The Annals of Statistics
1984, Vol. 12, No. 2, 766–774

SMOOTH OPTIMUM KERNEL ESTIMATORS OF DENSITIES, REGRESSION CURVES AND MODES[1]

BY HANS-GEORG MÜLLER

Universität Heidelberg

Several criteria concerning the choice of kernels for the nonparametric estimation of functions and their derivatives are discussed. A specific optimality criterion is described which applies to kernels of compact support and of different orders of smoothness. The solutions of the corresponding variational problems are explicitly given. Many kernels discussed previously are obtained as special cases.

1. Introduction: criteria for the choice of kernels. Kernel estimates for probability density functions introduced by Rosenblatt (1956) and Parzen (1962) as well as for regression curves in the fixed design case introduced by Priestley and Chao (1972) are in wide use by now. This is due mainly to their attractive properties and computational simplicity. The estimation of derivatives of these functions also has important applications (compare e.g. Johns and van Ryzin, 1972, R. S. Singh, 1977).

Let $\nu \geq 0$, $k > \nu + 1$ be given, assume that ν, k are both even or both odd, and define

$$M_{\nu,k} := \left\{ f \in L^2 : \int f(x) x^j dx = \begin{cases} 0 & 0 \leq j < k, \quad j \neq \nu \\ (-1)^\nu \nu! & j = \nu \end{cases} \right\}.$$

Let $X_1 \cdots X_n$ be i.i.d. observations with Lebesgue density f. Assume $f \in \mathscr{C}^k(I)$, where $I = [0, 1]$, $[0, \infty)$ or $(-\infty, \infty)$. In order to estimate $f^{(\nu)}$ we consider

$$f_{n,\nu}(x) := \frac{1}{nb_n^{\nu+1}} \sum_{i=1}^n \varphi\left(\frac{x - X_i}{b_n}\right) \tag{1.1}$$

where b_n is a sequence of bandwidths and the kernel φ is a bounded function, $\varphi \in M_{\nu,k}$. If $\nu = 0$, this is the Rosenblatt-Parzen estimate.

Let

$$Y_i = g(t_i) + \varepsilon_i, \qquad i = 1 \cdots n,$$

be noisy measurements taken from a function $g \in C^k([0, 1])$ at fixed points $0 \leq t_1 < t_2 < \cdots < t_n \leq 1$, where ε_i, $i = 1, 2 \cdots$ are i.i.d. random variables, $E\varepsilon_1 = 0$ (regression problem with a fixed design). An extension of the kernel estimate defined in Gasser and Müller (1979) and Cheng and Lin (1981) to the estimation

Received February 1983; revised November 1983.
[1] Research supported in part by Deutsche Forschungsgemeinschaft.
AMS 1980 *subject classification.* Primary 62G05; secondary 42C10.
Key words and phrases. Kernel density estimation, kernel regression estimation, estimation of derivatives, choice of kernels.

of $g^{(\nu)}$ is

(1.2) $$g_{n,\nu}(t) := \sum_{i=1}^{n} W_{i,\nu}(t) Y_i \qquad \text{where}$$

$$W_{i,\nu}(t) := \frac{1}{b_n^{\nu+1}} \int_{s_{i-1}}^{s_i} \varphi\left(\frac{t-x}{b_n}\right) dx$$

and $s_0 = 0$, $s_n = 1$, $t_i \le s_i \le t_{i+1}$, $i = 1 \cdots n$. We assume

$$\max_{1 \le i \le n} |s_i - s_{i-1}| = O(n^{-1}) \quad \text{and} \quad \varphi \in M_{\nu,k}.$$

For both estimates we require

(1.3) $$0 < b_n \to 0, \quad nb_n \to \infty \quad \text{as} \quad n \to \infty.$$

Kernels with compact support are advantageous from two points of view: there are considerable savings in computer time, and if the density to be estimated has compact support (which is always assumed for the regression function), estimation using a kernel with noncompact support will always be disturbed by boundary effects (Hominal and Deheuvels, 1979, Gasser and Müller, 1979). Another important characterization of a kernel is its smoothness, since it will be inherited by the estimated curve. It is necessary to use a μ times differentiable kernel, if the estimate $f_{n,\nu}$ resp. $g_{n,\nu}$ is to be differentiated μ times, $\mu \ge 0$. This is required e.g. for Silverman's testgraph method ($\nu = 0$, $\mu = 2$), devised to choose a good smoothing parameter b_n (Silverman, 1978). The number of vanishing moments of a kernel also has an effect on the performance of the estimate. This was pointed out by Bartlett (1963) and demonstrated in simulation studies by Schucany and Sommers (1977) and Gasser et al. (1982).

