The Annals of Statistics
1981, Vol. 9, No. 2, 334–346

ASYMPTOTICALLY OPTIMUM KERNELS FOR DENSITY ESTIMATION AT A POINT

BY JEROME SACKS[1] AND DONALD YLVISAKER[2]
Northwestern University and University of California, Los Angeles

Kernel estimation of $f(0)$ is considered where f is a density in some class $\mathscr{F}$ of d-dimensional densities, described in terms of a Taylor series expansion. A sequence of kernels which asymptotically minimizes the maximum mean square error of estimation over $\mathscr{F}$ is given. The shape of the kernel is fixed, the size of the window depends on $f(0)$, and an easily computed estimate is obtained to efficiently adapt the sequence to the unknown value of $f(0)$.

0. Introduction. Let $X_1, \ldots, X_n$ be a sample of n independent observations from a distribution F on R^d with density f. The problem of interest is that of estimating $f(0)$ (or $f(x_0)$ for some x_0). The focus of the paper is on kernel estimation, i.e., on estimates of the form $T_\Gamma = \frac{1}{n}\sum_{i=1}^n \Gamma(X_i)$ where Γ is called the kernel. Estimates of this type were introduced by Rosenblatt [6] and subsequently studied by a number of authors, e.g., Parzen [5] and Epanechnikov [2].

For a general estimate $\hat{f}_n(0)$ of $f(0)$, let the risk be given by

$$Q(f, \hat{f}_n(0), n) = E_f[\hat{f}_n(0) - f(0)]^2. \tag{0.1}$$

Our concern is with the asymptotic behavior of Q as $n \to \infty$ and we produce an easily calculated sequence of estimators which possesses asymptotically optimal properties in terms of the risk Q for certain families of f's. The asymptotic behavior of Q has been previously studied, and the relationship of the present results to those in the literature is most easily seen in the following context. Suppose the class of possible densities on R^1 is

$$\mathscr{F}_1 = \{f \mid f \geq 0, \int f = 1, f \in \mathscr{C}_2, \sup_x f(x) \leq \alpha_1, \sup_x |f''(x)| \leq 1\}. \tag{0.2}$$

For each n let the class of available kernels be

$$\mathscr{G}_{1n} = \{\Gamma \mid \Gamma(x) = b_n^{-1} W(xb_n^{-1}), b_n > 0, b_n \downarrow 0, W \geq 0, \int W = 1, \int xW = 0, \int x^2 W < \infty\}. \tag{0.3}$$

If $\Gamma_n \in \mathscr{G}_{1n}$, $n \geq 1$, it is straightforward to calculate (Rosenblatt [7], Equation (19)) that for $b_n = Kn^{-1/5}$,

$$\sup_{\mathscr{F}_1} Q(f, T_{\Gamma_n}, n) \sim q(b_n, W) = O(n^{-4/5}). \tag{0.4}$$

Epanechnikov [2] showed that $\inf_{b_n, W} q(b_n, W) = q(b_{0n}, W_0) \sim \frac{3}{4}(15)^{-1/5} \cdot \alpha_1^{4/5} n^{-4/5}$, where

$$W_0(x) = \frac{3}{4} 5^{-1/2}\left(1 - \frac{x^2}{5}\right)_+, \qquad b_{0n} = \left(\frac{\alpha_1}{n}\right)^{1/5}\left(\int W_0^2\right)^{1/5} \tag{0.5}$$

Received September, 1978; revised January, 1980.
[1] This author's research was supported in part by NSF Grant MCS 77-01657.
[2] This author's research was supported in part by NSF Grant MCS 77-02121.
AMS 1970 *subject classification.* Primary 62G20; secondary 62G05
Key words and phrases. Density estimation, mean square error, asymptotically minimax kernel estimates.

(This can be seen in [7] but in Equation (21) remove the factor 4 and in Equation (22) replace $2^{3/5}$ by 5/4.)

If

$$\mathscr{F}_2 = \{f \mid f \geq 0, \int f = 1, \sup_x f(x) \leq \alpha_1, \sup_x |f''(x)| \leq 1\}, \tag{0.6}$$

Farrell [3] has shown that the best attainable rate for $\sup_{\mathscr{F}_2} Q(f, T_n, n)$ is $O(n^{-4/5})$ for any sequence $\{T_n\}$ of estimators. Epanechnikov's kernel sequence retains its optimal character over $\mathscr{F}_2$ among nonnegative kernels but it is not minimax when kernels with negative values are permitted. Over $\mathscr{F}_1$ the best attainable rate is also $O(n^{-4/5})$ as recently shown by Stone [9] and again Epanechnikov's kernel loses its optimal character when the restriction to nonnegative kernels is removed.

To clarify the situation further, consider

$$\begin{aligned}\mathscr{F}_3 = \{f \mid f \geq 0, \int f = 1, f(x) = \\ f(0) + xf'(0) + r(x), f(0) \leq \alpha_1, |r(x)| \leq \tfrac{1}{2} x^2 \quad \text{for} \quad |x| \leq s\},\end{aligned} \tag{0.7}$$

where s and α_1 are positive numbers subject to the technical conditions $\int_{-s}^{s} (\alpha_1 + x^2/2) \leq 1$ and $\alpha_1 - s^2/2 \geq 0$. It follows from Theorem 1 in Section 1 that

$$\inf_{\mathscr{G}} \sup_{\mathscr{F}_3} Q(f, T_\Gamma, n) \sim \tfrac{3}{4} (15)^{-1/5} \alpha_1^{4/5} n^{-4/5} \tag{0.8}$$

where $\mathscr{G}$ is the collection of all (measurable) kernels, and that Γ_{0n} defined by (0.5) is asymptotically minimax in the sense that

$$\sup_{\mathscr{F}_3} Q(f, T_{\Gamma_{0n}}, n) \sim \inf_{\mathscr{G}} \sup_{\mathscr{F}_3} Q(f, T_\Gamma, n). \tag{0.9}$$

Thus, after enlarging the class of possible densities from $\mathscr{F}_1$ to $\mathscr{F}_3$, we are able to provide an explicit constant together with a rate in n which bounds the performance of *all* kernel estimates over the class, while demonstrating the asymptotic minimax character of a particular sequence of kernels.

The arguments used to get (0.8) and (0.9) require that 0 be an interior point of the support of f; if 0 is an endpoint of this support, the constant in (0.8) is different and Γ_{0n} does not have the property (0.9). For the latter situation a different sequence $\{\Gamma'_{0n}\}$ can be found to satisfy (0.9), see Section 3.

