The Annals of Statistics
1992, Vol. 20, No. 1, 598–602

NON-EXISTENCE OF AN ADAPTIVE ESTIMATOR FOR THE VALUE OF AN UNKNOWN PROBABILITY DENSITY

BY MARK G. LOW

University of Pennsylvania and University of California, Berkeley

A strong adaptive criteria is defined for density estimation problems. In a particular case it is shown that there is no strongly adaptive sequence of estimators. In contrast Woodroofe has shown that a weakly adaptive result holds.

1. Adaptive estimation of a probability density function. Let $X_1, \dots, X_n$ be i.i.d. random variables with common density f with respect to Lebesgue measure. We focus attention on the pointwise estimation problem that is estimating $f(x_0)$ for some point x_0. Without loss of generality, we take $x_0 = 0$. The measure of loss will be squared error. In what follows, estimators indexed by n will always be assumed to be measurable functions of $X_1, \dots, X_n$. In this setup asymptotic minimax linear estimators are well known when f is assumed to belong to $SY(\alpha, M)$, where

$$SY(\alpha, M) = \left\{ f\colon \left[-\tfrac{1}{2}, \tfrac{1}{2}\right] \to R\colon f \geq 0, \int f = 1, f(0) \leq \alpha, |f(x) - f(0)| \leq M|x| \right\}; \tag{1.1}$$

see Sacks and Ylvisaker (1981) and Donoho and Liu (1991).

By linear, we mean an estimator of the form

$$\hat{f}_n = \frac{1}{n} \sum \Gamma_n(X_i), \quad \text{where } \Gamma_n \text{ is a measurable function.} \tag{1.2}$$

Sacks and Ylvisaker showed that the asymptotic minimax linear estimator is a kernel estimator,

$$\hat{f}_n = \frac{1}{nh_n} \sum K\left[\frac{X_i}{h_n}\right], \tag{1.3}$$

$$\text{where } K(x) = \begin{cases} 1 - |x|, & |x| \leq 1, \\ 0, & |x| > 1, \end{cases} \quad \text{and} \quad h_n = 3^{1/3} M^{-2/3} \alpha^{1/3} n^{-1/3}.$$

For this sequence of estimators,

$$\lim_{n \to \infty} n^{2/3} \sup_{f \in SY(\alpha, M)} E_f\left(f(0) - \hat{f}_n\right)^2 = \alpha^{2/3} M^{2/3} 3^{-1/3}, \tag{1.4}$$

Received July 1989; revised April 1991.
AMS 1980 *subject classifications*. Primary 62G07; secondary 62C99
Key words and phrases. Minimax risk, adaptive estimation, density estimation.

where E_f indicates that the expectation is to be evaluated under the assumption that f is the true density.

In this set up a natural adaptive version of (1.4) is given by a sequence of estimators satisfying the following definition.

DEFINITION. A sequence $\{\hat{f}_n\}$ of estimators is strongly adaptive [over the class $SY(\alpha, M)$] if

$$\sup_{\substack{\alpha_1 \le \alpha \le \alpha_2 \\ M_1 \le M \le M_2}} \limsup_{n \to \infty} 3^{1/3} n^{2/3} \alpha^{-2/3} \sup_{f \in SY(\alpha, M)} E_f\big(f(0) - \hat{f}_n\big)^2 \le 1. \tag{1.5}$$

Sacks and Ylvisaker (1981) prove there is such a sequence of estimators if $M_1 = M_2$ and $\alpha_1 = 0$, $\alpha_2 = \infty$. Moreover for $\alpha_2 - \alpha_1$ sufficiently small and $M_2/M_1 > 1$ sufficiently near 1, there must exist strongly adaptive estimators in the sense of (1.5). This follows because for given α, M there exists an $\hat{f}_n$ such that

$$\lim_{n \to \infty} 3^{1/3} n^{2/3} \alpha^{-2/3} M^{-2/3} \sup_{f \in SY(\alpha, M)} E_f\big(f(0) - \hat{f}_n\big)^2 < 1$$

[see Sacks and Strawderman (1982)].

In Section 2 we show that if $\alpha_1 < \alpha_2$ and $M_2/M_1 > 3.1$, then strongly adaptive estimators do not exist. In particular, strongly adaptive estimators do not exist whenever $M_2 = \infty$ or $M_1 = 0$.

The uniform adaptively condition given in (1.5) can be contrasted to a pointwise criteria for densities $f \in SY(\alpha, M)$ satisfying

$$\frac{f(y) - f(0)}{|y|} \to \begin{cases} M_R & \text{as } y \to 0^+, \\ M_L & \text{as } y \to 0^- \end{cases} \tag{1.6}$$

for some M_R and M_L, where $M_R + M_L \ne 0$. Note that M_R and M_L may depend on f.