Therefore, we construct a class of kernels with compact support for the estimation of derivatives $\nu \ge 0$, exhibiting various degrees of smoothness $\mu \ge 0$ (implying that a generalized derivative of the corresponding order exists) and various numbers $(k - 2)$ of vanishing moments. The kernels to be constructed are optimal in the sense that they minimize the variance of the μth derivative of the estimate. In the following, kernels are assumed to have support $[-1, 1]$ if not stated otherwise.

It is not difficult to show for the mean square error (MSE), resp. integrated MSE (IMSE)

(1.4) $$\begin{aligned} \text{MSE/IMSE}(f_{n,\nu}) = {} & c_1[nb_n^{2\nu+1}]^{-1} \int \varphi^2(x)\, dx \\ & + c_2 b_n^{2(k-\nu)} \left(\int \varphi(x) x^k\, dx \right)^2 \\ & + o([nb_n^{2\nu+1}]^{-1} + b_n^{2(k-\nu)}) \end{aligned}$$

with constants c_1, c_2 depending neither on the bandwidth nor on the kernel, but on the function to be estimated (compare Deheuvels, 1977, and Müller and Gasser, 1979). The same holds true for $g_{n,\nu}$ (different constants), if we assume

$\varphi \in \text{Lip}([-1, 1])$, $t_i = i/n$, $i = 1 \cdots n$ and $E\varepsilon_1^2 < \infty$. If the support of the function to be estimated is limited, e.g. is [0, 1], there are boundary effects in the "boundary region" $R_n := [0, b_n) \cup (1 - b_n, 1]$ (compare Hominal and Deheuvels, 1979, Gasser and Müller, 1979, Rice, 1983). In this case, a necessary requirement that (1.4) is valid for IMSE integrated over the whole interval [0, 1] is that suitably modified kernels are used for estimation in $t \in R_n$. For the unmodified estimate, as defined by (1.1) or (1.2), (1.4) is valid for IMSE integrated over compact sets contained in the open interval (0, 1).

From (1.4) one can compute the optimal bandwidth by determining the minimum of the expression w.r.t. b_n. The dependence of MSE/IMSE on the kernel is then (up to some power) given by

$$T_1(\varphi) = \left(\int \varphi^2(x)\, dx \right)^{k-\nu} \left| \int \varphi(x) x^k\, dx \right|^{2\nu+1}.$$

In general, this expression can be made arbitrarily small (Deheuvels, 1977). In the case $\nu = 0$, $k = 2$, one can require the additional side condition $\varphi \geq 0$, which was adopted by Epanechnikov (1969) to derive an optimal kernel. For $k \leq 5$, Gasser et al. (1982) derived some further kernels minimizing T_1 under the side condition of a restricted number of sign changes of the kernel function.

A related problem was discussed by Eddy (1980) in the context of finding optimal kernels for the estimation of the mode by using the mode of the estimated density:

$$T_2(\varphi) = \left(\int \varphi^{(1)^2}(x)\, dx \right)^{k} \left| \int \varphi(x) x^k\, dx \right|^{3} \quad \text{under}$$

$$\varphi \in M_{0,k},\ \varphi(-1) = \varphi(1) = 0.$$

It may be shown that again the corresponding minimization problem has no solution. Instead, Eddy solves the problem of minimizing the variance w.r.t. the kernel:

$$\int \varphi^{(1)^2}(x)\, dx = \min \quad \text{under } \varphi \in M_{0,k},\ \varphi(-1) = \varphi(1) = 0.$$

He presents the solution to this problem for $k = 2, 4$ as Theorems 3.1, 3.2.

Similar ideas of minimizing the variance are discussed by Ramlau-Hansen (1983). Also, Gasser et al. (1982) consider the problem

$$\int \varphi^2(x)\, dx = \min \quad \text{under} \quad \varphi \in M_{\nu,k}.$$

The resulting kernels have also been discussed by Deheuvels (1977). These kernels minimize the variance of the estimate $f_{n,\nu}$ resp. $g_{n,\nu}$ but are discontinuous at $-1, 1$ (if $\nu = 0$, $k = 2$, one obtains the rectangular kernel). Using these kernels, estimated curves are not differentiable and indeed exhibit nasty jumps in practical applications especially if derivatives are estimated and the number of observations is small.