Theorem 1 is somewhat stronger than the statements at (0.8) and (0.9) indicate. If $\mathscr{F}_\alpha = \{f \in \mathscr{F}_3 \mid f(0) = \alpha\}$ for $0 < \alpha_0 \leq \alpha \leq \alpha_1 < \infty$ and $\{\Gamma_n\}$ is any sequence of kernels with $\sup_{\mathscr{F}_3} Q(f, T_{\Gamma_n}, n) = 0(n^{-4/5})$, then

$$\sup_{\mathscr{F}_\alpha} Q(f, T_{\Gamma_n}, n) \gtrsim \sup_{\mathscr{F}_\alpha} Q(f, T_{\Gamma_{\alpha,n}}, n) \sim q_\alpha^* n^{-4/5}. \tag{0.10}$$

Here q_α^* can be determined (and is given by (0.8) if $\alpha = \alpha_1$), $\Gamma_{\alpha,n}(x) = b_{\alpha,n}^{-1} W_0(x b_{\alpha,n}^{-1})$ where W_0 is given by (0.5), and the sequence $\{b_{\alpha,n}\}$ depends in an explicit way on α. This suggests the use of an adaptive two-stage estimation procedure which at the first stage estimates α by $\hat{\alpha}$, and at the second stage uses the estimate provided by $\Gamma_{\hat{\alpha},n}$. The result of Theorem 2 is that for a suitable $\hat{\alpha}$,

$$\sup_{\mathscr{F}_\alpha} Q(f, \Gamma_{\hat{\alpha},n}, n) \sim q_\alpha^* n^{-4/5} \tag{0.11}$$

for all $\alpha_0 \leq \alpha \leq \alpha_1$. The simplicity of $\hat{\alpha}$ and $\Gamma_{\hat{\alpha},n}$, together with (0.11), serves to recommend this procedure. Woodroofe [11] has shown that adaptive estimates can be useful for density estimation and Theorem 2 confirms this fact.

The theorems proved apply to classes like $\mathscr{F}_3$ for any dimension d and any order Taylor expansion. The bound on r is given a specific form (e.g., (0.7)) but the results can be proved for more general bounds with changes in the constant at (0.8). The estimation of a derivative of f at 0 can be analyzed by the methods in Sections 1 and 2 in order to obtain asymptotically optimum kernels, but we do not carry out such an analysis here. The kernel sequences found in Section 1 and those which can be obtained for estimation of derivatives all possess optimal rates of convergence (Stone [9]), but it is not known whether the constants we produce would allow improvement by using other than kernel estimates.

In Section 3B there is a listing of examples of optimal kernels. Entry 4 there is equivalent to (0.5) by rescaling. This kernel was shown by Epanechnikov to be optimal among d-dimensional kernels which are products of one-dimensional kernels. From Entry 11 in Table 1 it can be seen that the optimality of (0.5) does not carry over to dimension $d > 1$ since the optimal kernel is of the form $(a - b\|x\|^2)_+$ rather than $\prod_{i=1}^d (a - bx_i^2)_+$.

When 0 is an interior point of the support of f, the asymptotically optimal kernels may not give substantial improvement over standard kernels. Thus, for practical purposes, it would seem that one kernel is as good as another, an opinion which is generally held. The calculations in Section 3C give some explicit information on this point and does tend to support such an opinion at least when 0 is an interior point and the dimension of the space isn't large. However, the kernels of Section 3A *do* give explicitly useful methods for estimating the density when 0 is at or near an endpoint in which situation the standard kernels break down because they induce a large "bias" term.

1. The optimal kernels. For a d-tuple of nonnegative integers $j = (j_1, \ldots, j_d)$, set $|j| = j_1 + \ldots + j_d$ and $j! = j_1! \ldots j_d!$. For $x \in R^d$, take $x^j = x_1^{j_1} \ldots x_d^{j_d}$ and $\|x\|^2 = x_1^2 + \ldots + x_d^2$. Let D^j denote the operator $\partial^{j_1} \ldots \partial^{j_d}/(\partial x_1^{j_1} \ldots \partial x_d^{j_d})$.

Suppose m is a fixed positive number and $S_0 = \{x \mid \|x\| \le s_0\}$ is a fixed sphere about 0. For $0 \le \alpha_0 \le \alpha_1 \le \infty$ and $\alpha \in [\alpha_0, \alpha_1]$, let $\mathscr{F}_\alpha$ be the set of probability densities f on R^d which satisfy

$$
\text{(1.1)} \qquad \begin{aligned} f(x) &= \alpha + \textstyle\sum_{1 \le |j| \le k-1} D^j f(0) \dfrac{x^j}{j!} + r(x), && x \in R^d \\ |r(x)| &\le m\|x\|^k, && x \in S_0. \end{aligned}
$$

Thus a function in $\mathscr{F}_\alpha$ is suitably approximated by its Taylor series expansion at the origin. There is the complication at (1.1) that such an f be nonnegative and integrate to 1, and to ensure that these conditions are not overly restrictive we make

ASSUMPTION A. For each $\alpha \in [\alpha_0, \alpha_1]$ there are numbers $d_j = d_j(\alpha)$, $1 \le |j| \le k - 1$, so that $f \in \mathscr{F}_\alpha$ if

$$
\text{(1.2)} \qquad \begin{aligned} f(x) &= \alpha + \textstyle\sum_{1 \le |j| \le k-1} d_j \dfrac{x^j}{j!} + r(x), && x \in R^d \\ |r(x)| &\le m\|x\|^k, && x \in S_0. \end{aligned}
$$

Note that Assumption A is already satisfied if, for example, $\alpha_0 - ms_0^k \ge 0$ and $\int_{S_0} [\alpha_1 + m\|x\|^k] \le 1$ (take $d_j(\alpha) \equiv d_j \equiv 0$). Assumption A guarantees that 0 is an interior point of the support of f—see Section 3A in this connection.

A useful consequence of the fact that $\mathscr{F}_\alpha$ consists of densities is the following.

LEMMA 1. $\operatorname{Sup}_{\mathscr{F}_\alpha} |D^j f(0)| = B_j < \infty$, $1 \le |j| \le k - 1$.

PROOF. Suppose $B_j = +\infty$ for some j. Then there is a sequence $\{f_n\}$ in $\mathscr{F}_\alpha$ and a j_0, $1 \le |j_0| \le k - 1$, so that $D^{j_0} f_n(0) \to \pm\infty$, $|D^j f_n(0)/D^{j_0} f_n(0)| \le 1$ all j, n, and $\lim_{n\to\infty}(D^j f_n(0)/D^{j_0} f_n(0)) = \delta_j$ all j.

Suppose $D^{j_0}f_n(0) \to \infty$. Then for $x \in S_0$,

$$0 \le \lim_{n\to\infty} \frac{f_n(x)}{D^{j_0}f_n(0)} = \sum_{1\le|j|\le k-1} \delta_j \frac{x^j}{j!}$$

while Fatou's lemma gives

$$0 \le \int_{S_0} \sum_{1\le|j|\le k-1} \delta_j \frac{x^j}{j!} \le \lim_{n\to\infty} \int \frac{f_n(x)}{D^{j_0}f_n(0)} = 0.$$

Hence all δ_j must be zero. This contradicts $\delta_{j_0} = 1$ and the lemma follows from a similar argument applied to the case $D^{j_0}f_n(0) \to -\infty$.