We shall denote the class of densities $f \in SY(\alpha, M)$ satisfying (1.6) by $W(\alpha, M)$.

DEFINITION. A sequence $\{\hat{f}_n\}$ of estimators is weakly adaptive [over the class $W(\alpha, M)$] if

$$\lim_{n \to \infty} n^{2/3} (f(0))^{-2/3} \left(\frac{M_R + M_L}{2}\right)^{-2/3} E_f\big(f(0) - \hat{f}_n\big)^2 \le 3^{-1/3} \tag{1.7}$$

for each $f \in W(\alpha, M)$, where M_R and M_L are for each f, defined by (1.6).

The existence of a weakly adaptive sequence is essentially contained in Theorem 5.1 of Woodroofe (1970) applied to the kernel K given in (1.3). K however is not as required by that theorem twice continuously differentiable.

Instead look at a sequence K_j of twice continuously differentiable kernels such that $\int K_j^2(x)\,dx \to \int K^2(x)\,dx$ uniformly and $\int xK_j(x)\,dx \to \int xK(x)\,dx$ uniformly.

Woodroofe's theorem applies to each of these K_j and hence (1.7) holds with $3^{-1/3}$ replaced by $3^{-1/3} + \varepsilon_j$, where $\varepsilon_j \downarrow 0$. A simple diagonalization argument then shows the existence of a weakly adaptive sequence.

2. Main theorem.

THEOREM. *Suppose* $M_2/M_1 > 3.1$, *then*

$$\limsup_{n\to\infty} n^{2/3} \inf_{\hat{f}_n} \sup_{M_1 \le M \le M_2} \sup_{f \in W(1,M)} M^{-2/3} E_f\big(f(0) - \hat{f}_n\big)^2 > 3^{-1/3}. \tag{2.1}$$

PROOF. Consider the one parameter families

$$f_\theta^n(x) = 1 + \frac{12x^2 - 1}{n^{1/6}} + \theta\left(1 - \frac{|x| n^{1/3}}{d}\right)_+ - c_n(\theta), \qquad |\theta| \le M_2\, d n^{-1/3},$$

where d is a positive real number, $c_n(\theta) = \theta\, d n^{-1/3}$ and $(x)_+ = \max\{0, x\}$. Let

$$r_n(\theta) = \frac{1}{3^{1/3} n^{2/3}} \big(\min\{M: M_1 \le M \le M_2,\ f_\theta^n \in W(1, M)\}\big)^{2/3}.$$

If the theorem is false, then for any $\varepsilon > 0$ there is an estimator (sequence) δ_n such that

$$E\big(f_\theta(0) - \delta_n\big)^2 < (1 + \varepsilon) r_n(\theta)$$

for all n sufficiently large, say $n \ge n_0$ for some positive integer n_0. This together with the information inequality implies

$$(1 + \varepsilon) r_n(\theta) \ge \frac{\big(1 + \beta_n'(\theta)\big)^2}{n I_n(\theta)} + \beta_n^2(\theta), \qquad n \ge n_0,$$

where $I_n(\theta) = (2/3)\, d n^{-1/3}(1 + o(1))$ is the information function and $\beta_n(\theta) = E(\delta_n) - f_\theta(0)$ denotes the bias of the estimator δ_n. Hence

$$(1 + \varepsilon) r_n(\theta) \ge \frac{\big(1 + \beta_n'(\theta)\big)^2}{(2/3)\, d n^{2/3}(1 + \varepsilon)} + \beta_n^2(\theta), \tag{2.2}$$

$n \ge n_1$ for some positive integer n_1.

Now let

$$\phi = \frac{n^{1/3}\theta}{M_1^{1/3}}, \qquad D = M_1^{2/3} d,$$

and define the function c by

$$\frac{M_1^{1/3}}{n^{1/3}} c\left(\frac{n^{1/3}\theta}{M_1^{1/3}}\right) = \beta_n(\theta).$$

The theorem will be proved if we can show that for $M_2/M_1 > 3.1$, there is no solution to

$$(2.3)\quad \frac{1}{3^{1/3}}\left(\max\left(1, \frac{|\phi|}{D}\right)\right)^{2/3} \geq \frac{(1 + c'(\phi))^2}{(2/3)D} + c^2(\phi), \qquad |\phi| \leq \frac{M_2}{M_1}D.$$

Set $r_D(\phi) = (1/3^{1/3})(\max(1, \phi/D))^{2/3}$ Brown and Farrell (1990) have shown that (2.3) has a solution if and only if the differential equation

$$(2.4)\quad c'(\phi) = \left(\frac{2}{3}D\left(r_D(\phi) - c^2(\phi)\right)\right)^{1/2} - 1, \quad 0 \leq \phi \leq \frac{M_2}{M_1}D, \quad c(0) = 0.$$

has a solution.