Assume f resp. $g \in \mathscr{C}^{k+\mu}(\mathrm{I})$ for some $\mu \geq 0$ and let φ satisfy

$$\varphi \in M_{\nu,k} \cap \mathscr{C}^{\mu}([-1, 1]), \quad \varphi^{(j)}(-1) = \varphi^{(j)}(1) = 0, \quad j = 0 \cdots \mu - 1. \tag{1.5}$$

Then φ is $(\mu - 1)$ times differentiable on $\mathbb{R}$ and $\varphi^{(\mu-1)}$ is absolutely continuous. Therefore

$$f_{n,\nu}^{(\mu)}(x) = \frac{1}{nb_n^{\nu+\mu}} \sum_{i=1}^{n} \varphi^{(\mu)} \left(\frac{x - X_i}{b_n}\right) \text{ resp.}$$

$$g_{n,\nu}^{(\mu)}(t) = \frac{1}{b_n^{\nu+\mu}} \sum_{i=1}^{n} Y_i \int_{s_{i-1}}^{s_i} \varphi^{(\mu)} \left(\frac{t - x}{b_n}\right) dx.$$

In the same manner as (1.4) we obtain by applying Lemma 2.1 below

$$\begin{aligned} E(f_{n,\nu}^{(\mu)}(x) - f^{(\nu+\mu)}(x))^2 & \\ &= c_1'[nb_n^{2(\nu+\mu)+1}]^{-1} \int \varphi^{(\mu)^2}(x)\, dx + c_2' b_n^{2(k-\nu)} \int \varphi^{(\mu)}(x) x^k\, dx \\ &\quad + o([nb_n^{2(\nu+\mu)+1}]^{-1} + b_n^{2(k-\nu)}) \end{aligned} \tag{1.6}$$

(the same result holds for $g_{n,\nu}^{(\mu)}$ or MSE integrated over appropriate intervals; see the discussion following (1.4)).

The criterion which we will adopt for the choice of kernels is the minimization of the variance term of (1.6) leading to the problem

$$\int \varphi^{(\mu)^2}(x)\, dx = \min \quad \text{under (1.5)}. \tag{1.7}$$

It is clear that the solution of this problem yields kernels with compact support for any order of derivative $\nu \geq 0$ and different numbers of vanishing moments ($k - 2$) and degrees of smoothness $\mu \geq 0$. Using the optimal bandwidth for $f_{n,\nu}$ according to (1.4) will make the variance in (1.6) dominate. The problem of minimizing the variance of the first derivative of a density estimate is of interest in empirical Bayes procedures (Johns and van Ryzin, 1972). Eddy's problem is a special case of (1.7) where $\nu = 0$, $\mu = 1$.

In the next section, an explicit solution of (1.7) will be derived. These kernels are polynomials of degree $(k + 2\mu - 2)$ and are therefore simple to use and implement.

2. A hierarchy of smooth optimal kernels.

The solution of (1.7) requires several steps.

LEMMA 2.1. *Let $-\infty < a < b < \infty$ and $\mu \geq 0$ be given.*

(a). *For any function φ*

$$\varphi \in M_{\nu,k} \cap \mathscr{C}^{\mu}([a, b]) \quad \text{and} \tag{2.1}$$

$$\varphi^{(j)}(a) = \varphi^{(j)}(b) = 0, \quad j = 0 \cdots \mu - 1 \tag{2.2}$$

imply that $\varphi^{(\mu)} \in M_{\nu+\mu,k+\mu}$.

(b). *For any function $\psi \in M_{\nu+\mu,k+\mu} \cap \mathscr{L}([a, b])$ there is exactly one function $\varphi \in \mathscr{L}^{\mu}([a, b])$ which satisfies $\varphi^{(\mu)} = \psi$ and* (2.2). *Then φ also satisfies* (2.1).

PROOF. (a). Integration by parts;

(b). Define iterated integrals of ψ by $\psi_0 := \psi$, $\psi_j(x) := \int_a^x \psi_{j-1}(t)\, dt$, $1 \le j \le \mu$, and define $\varphi := \psi_\mu$. Obviously $\varphi^{(\mu)} = \psi$, $\varphi^{(j)}(a) = 0$,

$$\varphi^{(j)}(b) = \psi_{\mu-j}(b) = \frac{1}{(\mu - j - 1)!} \int_a^b (b - x)^{\mu-j-1} \psi(x)\, dx$$
$$= 0, j = 0 \cdots \mu - 1.$$

Let $\tilde{\varphi}$ satisfy (2.2) and $\tilde{\varphi}^{(\mu)} = \psi$. Then for all $j \ge 0$

$$\int_a^b \tilde{\varphi}(x) x^j\, dx = \frac{j!}{(j + \mu)!} \int \psi(x) x^{j+\mu}\, dx$$

and therefore $\tilde{\varphi}$ is unique.