Now consider kernel estimation of $f(0)$ assuming $f \in \mathscr{F} = \cup_{\alpha=\alpha_0}^{\alpha_1} \mathscr{F}_\alpha$. That is, if $X_1, \ldots, X_n$ is a sample from f, take $\tilde{f}(0) = \frac{1}{n}\sum_{i=1}^n \Gamma(X_i)$ for some (measurable) kernel Γ. Set

$$\begin{aligned} Q_{\alpha,n}(\Gamma) &= \sup_{\mathscr{F}_\alpha} E\left(\frac{1}{n}\sum_{i=1}^n \Gamma(X_i) - f(0)\right)^2 \\ &= \sup_{\mathscr{F}_\alpha}\left\{E\left(\frac{1}{n}\sum_{i=1}^n \Gamma(X_i) - \int \Gamma f\right)^2 + \left(\int \Gamma f - f(0)\right)^2\right\} \qquad (1.3)\\ &= \sup_{\mathscr{F}_\alpha}\left\{\frac{1}{n}\int \Gamma^2 f - \frac{1}{n}\left(\int \Gamma f\right)^2 + \left(\int \Gamma f - f(0)\right)^2\right\}, \end{aligned}$$

$$Q_n(\Gamma) = \sup_{\alpha_0\le\alpha\le\alpha_1} Q_{\alpha,n}(\Gamma). \qquad (1.4)$$

Note that the infimum over Γ's of $Q_{\alpha,n}(\Gamma)$ is 0 since one could select the kernel Γ with $\Gamma(x) \equiv \alpha$. To avoid such unsatisfactory estimates, let $\Delta_n^2 = \Delta^2 n^{-2k/(2k+d)}$ and set

$$q_{\alpha,n} = \inf_{\{\Gamma | Q_n(\Gamma) \le \Delta_n^2\}} Q_{\alpha,n}(\Gamma), \qquad (1.5)$$

Our results pertain to the asymptotic behavior as $n \to \infty$ of $q_{\alpha,n}$ and it will follow from Lemma 2 that the condition $Q_n(\Gamma) \le \Delta_n^2$ is satisfied for some Γ if Δ is taken large enough.

The first objective is an asymptotic upper bound on $q_{\alpha,n}$. This is accomplished by evaluating $Q_{\alpha,n}(\Gamma)$ along a specific sequence $\{\hat{\Gamma}_n\}$. Let $\{\Gamma_n\}$ be a sequence of kernels with Γ_n supported on $S_n = \{x \,|\, \|x\| \le s_n\}$, $s_0 \ge s_n \to 0$, and satisfying

$$\int \Gamma_n = 1, \int \Gamma_n x^j = 0, \quad 1 \le |j| \le k-1, \quad n \ge 1. \qquad (1.6)$$

For a given $\epsilon > 0$, if n is large enough, $|f(x) - \alpha| \le \epsilon$ on S_n for $f \in \mathscr{F}_\alpha$ (Lemma 1). Then

$$\begin{aligned} Q_{\alpha,n}(\Gamma_n) &\le \sup_{\mathscr{F}_\alpha}\left\{\frac{\alpha+\epsilon}{n}\int_{S_n} \Gamma_n^2 - \frac{(\alpha-\epsilon)^2}{n} + \left(\int_{S_n} \Gamma_n r\right)^2\right\} \qquad (1.7)\\ &\le \frac{\alpha+\epsilon}{n}\int \Gamma_n^2 - \frac{(\alpha+\epsilon)^2}{n} + \left(\int |\Gamma_n| m \|x\|^k\right)^2 \end{aligned}$$

LEMMA 2. *For all sufficiently large* Δ, $q_{\alpha,n} \le q_\alpha^* n^{-2k/(2k+d)}(1 + o(1))$ *where* q_α^* *is given at* (1.9).

PROOF. Write

$$M(x) = \left(\frac{m^2 n}{\alpha + \epsilon}\right)^{1/2} \|x\|^k = \rho \|x\|^k$$

and set

$$\hat{Q}(\Gamma) = \int \Gamma^2 + \left(\int |\Gamma| M\right)^2.$$

The problem of minimizing $\hat{Q}$ over Γ's satisfying (1.6) is solved in [8]; the solution is

$$\hat{\Gamma}(x) = \left(\sum_{0\le|j|\le k-1} b_j x^j - \lambda M(x)\right)_+ - \left(\sum_{0\le|j|\le k-1} b_j x^j + \lambda M(x)\right)_- .$$

where the coefficients b_j, λ are determined from (1.6) and the additional equation $\int |\hat{\Gamma}| M = \lambda$. Moreover, $\hat{Q}(\hat{\Gamma}) = b_0$. The situation is better described as follows: Make the change of variables

$$x_i = \rho^{-2/(2k+d)} u_i, \qquad b_j = \rho^{2(|j|+d)/(2k+d)} c_j, \qquad \lambda = \theta \rho^{d/(2k+d)}$$

and let

(1.8a) $$\hat{G}(u) = (\textstyle\sum c_j u^j - \theta \| u \|^k)_+ - (\sum c_j u^j + \theta \| u \|^k)_-.$$

The equations which determine the c_j's and θ are

(1.8b) $$\int \hat{G} = 1, \quad \int \hat{G} u^j = 0 \quad \text{for} \quad 1 \le |j| \le k-1, \quad \int \hat{G} \| u \|^k = \theta.$$

Then $\hat{\Gamma}(x) = \hat{\Gamma}_n(x) = \rho^{2d/(2k+d)} \hat{G}(\rho^{2/(2k+d)} x)$ and $\hat{Q}(\hat{\Gamma}_n) = b_0 = (m^2 n/(\alpha + \epsilon))^{d/(2k+d)} c_0$. Since $\hat{\Gamma}_n$ has bounded support and satisfies (1.6), it follows from (1.7) that

$$Q_{\alpha,n}(\hat{\Gamma}_n) \le \frac{\alpha + \epsilon}{n} \hat{Q}(\hat{\Gamma}_n) - \frac{(\alpha - \epsilon)^2}{n} = c_0 (\alpha + \epsilon)^{2k/(2k+d)} m^{2d/(2k+d)} n^{-2k/(2k+d)} - \frac{(\alpha - \epsilon)^2}{n}.$$

Now $b_0 = b_0(\epsilon)$ is a continuous function of (the arbitrary) ϵ. Therefore, if c_0^* is determined from (1.8) with $\epsilon = 0$,

$$Q_{\alpha,n}(\hat{\Gamma}_n) \le c_0^* \alpha^{2k/(2k+d)} m^{2d/(2k+d)} n^{-2k/(2k+d)} (1 + o(1)).$$

It is straightforward to show that $Q_n(\hat{\Gamma}_n) = O(n^{-2k/(2k+d)})$ and thus for suitably large Δ in (1.5), the desired result holds with

(1.9) $$q_\alpha^* = c_0^* \alpha^{2k/(2k+d)} m^{2d/(2k+d)}.$$

The main result (Theorem 1 below) establishes the optimum rate achievable by kernel estimates as the one determined in Lemma 2. Examples of $\hat{G}$ satisfying (1.8a) and (1.8b) are given in Section 3.

