The Annals of Statistics
1980, Vol. 8, No. 5, 1160–1163

ON IMPROVING CONVERGENCE RATES FOR NONNEGATIVE KERNEL DENSITY ESTIMATORS[1]

BY GEORGE R. TERRELL AND DAVID W. SCOTT
Baylor College of Medicine and Rice University

To improve the rate of decrease of integrated mean square error for nonparametric kernel density estimators beyond $0(n^{-\frac{4}{5}})$, we must relax the constraint that the density estimate be a bonafide density function, that is, be nonnegative and integrate to one. All current methods for kernel (and orthogonal series) estimators relax the nonnegativity constraint. In this paper we show how to achieve similar improvement by relaxing the integral constraint only. This is important in applications involving hazard function and likelihood ratios where negative density estimates are awkward to handle.

1. Introduction. Estimating rates of convergence for various nonparametric density estimators has been an important research topic over the last twenty-five years; see Tapia and Thompson (1978). Since the shape of the density is of most interest, the integrated mean square error (IMSE) is an appropriate criterion. For kernel density estimates, Parzen (1962) proved that one can choose kernels so that the IMSE $= 0(n^{-2r/(2r+1)})$ for positive integers r; however, if $r > 2$ the kernel estimate is not nonnegative. Thus for a bonafide density estimate (i.e., nonnegative and integrating to one), the kernel method is limited by the rate $0(n^{-\frac{4}{5}})$; see Farrell (1972). Wahba (1975) has shown that rates of convergence like $0(n^{-2r/(2r+1)})$ are possible for orthogonal series estimators, but again, these estimators are not nonnegative.

Generally these estimates are negative only in the tails and, thus, for many applications the lack of nonnegativity is unimportant. However for techniques like likelihood ratio estimation and hazard rate estimation, the presence of negative values poses theoretical as well as practical problems. A negative hazard rate implies the spontaneous reviving of the dead. In this note we give an example of a particular class of estimators based on ordinary kernel estimators that achieves the goal of faster rates of convergence by relaxing the integral constraint rather than the nonnegativity constraint. The method involved bias reduction in the logarithm of the estimator.

2. Bias reduction by geometric extrapolation. Rosenblatt (1956) first considered estimating a density function $f(x)$ given only a random sample $x_1, \cdots, x_n$ by the formula

$$(1)\qquad \hat{f}_h(x) = \frac{1}{nh}\Sigma_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$

Received May 1979; revised June 1979.
[1]Research supported in part by the National Heart, Lung, and Blood Institute under Grant NIH 17269.
AMS 1970 *subject classifications.* Primary 62G05.
Key words and phrases. Nonparametric density estimation, kernel estimation, rates of convergence.

where K is a kernel function integrating to one and h is a positive smoothing parameter. The bias of this estimator may be made of order h^r by choosing a kernel with "moments" $m_i = \int y^i K(y)\, dy = 0$ for $i = 1, \cdots, r-1$ since

$$(2) \qquad E\hat{f}_h(x) = n\int \frac{1}{nh} K\left(\frac{x-t}{h}\right) f(t)\, dt$$
$$= \int K(w) f(x-hw)\, dw$$
$$= f(x) + \Sigma_{i=1}^{r} \frac{(-1)^i}{i!} f^{(i)}(x) m_i h^i + 0(h^{r+1})$$

using Taylor's theorem and assuming the existence of the necessary derivatives. The variance is dominated by the term

$$(3) \qquad \operatorname{Var} \hat{f}_h(x) = n\int \left[\frac{1}{nh} K\left(\frac{x-t}{h}\right)\right]^2 f(t)\, dt$$
$$= \frac{1}{nh} \int K(w)^2 f(x-hw)\, dw$$
$$= \frac{f(x)}{nh} \int K(w)^2\, dw + 0(n^{-1}).$$

For a particular choice of r it follows from (2) and (3) that optimally IMSE = $0(n^{-2r/(2r+1)})$. However, for $r > 2$, K cannot be a nonnegative function and therefore $\hat{f}_h$ is not nonnegative for all x. Thus to improve over IMSE = $0(n^{-\frac{4}{5}})$ using these estimators we must be willing to accept negative estimates.