LEMMA 2.2. *The unique solution of the problem $\int \psi^2(x)\, dx = \min$ under $\psi \in M_{\nu+\mu,k+\mu} \cap \mathscr{L}([a, b])$ is a polynomial of degree $(k + \mu - 1)$. If $[a, b] = [-1, 1]$, the degree is $(k + \mu - 2)$ and the coefficients of the polynomial are given by*

$$(2.3) \qquad \gamma_i = \begin{cases} \dfrac{(-1)^{(i+\nu+\mu)/2}(k + \nu + 2\mu)!(k + \mu + i)!(k - \nu)(k + \mu - i)}{i!(i + \nu + \mu + 1)2^{2(k+\mu)+1}\left(\frac{k - \nu}{2}\right)!\left(\frac{k + \nu + 2\mu}{2}\right)!\left(\frac{k + \mu - i}{2}\right)!\left(\frac{k + \mu + i}{2}\right)!} & k + \mu + i \text{ even} \\ 0 & k + \mu + i \text{ odd.} \end{cases}$$

PROOF. The side conditions determine a polynomial $\tilde{\psi}$ of degree $(k + \mu - 1)$ uniquely. This polynomial is the unique solution, since for any other function $\psi \in M_{\nu+\mu,k+\mu} \cap \mathscr{L}([a, b])$:

$$\int \psi^2(x)\, dx = \int \tilde{\psi}^2(x)\, dx + 2 \int \tilde{\psi}(x)(\psi - \tilde{\psi})(x)\, dx + \int (\psi - \tilde{\psi})^2(x)\, dx$$
$$> \int \tilde{\psi}^2(x)\, dx.$$

If $[a, b] = [-1, 1]$, the polynomial is of degree $(k + \mu - 2)$, since any solution of the problem is symmetric if $(k + \mu)$ is even, and antisymmetric if $(k + \mu)$ is odd. This may be seen by symmetrization or antisymmetrization of a possible solution and application of Cauchy-Schwarz inequality. The coefficients of this polynomial under the side condition $\psi \in M_{\nu,k} \cap \mathscr{L}([-1, 1])$ are given in Gasser et al. (1982), Theorem 1, based on a Legendre expansion. Replacing ν by $(\nu + \mu)$ and k by $(k + \mu)$ yields (2.3).

LEMMA 2.3. *Let ν, k, μ, j, m be nonnegative integers satisfying $\nu + 2 \le k$, ν, k*

both even or both odd, and $(k+j-1)/2 \le m$.

a. *Let* $k+j$ *be even. For odd* λ *satisfying* $1 \le \lambda \le k+j-1$ *we have*

$$\sum_{i=0}^{m} (-1)^i \frac{1}{2i+\lambda} \frac{(2i+k+j)!}{((2i+k+j)/2)!(m-i)!(2i)!} = 0.$$

b. *Let* $k+j$ *be odd. For even* λ *satisfying* $2 \le \lambda \le k+j-1$ *we have*

$$\sum_{i=0}^{m} (-1)^i \frac{1}{2i+1+\lambda} \frac{(2i+1+k+j)!}{((2i+1+k+j)/2)!(m-i)!(2i+1)!} = 0.$$

PROOF. a. For $0 \le p \le (k+j-2)/2 < m$:

$$\sum_{i=0}^{m} (-1)^i \frac{i^p}{(2i+1)(2i+3)\cdots(2i+k+j-1)} \cdot \frac{(2i+k+j)!}{((2i+k+j)/2)!(m-i)!(2i)!}$$
$$= \frac{2^{(k+j)/2}}{m!} \sum_{i=0}^{m} (-1)^i i^p \binom{m}{i} = 0$$

and therefore

$$\sum_{i=0}^{m} (-1)^i \cdot \frac{(2i+1)(2i+3)\cdots(2i+\lambda-2)(2i+\lambda+2)(2i+\lambda+4)\cdots(2i+k+j-1)(2i+k+j)!}{(2i+1)(2i+3)\cdots(2i+k+j-1)((2i+k+j)/2)!(m-i)!(2i)!} = 0$$

which implies the result. Proof of b. is analogous.