THEOREM 1. *Suppose Assumption* A *holds. Then for all sufficiently large* Δ *and all* $\alpha \in [\alpha_0, \alpha_1]$,

$$q_{\alpha,n} \sim q_\alpha^* n^{-2k/(2k+d)}.$$

Moreover, if $\hat{G}$ *satisfies* (1.8a) *and* (1.8b) *with* $\epsilon = 0$ *and* $\hat{\Gamma}_{\alpha,n}(x) = \rho^{2d/(2k+d)} \hat{G}(x \rho^{2/(2k+d)})$, *then*

$$Q_{\alpha,n}(\hat{\Gamma}_{\alpha,n}) \sim q_{\alpha,n}.$$

PROOF. Because of Lemma 2 it is necessary to show that $q_{\alpha,n} \ge q_\alpha^* n^{-2k/(2k+d)}(1 + o(1))$. To accomplish this we first obtain a lower bound on $q_{\alpha,n}$ (see (1.16)). An analysis of the bound then gives the desired result.

Let $\{\Gamma_n\}$ be a sequence of kernels satisfying $Q_n(\Gamma_n) \le \Delta_n^2$ and therefore

$$\left|\int \Gamma_n f - f(0)\right| \le \Delta_n, \qquad f \in \mathscr{F}. \tag{1.10}$$

Take f at (1.2) with $\alpha = \alpha_0$, $r(x) \equiv 0$ on S_0 and $\int_{S_o} f = (1-p) < 1$. Then $g = (1+\tau)f$ on S_0 is in $\mathscr{F}_{\alpha_{o}(1+\tau)}$ for τ positive and sufficiently small. Let $\bar{f}$ be any density with support on S_o^c. Extend f to $p\bar{f}$ on S_o^c and extend g to $(p(1+\tau)-\tau)\bar{f}$ on S_o^c. Then with the use of (1.10)

$$\begin{aligned}(2+\tau)\Delta_n &\ge \left|(1+\tau)\left(\int \Gamma_n f - f(0)\right) - \left(\int \Gamma_n g - g(0)\right)\right| \\ &= \left|p(1+\tau)\int_{S_o^c} \Gamma_n \bar{f} - (p(1+\tau)-\tau)\int_{S_o^c} \Gamma_n \bar{f}\right| = \tau\left|\int_{S_o^c} \Gamma_n \bar{f}\right|.\end{aligned}$$

Since $\bar{f}$ is an arbitrary density on S_o^c it follows that $\operatorname{ess\,sup}_{S_o^c}|\Gamma_n| \le ((2+\tau)/\tau)\Delta_n$. Write

$$\left|\left|\int_{S_o^c} \Gamma_n f\right| - \left|\int_{S_o} \Gamma_n f - f(0)\right|\right| \le \left|\int \Gamma_n f - f(0)\right|$$

and find

$$\left|\int_{S_o} \Gamma_n f - f(0)\right| \le \left(\frac{2+\tau}{\tau} + 1\right)\Delta_n = 2\left(\frac{1+\tau}{\tau}\right)\Delta_n \qquad \text{for } f \in \mathscr{F}. \tag{1.11}$$

Given $\epsilon > 0$, choose $S = \{x \mid \|x\| \le s\} \subset S_0$ so that

$$|f(x) - \alpha| \le \epsilon \qquad \text{on } S, \quad f \in \mathscr{F}_\alpha. \tag{1.12}$$

This is possible by Lemma 1. From (1.10)

$$\left|\int \Gamma_n f\right| \le \Delta_n + \alpha, \qquad f \in \mathscr{F}_\alpha, \quad n \ge 1. \tag{1.13}$$

Let $\bar{f}$ be a density concentrated on S_o^c and write $\int \Gamma_n \bar{f} = \gamma_n$ so that $\gamma_n \in [\operatorname{ess\,inf}_{S_o^c} \Gamma_n, \operatorname{ess\,sup}_{S_o^c} \Gamma_n]$. If f is any density in $\mathscr{F}_\alpha$ which is conditionally $\bar{f}$ on S_o^c, then

$$\begin{aligned}\left|\int \Gamma_n f - f(0)\right| &= \left|\int \Gamma_n f - f(0) + \left(1 - \int_{S_o} f\right)\gamma_n\right| \\ &= \left|\int_{S_o} (\Gamma_n - \gamma_n) f + \gamma_n - f(0)\right|.\end{aligned} \tag{1.14}$$

Take $r(x) = \pm I_S \operatorname{sgn}(\Gamma_n - \gamma_n) m \|x\|^k$ in (1.2) and use (1.14) to obtain

$$\begin{aligned}\sup_{\mathscr{F}_\alpha}\left|\int \Gamma_n f - f(0)\right| \ge \max\Bigg\{&\left|\int_{S_o} (\Gamma_n - \gamma_n)\left(\alpha + \sum \frac{d_j}{j!} x^j\right)\right. \\ &\left. + \int_S |\Gamma_n - \gamma_n| m \|x\|^k + \gamma_n - f(0)\right|, \\ &\left|\int_{S_o} (\Gamma_n - \gamma_n)\left(\alpha + \sum \frac{d_j}{j!} x^j\right)\right.\end{aligned} \tag{1.15}$$

$$- \int_S \left| \Gamma_n - \gamma_n \right| m \| x \|^k + \gamma_n - f(0) \Big| \Big\}$$

$$\geq \int_S \left| \Gamma_n - \gamma_n \right| m \| x \|^k.$$

It follows from (1.12), (1.13) and (1.15) that

$$Q_{\alpha,n}(\Gamma_n) \geq \sup_{\mathscr{F}_\alpha} \left\{ \frac{1}{n} \int_S \Gamma_n^2 f - \frac{1}{n} \left(\int \Gamma_n f \right)^2 + \left(\int \Gamma_n f - f(0) \right)^2 \right\}$$

$$(1.16) \qquad \geq \frac{\alpha - \epsilon}{n} \int_S \Gamma_n^2 - \frac{(\Delta + \alpha)^2}{n} + \left(\int_S \left| \Gamma_n - \gamma_n \right| m \| x \|^k \right)^2$$

$$= \frac{\alpha - \epsilon}{n} \int_S (\Gamma_n - \gamma_n)^2 + \left(\int_S \left| \Gamma_n - \gamma_n \right| m \| x \|^k \right)^2 + O\left(\frac{1}{n}\right),$$

where the last equality follows simply from the fact that $|\gamma_n| \leq ((2 + \tau)/\tau)\Delta_n$ which was noted before (1.11).

If $\tilde{\Gamma}_n = I_S(\Gamma_n - \gamma_n)$ satisfied (1.6) for each n, the result of Theorem 1 would follow easily from (1.16). It will be shown that (1.6) is nearly satisfied by $\tilde{\Gamma}_n$ (see (1.23)), and that this is sufficient to conclude the proof.