It is of interest to obtain improved rates like $0(n^{-\frac{8}{9}})$ by relaxing the integral constraint rather than the nonnegativity constraint. We have done this by considering the logarithm of estimator (1). We shall limit our kernels to symmetric nonnegative kernels so that the odd "moments" in equation (2) are all zero. Let $I_h(x) = E\hat{f}_h(x)$. Then we may rewrite equation (2) in the form

$$(4) \qquad I_h(x) = f(x)\left[1 + \frac{a_2}{f(x)} h^2 + \frac{a_4}{f(x)} h^4 + \cdots\right]$$

where $a_i = (-1)^i f^{(i)}(x) m_i / i!$. Taking logarithms, we have

$$\log[I_h(x)] = \log f(x) + \frac{a_2}{f(x)} h^2 + \frac{a_4 f(x) - \frac{1}{2} a_2^2}{f(x)^2} h^4 + \cdots$$

by careful application of the series expansion for natural logarithms. We may reduce the bias for a fixed kernel by eliminating powers of h to obtain, for example,

$$\frac{4}{3} \log I_h(x) - \frac{1}{3} \log I_{2h}(x) = \log f(x) - \frac{4a_4 f(x) - 2a_2^2}{f(x)^2} h^4 + \cdots$$

or taking exponentials

$$(5) \qquad I_h(x)^{\frac{4}{3}} I_{2h}(x)^{-\frac{1}{3}} = f(x) + \frac{2a_2^2 - 4a_4 f(x)}{f(x)} h^4 + \cdots$$

using a series expansion for the exponential function.

The estimator we are proposing is actually the ratio of two ordinary nonnegative kernel estimators with $r = 2$

$$\hat{f}_h^*(x) = \hat{f}_h(x)^{\frac{4}{3}} \hat{f}_{2h}(x)^{-\frac{1}{3}} \tag{6}$$

where we take $\hat{f}_h^*(x) = 0$ whenever $\hat{f}_h(x) = 0$. We can easily check that this geometric estimator has the properties we claim; namely, bias $= 0(h^4)$ and variance $= 0(1/nh)$. Writing

$$\hat{f}_h(x) = I_h(x) + Z$$

and

$$\hat{f}_{2h}(x) = I_{2h}(x) + W,$$

we have shown in equations (2) and (3) that the random variables Z and W have expectation 0 while the variances and thus the covariances of Z and W are of order $1/nh$. Our estimator (6) may be written in the factored form

$$\begin{aligned} \hat{f}_h(x)^{\frac{4}{3}} \hat{f}_{2h}(x)^{-\frac{1}{3}} &= I_h(x)^{\frac{4}{3}} \left[1 + \frac{Z}{I_h(x)} \right]^{\frac{4}{3}} I_{2h}(x)^{-\frac{1}{3}} \left[1 + \frac{W}{I_{2h}(x)} \right]^{-\frac{1}{3}} \\ &= I_h(x)^{\frac{4}{3}} I_{2h}(x)^{-\frac{1}{3}} + \tfrac{4}{3} Z \left(\frac{I_h(x)}{I_{2h}(x)} \right)^{\frac{1}{3}} \\ &\quad - \tfrac{1}{3} W \left(\frac{I_h(x)}{I_{2h}(x)} \right)^{\frac{4}{3}} + 0\left[(Z + W)^2 \right] \end{aligned} \tag{7}$$

using the expansion $(1 + a)^x = 1 + ax + 0(a^2)$. From equation (4) it follows that $I_h(x)/I_{2h}(x) = 1 + 0(h^2)$ by simple division. That the bias is $0(h^4)$ follows from equations (5) and (7). From equation (7) we may write

$$\begin{aligned} \operatorname{Var} \hat{f}_h^*(x) &= E\left[\tfrac{4}{3} Z - \tfrac{1}{3} W \right]^2 + 0(n^{-1}) \\ &= \operatorname{Var}\left[\tfrac{4}{3} \hat{f}_h(x) - \tfrac{1}{3} \hat{f}_{2h}(x) \right] + 0(n^{-1}), \end{aligned} \tag{8}$$

which is simply $0(1/nh)$. Thus the optimal IMSE for the estimator (6) is $0(n^{-\frac{8}{9}})$.