THEOREM 2.4. *The unique solution of* (1.7) *is a polynomial of degree* $(k+2\mu-2)$, *restricted to* $[-1, 1]$. *The coefficients of this polynomial* $P_{\nu,k,\mu}$ *are given by*

$$(2.4) \qquad \gamma_i = \begin{cases} \dfrac{(-1)^{(i+\nu)/2}(k+\nu+2\mu)!(k+i)!(k-\nu)(k+2\mu-i)}{i!(i+\nu+1)2^{2(k+\mu)+1}\left(\frac{k-\nu}{2}\right)!\left(\frac{k+\nu+2\mu}{2}\right)!\left(\frac{k+2\mu-i}{2}\right)!\left(\frac{k+i}{2}\right)!} & k+i \text{ even} \\ 0 & k+i \text{ odd} \end{cases}$$

$0 \le i \le k+2\mu-2$.

PROOF. By Lemma 2.1, 2.2, the solution is a unique polynomial of degree $(k+2\mu-2)$. If $\mu = 0$, the coefficients are given in Lemma 2.2. If $\mu > 0$, by Lemma 2.1, 2.2, it is sufficient to show that the coefficients (2.4) of $P_{\nu,k,\mu}$ satisfy the requirements

$$\text{(a)} \qquad P_{\nu,k,\mu}^{(\mu)} = P_{\nu+\mu,k+\mu,0}$$

and

(b) $$P^{(j)}_{\nu,k,\mu}(-1) = P^{(j)}_{\nu,k,\mu}(1) = 0, \quad 0 \le j < \mu.$$

(a): To be seen by some simple algebra. (b): Let $\gamma_i^{(j)}$ be the coefficient of x^i of the polynomial $P^{(j)}_{\nu,k,\mu}$, $j \ge 0$. As $P^{(j)}_{\nu,k,\mu}$ is always symmetric or antisymmetric, it is enough to show that $\sum_{i=0}^{k+2\mu-2-j} \gamma_i^{(j)} = 0, \quad 0 \le j < \mu$.

If $(k+j)$ is even ($(j+\nu)$, $(k-\nu)$ are then even, too), it is equivalent to show

$$\sum_{i=0}^{(k+2\mu-2-j)/2} \frac{(-1)^i (k+2i+j)!(k+2\mu-2i-j)}{(2i)!(2i+j+\nu+1)\left(\frac{k+2\mu-2i-j}{2}\right)!\left(\frac{k+2i+j}{2}\right)!} = 0.$$

Define $m := (k+2\mu-2-j)/2$. As $0 \le j < \mu$, the assumptions of Lemma 2.3 a. are satisfied. The case that $(k+j)$ is odd is treated in the same way using Lemma 2.3 b.

Polynomial kernel functions discussed previously are special cases of (2.4) and

TABLE 1
Smooth optimum kernels (μ = 2, 3) for important values of parameters ν and k.

ν	**k**	μ	**Kernel on [−1, 1]**
0	2	2	$\frac{15}{16}(1 - 2x^2 + x^4)$
0	2	3	$\frac{35}{32}(1 - 3x^2 + 3x^4 - x^6)$
0	4	2	$\frac{105}{64}(1 - 5x^2 + 7x^4 - 3x^6)$
0	4	3	$\frac{315}{512}(3 - 20x^2 + 42x^4 - 36x^6 + 11x^8)$
0	6	2	$\frac{315}{2048}(15 - 140x^2 + 378x^4 - 396x^6 + 143x^8)$
0	6	3	$\frac{3465}{4096}(3 - 35x^2 + 126x^4 - 198x^6 + 143x^8 - 39x^{10})$
1	3	2	$\frac{105}{16}(-x + 2x^3 - x^5)$
1	3	3	$\frac{315}{32}(-x + 3x^3 - 3x^5 + x^7)$
1	5	2	$\frac{315}{64}(-5x + 21x^3 - 27x^5 + 11x^7)$
1	5	3	$\frac{3465}{512}(-5x + 28x^3 - 54x^5 + 44x^7 - 13x^9)$
2	4	2	$\frac{315}{32}(-1 + 9x^2 - 15x^4 + 7x^6)$
2	4	3	$\frac{3465}{256}(-1 + 12x^2 - 30x^4 + 28x^6 - 9x^8)$
2	6	2	$\frac{3465}{512}(-5 + 84x^2 - 270x^4 + 308x^6 - 117x^8)$
2	6	3	$\frac{45045}{1024}(-1 + 21x^2 - 90x^4 + 154x^6 - 117x^8 + 33x^{10})$