The first step is to get the estimates at (1.21) and (1.22). Take $r(x) = I_{S_o - S}(\operatorname{sgn} \Gamma_n) m \| x \|^k$ in (1.2) and use the triangle inequality in conjunction with (1.11) to get

$$(1.17) \qquad \int_{S_o - S} |\Gamma_n| m \| x \|^k \leq 2\left(\frac{1 + \tau}{\tau}\right)\Delta_n.$$

For $0 \leq |j| \leq k - 1$ and $x \in S_o - S$ write

$$|x^j| \leq |\max x_\iota|^{|j|} \leq s_o^{|j|} = \left(\frac{s_o^{|j|}}{ms^k}\right) ms^k \leq \left(\frac{s_o^{|j|}}{ms^k}\right) m \| x \|^k,$$

and use (1.17) to get

$$(1.18) \qquad \left| \int_{S_o - S} \Gamma_n x^j \right| \leq \left(\frac{s_o^{|j|}}{ms^k}\right) 2 \cdot \left(\frac{1 + \tau}{\tau}\right)\Delta_n, \qquad 0 \leq |j| \leq k - 1.$$

It is an easily checked consequence of Assumption A that for any $1 \leq |j_o| \leq k - 1$ there are densities $f \in \mathscr{F}_\alpha$ with

$$(1.19) \qquad f(x) = \alpha + \sum_{1 \leq |j| \leq k-1} d_j \frac{x^j}{j!} \pm \delta x^{j_o}, \qquad x \in S_o,$$

provided δ is chosen sufficiently small. Use the f's at (1.19) in (1.11) and the triangle inequality to find

$$(1.20) \qquad \left| \int_{S_o} \Gamma_n x^j \right| \leq \delta^{-1} 2\left(\frac{1 + \tau}{\tau}\right)\Delta_n, \qquad 1 \leq |j_o| \leq k - 1.$$

From (1.18) and (1.20) it follows that

$$(1.21) \qquad \left| \int_S \Gamma_n x^j \right| \leq \left(\frac{s_o^{|j|}}{ms^k} + \delta^{-1}\right) 2\left(\frac{1 + \tau}{\tau}\right)\Delta_n, \qquad 1 \leq |j| \leq k - 1.$$

Finally, use (1.2) in (1.11) with $r \equiv 0$, apply (1.18) and (1.20) and get

$$(1.22) \qquad \left| \int_S \Gamma_n - 1 \right| \leq \alpha^{-1} 2 \cdot \left(\frac{1 + \tau}{\tau}\right)\Delta_n \left(1 + \frac{\alpha}{ms^k} + \delta^{-1} \sum_{1 \leq |j| \leq k-1} \frac{|d_j|}{j!}\right).$$

Let $\tilde{\Gamma}_n = I_S(\Gamma_n - \gamma_n)$ and $g_j = \int \tilde{\Gamma}_n x^j$, $0 \le |j| \le k - 1$. It follows directly from (1.21), (1.22) and the fact that $\gamma_n = O(\Delta_n) = O(n^{-k/(2k+d)})$ that

$$g_o = \int \tilde{\Gamma}_n = 1 + O(n^{-k/(2k+d)}), \qquad g_j = \int \tilde{\Gamma}_n x^j = O(n^{-k/(2k+d)}), \qquad 1 \le |j| \le k-1. \tag{1.23}$$

Take $M(x) = (m^2 n/(\alpha - \epsilon))^{1/2} \| x \|^k = \tilde{\rho} \| x \|^k$ and again set

$$\hat{Q}(\Gamma) = \int \Gamma^2 + \left(\int |\Gamma| M \right)^2.$$

The problem of minimizing $\hat{Q}$ over Γ's satisfying (1.23) has been solved in [8] with the result that

$$\tilde{\Gamma}(x) = (\textstyle\sum_{0 \le |j| \le k-1} \tilde{b}_j x^j - \tilde{\lambda} M(x))_+ - (\sum_{0 \le |j| \le k-1} \tilde{b}_j x^j + \tilde{\lambda} M(x))_-,$$

where the coefficients $\tilde{b}_j$, $\tilde{\lambda}$ are determined from the equations

$$\int \tilde{\Gamma} x^j = g_j, \qquad 0 \le |j| \le k - 1, \qquad \int |\tilde{\Gamma}| M = \tilde{\lambda}.$$

Moreover, $\hat{Q}(\tilde{\Gamma}) = \sum_{0 \le |j| \le k-1} g_j \tilde{b}_j$. Make the change of variables

$$x_i = \tilde{\rho}^{-2/(2k+d)} u_i, \qquad \tilde{b}_j = \tilde{\rho}^{2(|j|+d)/(2k+d)} \tilde{c}_j,$$
$$\tilde{\lambda} = \tilde{\theta} \tilde{\rho}^{d/(2k+d)}, \qquad h_j = \tilde{\rho}^{2|j|/(2k+d)} g_j$$

and set $\hat{G}(u) = (\sum \tilde{c}_j u^j - \tilde{\theta} \| u \|^k)_+ - (\sum \tilde{c}_j u^j + \tilde{\theta} \| u \|_k)_-$. The equations that determine $\tilde{c}_j$, $\tilde{\theta}$ are

$$\int \hat{G} u^j = h_j, \qquad 0 \le |j| \le k - 1, \qquad \int \hat{G} \| u \|^k = \tilde{\theta} \tag{1.24}$$

From (1.23), $h_o = g_o = 1 + O(n^{-k/(2k+d)})$ and $h_j = \tilde{\rho}^{2|j|/(2k+d)} g_j = O(n^{(|j|-k)/(2k+d)})$, $1 \le |j| \le k - 1$. Then if the $\tilde{c}_j$ are determined from (1.24) and c_j's are determined by (1.8) with $\rho = \tilde{\rho}$ there, $|\tilde{c}_j - c_j| = o(1)$ for all j. Therefore

$$\hat{Q}(\tilde{\Gamma} = \sum g_j \tilde{b}_j = \tilde{\rho}^{2d/(2k+d)} \sum g_j \tilde{c}_j = (c_o + o1))\tilde{\rho}^{2d/(2k+d)}.$$

From (1.16)

$$Q_{\alpha,n}(\Gamma_n) \le \frac{\alpha - \epsilon}{n} \hat{Q}(\Gamma_n - \gamma_n) + O\left(\frac{1}{n}\right)$$
$$\ge \frac{\alpha - \epsilon}{n}(c_o + o(1))\left(\frac{m^2 n}{\alpha - \epsilon}\right)^{d/(2k+d)} + O\left(\frac{1}{n}\right).$$

Now $c_o = c_o(\epsilon)$ is continuous in (the arbitrary) ϵ so this lower bound, together with Lemma 2, establishes the theorem.

REMARK 1. A slightly simpler approach to the result in Theorem 1 is possible if the class of kernels used is $\mathscr{G}_o = \{\Gamma \mid 0 \in [\operatorname{ess\,inf}_{S_0} \Gamma, \operatorname{ess\,sup}_{S_0} \Gamma]\}$. The kernel Γ with $\Gamma(x) \equiv \alpha$ is then automatically ruled out and if $q_{\alpha,n}$ is defined by

$$q_{\alpha,n} = \inf_{\mathscr{G}_o} Q_{\alpha,n}(\Gamma), \tag{1.5'}$$

the proof of the theorem can proceed as before with $\gamma_n = 0$, and it does not require the argument leading to (1.11).