The previous argument can be generalized as follows: Given a sequence of nonnegative and symmetric kernel density estimates for a fixed point with smoothing parameters whole multiples of a fixed parameter $h, \hat{f}_{ih}(x)$ for $i = 1, \cdots, s$, take a multiplicative combination of these with ith exponent

$$(-1)^{i-1} \frac{2s(s-1) \cdots (s-i+1)}{(s+1)(s+2) \cdots (s+i)}. \tag{9}$$

The resulting nonnegative estimate for $f(x)$ can be shown to have asymptotic integrated mean square error $0(n^{-4s/(4s+1)})$.

3. Example. Consider the geometric estimate of a standard Gaussian density using Rosenblatt's symmetric boxcar kernel $K(t) = 1$ for $|t| < \frac{1}{2}$ and zero elsewhere. From equation (4) the constants $a_2 = h^2 f''(x)/6$ and $a_4 = h^4 f^{(iv)}(x)/120$.

Thus the bias of $\hat{f}_h^*(x)$ is given by equation (5) and the variance is $25f(x)/(36nh)$ by equation (8). A straightforward computation of the integrated mean square error for the case of standard Gaussian samples gives

$$\text{IMSE} = \frac{25}{36nh} + \frac{17}{32\pi^{\frac{1}{2}}}\frac{h^8}{81} + 0(n^{-1})$$

which is asymptotically minimized by the choice

$$h = \left(\frac{225\pi^{\frac{1}{2}}}{17}\right)^{\frac{1}{9}} n^{-\frac{1}{9}} \tag{10}$$

with resulting IMSE $\doteq .550\, n^{-\frac{8}{9}}$. For the $0(n^{-\frac{4}{5}})$ estimator (1) using the Rosenblatt boxcar kernel it is well-known that IMSE $\doteq .339\ n^{-\frac{4}{5}}$, so the geometric estimate represents an improvement over the boxcar kernel estimate for $n \geqslant 232$. A similar comparison using a symmetric triangle kernel shows improvement for $n \geqslant 142$.

We performed some preliminary numerical experiments to observe the behavior of the geometric estimator using (10) for simulated standard Gaussian data. For twenty-five repetitions for the sample sizes 25, 100, and 500, we computed the actual area of the density estimate. The average areas were 1.052 (.009), 1.028 (.006), and 1.015 (.003), respectively, with the estimated standard deviation given in parentheses. It can be shown that the area will always converge to one from above for any sampling density.

4. Discussion. Our approach to improved convergence rates was suggested by extrapolation to the limit methods from numerical analysis [2]. In fact a linear combination of estimates $\hat{f}_{ih}$, $i = 1, \cdots, s$, using a fixed symmetric kernel with coefficients from (9) gives a Parzen kernel estimator with $r = 2s$. The method of this note is simply a multiplicative analogue. Our estimators are not unique. Geometric estimates may be constructed from kernel estimators other than $\hat{f}_h$ and $\hat{f}_{2h}$, for example $\hat{f}_h$ and $\hat{f}_{3h}$. For the Rosenblatt boxcar kernel, the pair $\hat{f}_h$ and $\hat{f}_{2.15h}$ is optimal but results in less than a 1% improvement in IMSE and is much less simple to use in practice.

REFERENCES

[1] Farrell, R. H. (1972). On best obtainable asymptotic rates of convergence in estimation of a density function at a point. *Ann. Math. Statist.* **43**, 170–180.

[2] Henrici, P. (1964). *Elements of Numerical Analysis*. Wiley, New York.

[3] Parzen, E. (1962). On the estimation of probability density and mode. *Ann. Math. Statist.* **33** 1065–1076.

[4] Rosenblatt, M. (1956). Remarks on some non-parametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837.

[5] Tapia, R. A. and Thompson, J. R. (1978). *Nonparametric Probability Density Estimation*. Johns Hopkins Univ. Press, Baltimore.

[6] Wahba, G. (1975). Optimal convergence properties of variable knot, kernel, and orthogonal series methods for density estimation. *Ann. Statist.* **3** 15–29.

Department of Community Medicine
Baylor College of Medicine
Houston, Texas 77036

Department of Mathematical Sciences
Rice University
Houston, Texas 77001