therefore share the same optimality property. These include:
Rectangular kernel ($\nu = 0$, $k = 2$, $\mu = 0$)
Epanechnikov kernel ($\nu = 0$, $k = 2$, $\mu = 1$)
Deheuvels: Legendre kernel of order j ($\nu = 0$, $k = 2j + 2$, $\mu = 0$)
Eddy: $\nu = 0$, any k, $\mu = 1$ (explicit solutions: $k = 2, 4$)
Ramlau-Hansen: $\nu = 0$, any k, any μ (explicit solutions: $k = 2, 4$)
Gasser et al.: any ν, any k, $\mu = 0, 1$ (explicit solutions).

Table 1 (based on (2.4)) contains some smooth optimum kernels ($\mu = 2, 3$).

Kernels for $\nu = 0$, $k = 2, 4$ of this table have also been explicitly computed by Ramlau-Hansen. Lemmata 2.1, 2.2 show that (1.7) has a unique solution also in case that the kernels are assumed to have asymmetric support. This allows the construction of smooth modified kernels needed for estimation within the boundary region R_n (see discussion following (1.4)). These kernels are polynomials of degree $(k + 2\mu - 1)$ and constitute a unique modification of kernels (2.4). In case that the support of the function is e.g. $[0, 1]$, using these kernels within R_n will make (1.4) and (1.6) hold for MSE integrated over the whole interval of support $[0, 1]$.

Acknowledgement. I wish to thank the referees for their comments which led to a considerable improvement of the presentation.

REFERENCES

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

CHENG, K. F. and LIN, P. E. (1981). Nonparametric estimation of a regression function. *Z. Wahrsch. verw. Gebiete* **57** 223–233.

DEHEUVELS, P. (1977). Estimation non paramétrique de la densité par histogrammes généralisés. *Revue de Statistique Appliquée* **25** No. 3, 5–42.

EDDY, W. F. (1980). Optimum kernel estimators of the mode. *Ann. Statist.* **8** 870–882.

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

GASSER, TH. and MÜLLER, H. G. (1979). Kernel estimation of regression functions. Proceedings Heidelberg 1979. *Lecture Notes in Math.* **757** 23–68.

GASSER, TH., MÜLLER, H. G., and MAMMITZSCH, V. (1982). Kernels for nonparametric curve estimation. Preprint No. 210, Sonderforschungsbereich 123. Universität Heidelberg.

HOMINAL, P. and DEHEUVELS, P. (1979). Estimation non paramétrique de la densité compte-tenu d'informations sur le support. *Revue de Statistique Appliquée* **27** No. 3, 47–59.

JOHNS, M. V., JR. and VAN RYZIN, J. (1972). Convergence rates for empirical Bayes two-action problems. II. Continuous Case. *Ann. Math. Statist.* **43** 934–947.

MÜLLER, H. G. and GASSER, TH. (1979). Optimal convergence properties of kernel estimates of derivatives of a density function. Proceedings Heidelberg 1979. *Lecture Notes in Math.* **757** 144–154.

PARZEN, E. (1962). On estimation of a probability density and mode. *Ann. Math. Statist.* **35** 1065–1076.

PRIESTLEY, M. B. and M. T. CHAO (1972). Nonparametric function fitting. *J. Royal Statist. Soc. B* **34** 385–392.

RAMLAU-HANSEN, H. (1983). The choice of a kernel function in the graduation of counting process intensities. *Scand. Actuarial J.* 165–182.

RICE, J. (1983). Boundary modification for kernel regression. Manuscript, University of California, San Diego.

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

SCHUCANY, W. R. and J. P. SOMMERS (1977). Improvement of kernel type density estimators. *J. Amer. Statist. Assoc.* **72** 420–423.

SILVERMAN, B. W. (1978). Choosing the window width when estimating a density. *Biometrika* **65** 1–11.

SINGH, R. S. (1977). Applications of estimators of a density and its derivatives to certain statistical problems. *J. Royal Statist. Soc. B* **39** 357–363.

INSTITUT FÜR ANGEWANDTE MATHEMATIK
UNIVERSITÄT HEIDELBERG
IM NEUENHEIMER FELD 294
6900 HEIDELBERG
WEST GERMANY