COROLLARY. *If Assumption* A *holds, then*

$$\inf_{\mathscr{G}} \sup_{\alpha \le \alpha_1} Q_{\alpha,n}(\Gamma) \sim Q_{\alpha_1,n}(\hat{\Gamma}_{\alpha_1,n}).$$

PROOF. It is easy to show that q_α^* increases with α and then, from Theorem 1, that

$$\sup_{\alpha\in[\alpha_0,\alpha_1]} Q_{\alpha,n}(\Gamma) \ge \sup_{\alpha\in[\alpha_0,\alpha_1]} q_\alpha^* n^{-2k/(2k+d)}(1+o(1))$$
$$= q_{\alpha_1}^* n^{-2k/(2k+d)}(1+o(1)).$$

If $0 \le \alpha \le \bar{\alpha}$, one gets from (1.7) that

$$Q_{\alpha,n}(\hat{\Gamma}_{\bar{\alpha},n}) \le \frac{\bar{\alpha}+\epsilon}{n}\int \hat{\Gamma}^2_{\bar{\alpha},n} + \left(\int |\hat{\Gamma}_{\bar{\alpha},n}|\, m \|x\|^k\right)^2 + O\left(\frac{1}{n}\right)$$
$$\le q_{\bar{\alpha},n}(1+o(1)),$$

and this is enough to establish the corollary.

2. Adaptive estimate. The asymptotically optimum sequence $\{\hat{\Gamma}_{\alpha,n}\}$ given in Theorem 1 depends on the unknown value of $\alpha = f(0)$. In this section we show that a two-stage procedure which first estimates α can give the convergence rate associated with $\hat{\Gamma}_{\alpha,n}$ for all $\alpha_0 \le \alpha \le \alpha_1$.

Suppress the dependence on n and write $\hat{\Gamma}_{\alpha,n} = \hat{\Gamma}_\alpha$ and $\hat{f}_\alpha(0) = \frac{1}{n}\sum_{i=1}^n \hat{\Gamma}_\alpha(X_i)$. A two-stage adaptive procedure is defined as follows: Take a starting value $\bar{\alpha} \in [\alpha_0, \alpha_1]$, compute $\hat{\alpha}$ by

$$\begin{aligned}\hat{\alpha} &= \alpha_0 && \text{if } \hat{f}_{\bar{\alpha}}(0) \le \alpha_0\\ &= \hat{f}_{\bar{\alpha}}(0) && \text{if } \hat{f}_{\bar{\alpha}}(0) \in [\alpha_0, \alpha_1]\\ &= \alpha_1 && \text{if } \hat{f}_{\bar{\alpha}}(0) \ge \alpha_1,\end{aligned}$$

and then define $\hat{f}_{\hat{\alpha}}(0) = \frac{1}{n}\sum_{i=1}^n \hat{\Gamma}_{\hat{\alpha}}(X_i)$ to be the estimate of $f(0)$.

THEOREM 2. $\operatorname{Sup}_{\mathscr{F}_\alpha} E(\hat{f}(0) - f(0))^2 = q_\alpha^* n^{-2k/(2k+d)}(1+o(1))$ *for all* $\alpha \in [\alpha_0, \alpha_1]$ *and any starting value* $\bar{\alpha} \in [\alpha_0, \alpha_1]$.

PROOF. It is enough to show that

$$\sup_{\mathscr{F}_\alpha} E(\hat{f}(0) - \hat{f}_\alpha(0))^2 = o(n^{-2k/(2k+d)}), \qquad \alpha \in [\alpha_0, \alpha_1] \tag{2.1}$$

because of the asymptotic optimality of $\hat{f}_\alpha(0)$ proved in Theorem 1.

Consider the function $\hat{G}$ defined at (1.8a) and (1.8b). $\hat{G}$ vanishes outside some sphere $\|x\| \le \sigma_1$, $|\hat{G}(x)|$ is bounded by some constant σ_2 for all $x \in R^d$ and, for some constant σ_3, $|\hat{G}(x) - \hat{G}(y)| \le \sigma_3 \|x - y\|$ for all $x, y \in R^d$. If $\beta \in [\alpha_0, \alpha_1]$ and $\hat{\Gamma}_\beta(x) = (m^2 n/(\beta))^{d/(2k+d)} \cdot \hat{G}((m^2 n/(\beta))^{1/(2k+d)} x)$, then $\hat{\Gamma}_\beta(x)$ vanishes outside the sphere $\|x\| \le (\alpha_1/(m^2 n))^{1/(2k+d)}\sigma_1 = s_n$, say. For $\beta, \gamma \in [\alpha_0, \alpha_1]$,

$$\begin{aligned}|\hat{\Gamma}_\beta(x) - \hat{\Gamma}_\gamma(x)| \le{}& \left(\frac{m^2 n}{\beta}\right)^{d/(2k+d)} \left|\hat{G}\left(\left(\frac{m^2 n}{\beta}\right)^{1/(2k+d)} x\right) - \hat{G}\left(\left(\frac{m^2 n}{\gamma}\right)^{1/(2k+d)} x\right)\right|\\ &+ \left|\left(\frac{m^2 n}{\beta}\right)^{d/(2k+d)} - \left(\frac{m^2 n}{\gamma}\right)^{d/(2k+d)}\right| \left|\hat{G}\left(\left(\frac{m^2 n}{\gamma}\right)^{1/(2k+d)} x\right)\right|\\ \le{}& \sigma_4 n^{d/(2k+d)} |\beta - \gamma|\end{aligned} \tag{2.2}$$

for $\|x\| \le s_n$. For large n, $s_n \le s_o$ and then (2.2) implies

$$
\begin{aligned}
(2.3)\qquad \left|\int (\hat\Gamma_\beta - \hat\Gamma_\gamma) f\right| &= \left|\int_{\|x\|\le s_n} (\hat\Gamma_\beta - \hat\Gamma_\gamma) f\right| = \left|\int_{\|x\|\le s_n} (\hat\Gamma_\beta - \hat\Gamma_\gamma) r\right| \\
&\le \sigma_4 n^{d/(2k+d)} |\beta - \gamma| \int_{\|x\|\le s_n} m \|x\|^k \le \sigma_5 n^{-k/(2k+d)} |\beta - \gamma|
\end{aligned}
$$

for any $f \in \mathscr{F}_\alpha$.

