and n are nondecreasing, and a corresponding version of Lemma 1, the proof follows by similar arguments as above.

Finally, for a general kernel K, express it as $K = \alpha K_1 + (1 - \alpha) K_2$, where K_1 and K_2 are kernels vanishing outside $(-\infty, 0]$ and $[0, \infty)$, respectively. Then the result for K follows from that for K_1 and K_2. □

Combining Lemmas 2 and 3, we get the following theorem.

THEOREM 3. *For any kernel K, the estimator given by* (2.1) *satisfies*

$$\sup_{f \in V_B} \int_0^1 \left| \hat{f}_n(x) - f(x) \right| dx \le (2B + 1) k_1 h + \left(\frac{k_2}{nh} \right)^{1/2} \quad \textit{for all } n \textit{ and } h.$$

Consequently,

$$\inf_h \sup_{f \in V_B} \int_0^1 \left| \hat{f}_n(x) - f(x) \right| dx \le 2 (k_1 k_2)^{1/3} (2B + 1)^{1/3} n^{-1/3} \quad \textit{for all } n.$$

PROOF. The first inequality obtains directly from Lemmas 2 and 3. To get the nonasymptotic minimax bound, minimize the first bound over h. □

REMARK 6. Let W_B be the class of all Lipschitz densities on $[0, 1]$ with Lipschitz constant not exceeding B. It is not hard to check that W_B is nonempty if and only if $B \ge 4$. Suppose that B is bounded away from 4, say $B \ge 5$. Then Theorem 5.4 in Devroye (1987) says that

$$\inf_{f_n} \sup_{f \in W_B} E\left(\int \left| f_n - f \right| \right) \ge (c + o(1)) B^{1/3} n^{-1/3}$$

for some constant c (not depending on B). Since $W_B \subset V_B$, the same minimax lower bound holds for V_B also. Thus the minimax upper bound for V_B using kernel estimators given by Theorem 3 comes to within a constant multiple of the minimax lower bound for V_B. In particular, it shows that kernel estimators (and their restrictions) with proper choice of bandwidth are minimax optimal for V_B.

Recently Engel (1990) obtained an $O(n^{-1/3})$ L_1 rate result for an orthogonal series estimator based on the system of Haar functions when $f \in V_B$. However, his result is for individual densities and is truly asymptotic. In another result he obtained a bound on the main term under additional smoothness condition on the density.

Acknowledgments. I would like to express my sincere thanks to Professor Luc Devroye for his helpful and encouraging comments on an earlier version of the paper. Thanks are also due to an anonymous referee for some very constructive criticisms which led to substantial improvement of the paper.

REFERENCES

APOSTOL, T. M. (1974). *Mathematical Analysis*, 2nd ed. Addison-Wesley, Reading, Mass.

BIRGÉ, L. (1987a). Estimating a density under order restrictions: Nonasymptotic minimax risk. *Ann. Statist.* **15** 995–1012.

BIRGÉ, L. (1987b). On the risk of histograms for estimating decreasing densities. *Ann. Statist.* **15** 1013–1022.

BIRGÉ, L. (1989). The Grenander estimator: A nonasymptotic approach. *Ann. Statist.* **17** 1532–1549.

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

CLINE, D. B. H. (1988). Admissible kernel estimators of a multivariate density. *Ann. Statist.* **16** 1421–1427.

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

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

ENGEL, J. (1990). Density estimation with Haar series. *Statist. Probab. Lett.* **9** 111–117.

GRENANDER, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuar.* **39** 125–153.

GROENEBOOM, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (L. M. Le Cam and R. A. Olshen, eds.) **2** 539–555. IMS, Hayward, Calif.

PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31** 23–36.

DEPARTMENT OF STATISTICS
STATISTICS BUILDING
UNIVERSITY OF GEORGIA
ATHENS, GEORGIA 30602