Now suppose that we can find a function $c_1(t)$ such that (i) if $c(t)$ is a solution to (2.4) on any interval $[0, T]$, then $c(t) \leq c_1(t)$ on $[0, T]$ (ii) for some $T' < (M_2/M_1)D$, $c_1(T') < -(r_D(T'))^{1/2}$.

It would then follow that $T < T'$ otherwise $c^2(T') > r_D(T')$ which contradicts the inequality in (2.3) and the equality in (2.4). Hence we would have proved that there is no solution to (2.4) on the whole interval $[0, (M_2/M_1)D]$ and the theorem would be proved. We shall now construct such a function c_1 for the special case of $D = 1$. For that special case we write $r(\phi)$ instead of $r_1(\phi)$.

The construction is based on a modification of the Euler method for approximating solutions to first order differential equations. First fix a step size h. Then define c_1 recursively by $c_1(0) = 0$

$$(2.5)\quad c_1(jh + x) = c_1(jh) + \left(\left(\tfrac{2}{3}\left(r((j+1)h) - c_1^2(jh)\right)\right)^{1/2} - 1\right)x \quad \text{for } 0 \leq x \leq h,\ j = 0, 1, \ldots.$$

Note that since $r(\phi)$ is a nondecreasing function of ϕ,

$$(2.6)\quad r((j+1)h) \geq r(jh + x) \qquad \forall\, 0 \leq x \leq h.$$

Also note that any solution c to (2.4), (with $D = 1$) satisfies $c'(t) < 0$ (at least) if

$$\left(\frac{2}{3}r_1(t)\right) < 1, \quad \text{that is, if} \quad t < \frac{3^2}{2^{3/2}} \equiv 3.18.$$

Hence $c(t)$ is a decreasing function of t for $0 \leq t \leq 3.18$ and so if $h < 3.18$,

$$(2.7)\quad c(t) \leq c_1(0), \qquad 0 \leq t \leq h.$$

Now (2.5), (2.6) and (2.7) taken together show that $c'(t) \leq c_1'(t)$ on $[0, h)$ and hence

$$(2.8)\quad c(t) \leq c_1(t) \quad \text{on } [0, h].$$

Now suppose for some positive integer j that $c(jh) \leq c_1(jh)$ and assuming $(j+1)h < 3.18$, then once again noting that $c(t)$ is decreasing and that (2.6) holds, we have

$$c(t) \leq c_1(t) \quad \text{on } [jh, (j+1)h].$$

Hence this construction insures that c_1 satisfies condition (i).

For $h = 0.001$, $c(3.086) = -1.212$ and $c^2(3.086) = 1.471 > r(3.086) = 1.470$. Hence if $M_2/M_1 > 3.086$, there is no solution to (2.3) when $D = 1$ and the theorem is proved. □

REMARK. The construction of the function c_1 makes it clear that if $h_1 < h_2$, then $c_i^{h_1}(t) < c_1^{h_2}(t)$, where $c_1^h(t)$ is the c_1 function corresponding to the step size h. Hence the value of T' given in condition (ii), corresponding to h_1, will be smaller than that corresponding to h_2. In particular if we take $h = 0.01$ in our above example, $c_1^{0.01}(3.12) = -1.22$ and $(1.22)^2 = 1.4884 > r(3.12) = 1.4805$.

Acknowledgments. The author would like to thank L. D. Brown and an anonymous referee for comments on an earlier draft of this paper which helped to improve the final presentation. This paper was partly written when the author was visiting the University of Illinois, Champaign-Urbana. The advice of J. Sacks is gratefully acknowledged.

REFERENCES

BROWN, L. D. and FARRELL, R. H. (1990). A lower bound for the risk in estimating the value of a probability density. *J. Amer. Statist. Assoc.* **85** 1147–1153.

DONOHO, D. L. and LIU, R. C. (1991). Geometrizing rates of convergence, III. *Ann. Statist.* **19** 668–701.

SACKS, J. and STRAWDERMAN, W. E. (1982). Improvements on linear minimax estimators. In *Statistical Decision Theory and Related Topics III* (S. S. Gupta and J. O. Berger, eds) **2** 287–304. Academic, New York.

SACKS, J. and YLVISAKER, D. (1981). Asymptotically optimal kernels for density estimation at a point. *Ann. Statist.* **9** 334–346.

WOODROOFE, M. (1970). On choosing a delta sequence. *Ann. Math. Statist.* **41** 1665–1671.

DEPARTMENT OF STATISTICS
THE WHARTON SCHOOL
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA 19104-6302