For a fixed $f \in \mathscr{F}_\alpha$, write

$$
(2.4)\qquad E(\hat f(0) - \hat f_\alpha(0))^2 = \frac{1}{n^2} \sum_i \sum_j E(\hat\Gamma_{\hat\alpha}(X_i) - \hat\Gamma_\alpha(X_i))(\hat\Gamma_{\hat\alpha}(X_j) - \hat\Gamma_\alpha(X_j))
$$

and consider an individual term in the sum, with $i \ne j$. Let $\tilde\alpha = (1/(n-2)) \sum_{l\ne i,j} \hat\Gamma_{\tilde\alpha}(X_l)$ if this average is in $[\alpha_0, \alpha_1]$, $\tilde\alpha = \alpha_0$ if the average is $<\alpha_0$, and $\tilde\alpha = \alpha_1$ otherwise. Then

$$
\begin{aligned}
&E(\hat\Gamma_{\hat\alpha}(X_i) - \hat\Gamma_\alpha(X_i))(\hat\Gamma_{\hat\alpha}(X_j) - \hat\Gamma_\alpha(X_j)) \\
(2.5)\qquad &\quad = E\{[(\hat\Gamma_{\tilde\alpha}(X_i) - \hat\Gamma_\alpha(X_i)) + (\hat\Gamma_{\hat\alpha}(X_i) - \hat\Gamma_{\tilde\alpha}(X_i))] \\
&\qquad\qquad \times [(\hat\Gamma_{\tilde\alpha}(X_j) - \hat\Gamma_\alpha(X_j)) + (\hat\Gamma_{\hat\alpha}(X_j) - (\hat\Gamma_{\tilde\alpha}(X_j))].
\end{aligned}
$$

To evaluate $E(\hat\Gamma_{\tilde\alpha}(X_k) - \hat\Gamma_\alpha(X_i))(\hat\Gamma_{\tilde\alpha}(X_j) - \hat\Gamma_\alpha(X_j))$, condition on X_l, $l \ne i, j$, and the expectation is bounded by $\sigma_5^2 n^{-2k/(2k+d)} |\tilde\alpha - \alpha|^2$ because of (2.3). According to Lemma 2, $E(\hat\alpha - \alpha)^2 = O(n^{-2k/(2k+d)})$ so the unconditional expectation is $O(n^{-4k/(2k+d)})$. One can calculate directly with the use of (2.2) that $|\hat\alpha - \tilde\alpha| \le \sigma_6 n^{-2k/(2k+d)}$. Observe that if $I_i = I_{\{\|X_i\|\le s_n\}}$,

$$
\begin{aligned}
&|E(\hat\Gamma_{\hat\alpha}(X_i) - \hat\Gamma_{\tilde\alpha}(X_i))(\hat\Gamma_{\hat\alpha}(X_j) - \hat\Gamma_{\tilde\alpha}(X_i))| \\
&\qquad = |E(\hat\Gamma_{\hat\alpha}(X_i) - \hat\Gamma_{\tilde\alpha}(X_i))(\hat\Gamma_{\hat\alpha}(X_j) - \hat\Gamma_{\tilde\alpha}(X_j)) I_i I_j| \\
&\qquad \le \sigma_4^2 n^{2d/(2k+d)} \sigma_6^2 n^{-4k/(2k+d)} E I_i I_j \le \sigma_7 n^{-4k/(2k+d)}
\end{aligned}
$$

with the last inequality coming from the estimate $\int_{\|x\|\le s_n} f \le \sigma_8 n^{-d/(2k+d)}$, a consequence of Lemma 1. Therefore the term at (2.5) is $O(n^{-4k/(2k+d)})$. Since the diagonal terms in (2.4) can be bounded in the same way, (2.4) is bounded by a term of order $n^{-4k/(2k+d)} = o(n^{-2k/(2k+d)})$, uniformly over $\mathscr{F}_\alpha$. The theorem is proved.

3. Remarks and examples.

A. Estimating f at or near a boundary point. Under Assumption A of Section 1, 0 is an interior point of the support of f. If the S_0 at (1.1) and (1.2) is taken to be a "one-sided" interval it is possible to obtain corresponding results when 0 is a boundary point of this support set. The situation is most easily described in dimension 1, with 0 a left endpoint of the support of f, say. The asymptotics governing the development to (1.8) will now produce integrals over $[0, \infty)$ rather than $(-\infty, \infty)$ and the resulting $\hat G$ will differ accordingly (see the table in Section 3B). Similar changes occur in higher dimensional settings and there one can find asymptotically optimum kernels that depend on the shape of the boundary of the support set near 0.

If 0 is again a left endpoint and the problem is to estimate $f(x_0)$ for some $x_0 > 0$, the estimator from Section 1 is

$$
(3.1)\qquad \hat f_\alpha(x_0) = \frac{1}{n} \sum_{i=1}^n \rho^{2d/(2k+d)} \hat G(\rho^{2/(2k+d)}(X_i - x_0)).
$$

Suppose then that x_o varies with n so that $\rho^{2/(2k+d)} x_o \to \xi$ as $n \to \infty$. An analysis like that in Section 1 would now determine an asymptotically minimax estimator having the form (3.1), but with $\hat G$ satisfying (1.8) with integration extending over the interval $[-\xi, \infty)$. For a finite sample size illustration of this effect, suppose $d = 1$, $k = 2$, $x_o = ½$, $m = ½$, $\alpha = ¼$ and $n = 32$. A simple calculation gives $\xi = 1$ and one might therefore employ the $\hat{G}$ listed as Entry 7 in Table 1. For the same values of d, k, m, α and n, $x_o = 1$ would not be judged as near the boundary since, with $\xi = 2$, Entry 4 of Table 1 now applies.

B. EXAMPLES OF $\hat{G}$. (1.8) can be solved by hand in some simple cases, and generally one can resort to a computer and the algorithm of [8] in order to obtain a solution. In Table 1 the column "Interval" should be understood in the sense of Section 3A. By "Positive part of $\hat{G}$" is meant the function $\sum c_j u^j - \theta \| u \|^k$ that appears in (1.8a) $-\hat{G}$ itself is easily obtained when this is given. In this connection note that the kernels given as entries 1, 2, 3, 4, 7, 10 and 11 do not have a negative part (e.g., for Entry 7, $.586 - .056\, u + .308\, u^2$ is never negative). Entry 3 can be obtained as follows: For $0 < \xi < 3^{1/3}$ solve $\int_{-\xi}^{\infty} (c_o - \theta | u |)_+ \, du = \theta$ and obtain c_o/θ as the unique real root of $z^3 + 3z\xi^2 - 2\xi^3 - 6 = 0$. Then $\int_{-\xi}^{c_o/\theta} \left(\frac{c_o}{\theta} - | u |\right) du = 1/\theta$ determines θ.

C. ASYMPTOTIC EFFICIENCIES. To compare the asymptotic behavior of the optimal kernels with some standard ones, we will only deal with the important case $k = 2$.

According to Theorem 1, the kernels $\hat{\Gamma}_n$ have asymptotic mean square error

$$Q_{\alpha,n}(\hat{\Gamma}_n) \sim m^2 \left(\frac{\alpha}{nm^2}\right)^{4/(d+4)} c_o^*, \tag{3.2}$$

where (see (1.8) and remarks following) c_o^* is given by the constant coefficient in Table 1 and, from Entry 11 (which is the pertinent one), we obtain

$$c_o^* = v_o^{-4/(d+4)} \left(\frac{d+2}{2}\right)^{4/(d+4)} \left(\frac{d}{4+d}\right)^{d/(d+4)} \tag{3.3}$$

The Epanechnikov kernel Γ_{on} is given at (0.5) when $d = 1$ and, for general $d, \Gamma_{on}^{(d)} = \prod_{j=1}^{d} \Gamma_{on}(x_j)$. It is straightforward to calculate that

TABLE 1

Entry	Dimension (d)	Degree (k)	Interval	Positive Part of $\hat{G}$
1	1	1	$(-\infty, \infty)$ or $[-\xi, \infty)$, $\xi \geq 3^{1/3}$	$3^{-1/3}(1 - 3^{-1/3} \mid u \mid) = .693 - .481 \mid u \mid$
2	1	1	$[0, \infty)$	$2.6^{-1/3}(1 - 6^{-1/3} \mid u \mid) = 1.11 - 6.06 \mid u \mid$
3	1	1	$[-\xi, \infty)$, $\xi < 31/3$	(see text in Section 3B)
4	1	2	$(-\infty, \infty)$	$\frac{3}{4}\left(\frac{4}{15}\right)^{1/5}\left(1 - \left(\frac{4}{15}\right)^{2/5} u^2\right) = .576 - .339u^2$
5	1	2	$[0, \infty)$	$2.81 - 3.01u - .75u^2$
6	1	2	$[-.5, \infty)$	$.902 - .766u - .159u^2$
7	1	2	$[-1, \infty)$	$.586 - .056u - .308u^2$
8	1	3	$(-\infty, \infty)$	$.901 - .945u^2 - .359\mid u\mid^3$
9	1	4	$(-\infty, \infty)$	$.96 - 1.2u^2 - .33u^4$
10	d	1	R^d	$\frac{(v_1 - v_2)^{d/(d+2)}}{(v_0 - v_1)} (1 - (v_1 - v_2)^{1/(d+2)} \|u\|)^*$
11	d	2	R^d	$\frac{(v_2 - v_4)^{d/(d+4)}}{(v_0 - v_2)} (1 - (v_2 - v_4)^{2/(d+4)} \|u\|^2)^*$

* $v_j = \int_{\|u\| \leq 1} \|u\|^j \, du = \frac{d}{j+d} v_0$, $v_0 = \frac{\pi^{d/2}}{(d/2)!}$ if d is even, $v_0 = \pi^{(d-1)/2} 2^d \left(\frac{d-1}{2}\right)!/d!$ if d is odd.

$$(3.4)\qquad Q_{\alpha,n}(\Gamma^d_{on}) \sim m^2\left(\frac{\alpha}{nm^2}\right)^{4/(d+4)}(d+4)d^{d/(d+4)}4^{-4/(d+4)}\left(\frac{3}{5^{3/2}}\right)^{4d/(d+4)}.$$

The ratio of (3.2) to (3.4) produces the asymptotic relative efficiencies given as the entries in the column of Table 2 headed by Epanechnikov.

The other kernel we compare is Rosenblatt's kernel which is obtained by taking $W(u) = 1$ on $[-\frac{1}{2}, \frac{1}{2}]$, $W(u) = 0$ otherwise, $b_n = \left(\frac{9}{8n}\frac{\alpha}{m^2}\right)^{1/5}$ and forming the 1-dimensional kernel $\Gamma(u) = b_n^{-1}W(ub_n^{-1})$ (see (0.3)). For d-dimensions we use $\Gamma_{R,n} = \Gamma(u_1)\cdots\Gamma(u_d)$. Another straightforward calculation shows that

$$(3.5)\qquad Q_{\alpha,n}(\Gamma_{R,n}) \sim m^2\left(\frac{\alpha}{nm^2}\right)^{4/(d+4)} d^{d/(d+4)}\frac{d+4}{4}(36)^{-d/(d+4)}$$

The ratio of (3.2) to (3.5) produces the last column of Table 2.

D. Other classes of f's. For the special setting discussed in the introduction the assumption $|f''| \le 1$ leads to a problem different from those of the present paper, as well as from those considered in [2] and [7]. For simplicity in the present discussion suppose this is the relevant assumption and that 0 is a left endpoint of the support of f with $f(0) \le 1$. For Γ satisfying $\int \Gamma = 1$ and $\int x\Gamma = 0$, the bias term $\int \Gamma f - f(0)$ can be rewritten as $\int \Gamma(x)(f(x) - f(0) - xf'(0))\,dx$. After an integration by parts the minimax problem can be related to that of

$$(3.6)\qquad \min_{H|H(0)=0,H'(0)=1}\left\{\frac{1}{n}\int_0^\infty H''^2 + \left(\int_0^\infty |H|\right)^2\right\}$$

with $H'' = \Gamma$. Variational problems of this type are characterized by considerble mathematical difficulty and rather complicated solutions, as is evidenced by the work of Berkovitz and Pollard [1] and Hestenes and Redheffer [4]. For the present case, Theorem 1 of [4] implies that the minimizing Γ has a continuous derivative and this smoothness means the kernel given as Entry 5 in Table 1 cannot solve the problem with $|f''| \le 1$. Thus, while this kind of restriction on f is appealing, the difficulty of obtaining workable solutions makes it reasonable to retain the assumptions about f as they are given in Section 1. Moreover, it appears that there is not much loss of efficiency when one uses the kernels found here as suboptimal solutions to problems like (3.6) (see [8] for some remarks about this).

E. Choosing m. To implement the procedures of this paper it is necessary to specify m. Some scant information obtained in regression contexts indicates that crude methods may be adequate since the estimate $\hat{f}(0)$ seems to be insensitive to misspecification of m. We hope to address these problems later.

TABLE 2

	ARE	
d	Epanechnikov	Rosenblatt
1	1	.94
2	.99	.89
4	.95	.83
8	.91	.75
∞	.67	.49

REFERENCES

[1] BERKOVITZ, L. D. and POLLARD, H. (1967). A non-classical variational problem arising from an optimal filter problem, *Arch. Rational Mech. Anal.* **26** 281–304.
[2] EPANECHNIKOV, V. A. (1969). Nonparmetric estimates of a multivariate probability density. *Theor. Probability Appl.* **14** 153–158.
[3] FARRELL, R. H. (1972). On the best obtainable asymptotic rates of convergence in estimation of a density function at a point, *Ann. Math. Statist.* **43** 170–180.
[4] HESTENES, M. and REDHEFFER, R. (1974). On the minimization of certain quadratic functionals. I. *Arch. Rational Mech. Anal.* **56** 1–14.
[5] PARZEN, E. (1962). On the estimation of a probability density and mode, *Ann. Math. Statist.* **33** 1065–1076.
[6] ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–835.
[7] ROSENBLATT, M. (1971). Curve estimates. *Ann. Math. Statist.* **42** 1815–1842.
[8] SACKS, J. and YLVISAKER, D. (1978). Linear estimation for approximately linear models. *Ann. Statist.* **6** 1122–1137.
[9] STONE, C. J. (1978). Optimal rates of convergence for nonparametric estimators. To appear in *Ann. Statist.*
[10] TAPIA, RICHARD A. and THOMPSON, JAMES R. *Nonparametric Probability Density Estimation.* Johns Hopkins Univ. Press. Baltimore.
[11] WOODROOFE, M. (1970). On choosing a delta-sequence, *Ann. Math. Statist.* **41** 1665–1671.

DEPARTMENT OF MATHEMATICS
NORTHWESTERN UNIVERSITY
EVANSTON, ILLINOIS 60201

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA, LOS ANGELES
LOS ANGELES, CALIFORNIA 90024