The Annals of Statistics
1978, Vol. 6, No. 4, 932–934

LOWER BOUNDS FOR NONPARAMETRIC DENSITY ESTIMATION RATES

By David W. Boyd[1] and J. Michael Steele

University of British Columbia

If $f_n(x)$ is *any* estimator of the density $f(x)$, it is proved that the mean integrated square error is no better than $O(n^{-1})$.

1. Introduction. In Wegman's paper [5] on nonparametric density estimation, he states that it would be interesting to show that there is no density estimator which has mean integrated square rate better than $O(n^{-1})$. The object of this note is to prove such a result, making no arbitrary assumptions about the specific form of the estimator. This proof is given in Section 2. Our method applies to some other measures of error, as we point out in Section 3.

To be precise, a density estimator $\hat{f}_n(x) = \hat{f}_n(x; x_1, \cdots, x_n)$ is a sequence of Borel functions defined on R^{n+1}. If $X_1, X_2, \cdots$ are independent identically distributed random variables with density $f(x)$, then $\hat{f}_n(x; X_1, \cdots, X_n)$ provides an estimate for $f(x)$. The mean integrated square error is defined to be

$$\text{MISE}(n) = E_f \int_{-\infty}^{\infty} (f(x) - \hat{f}_n(x))^2\, dx\,, \tag{1}$$

where E_f denotes expectation according to the density f.

Tartar and Kronmal [3], and Wegman [4, 5] give a nice review of the extensive literature of such estimators.

2. The main result.

Theorem. *For any density estimator $\hat{f}_n(x)$, there is a square integrable density f, and a constant $c > 0$ such that*

$$E_f \int_{-\infty}^{\infty} (f(x) - f_n(x))^2\, dx \geqq c/n\,, \tag{2}$$

for infinitely many n. Thus there is no density estimator for which MISE (n) *is better than $O(n^{-1})$. In* (2), *f can be chosen to be a normal density with mean zero.*

Proof. We shall introduce a parametric family of densities $f_\theta(x)$, and use $\hat{f}_n(x)$ to construct an estimator $\hat{\theta}_n$ for the parameter θ. Specifically, if $n(x; 0, \sigma^2)$ is a normal density with mean 0, we define

$$\theta = \int_0^1 n(x; 0, \sigma^2)\, dx = \phi(\sigma)\,, \tag{3}$$

and let $J = \{\theta : \theta = \phi(\sigma), 1 \leqq \sigma \leqq 2\}$, a closed interval. For each $\theta \in J$ there is a unique $\sigma \in [1, 2]$ for which (3) holds; we denote the corresponding density

Received March 1977; revised October 1977.

[1] Supported in part by the National Research Council of Canada and the I. W. Killam Foundation.

AMS 1970 *subject classifications.* Primary 62G20; Secondary 62F20.

Key words and phrases. Nonparametric, density estimation, mean integrated square error, Cramér-Rao inequality.

$n(x; 0, \sigma^2)$ by $f_\theta(x)$. Thus $\theta = \int_0^1 f_\theta(x)\,dx$. Let $\hat{g}_n(x)$ be obtained from $\hat{f}_n(x)$ by truncating in the following way: $\hat{g}_n(x) = \min(\max(\hat{f}_n(x), 0), 1)$ so that $0 \leqq \hat{g}_n(x) \leqq 1$. We use this to construct the following estimator of θ:

$$\hat{\theta}_n = \int_0^1 \hat{g}_n(x)\,dx\,. \tag{4}$$

The basic observation is that by Schwarz' inequality we have

$$\begin{aligned}(\theta - \hat{\theta}_n)^2 &= \{\textstyle\int_0^1 (f_\theta(x) - \hat{g}_n(x))\,dx\}^2 \\ &\leqq \textstyle\int_0^1 (f_\theta(x) - \hat{g}_n(x))^2\,dx \leqq \int_0^1 (f_\theta(x) - \hat{f}_n(x))^2\,dx\,.\end{aligned} \tag{5}$$

Thus, writing $E_\theta = E_{f_\theta}$,

$$E_\theta(\theta - \hat{\theta}_n)^2 \leqq E_\theta \int_{-\infty}^{\infty} (f_\theta(x) - \hat{f}_n(x))^2\,dx\,. \tag{6}$$

The theorem will thus be proved by exhibiting a θ^* such that

$$E_{\theta^*}(\theta^* - \hat{\theta}_n)^2 \geqq c/n\,, \tag{7}$$

for infinitely many n.

By the Cramér–Rao inequality [6], page 188, one has

$$E_\theta(\theta - \hat{\theta}_n)^2 \geqq B(\theta)^2 + (1 + B'(\theta))^2/nI(\theta)\,, \tag{8}$$

where $B(\theta) = E_\theta(\hat{\theta}_n) - \theta$, and $I(\theta) = E_\theta((\partial/\partial\theta)\log f_\theta(x))^2$. The validity of (8) may be justified most easily by checking that the steps of the proof of (8) in [6], pages 182–188, are valid for the density f_θ used here. It is easily checked that $B(\theta)$, $B'(\theta)$ and $I(\theta)$ are continuous functions of θ. Since J is closed, we have $\sup_{\theta \in J} I(\theta) = M < \infty$. Let $J_1 = [a, b]$ be any closed interval in the interior of J, and let n_1 satisfy $n_1^{-\frac{1}{2}} \leqq (b - a)/8$. Let $S^2 = \sup_{\theta \in J_1}(1 + B'(\theta))^2$. If $S^2 \geqq \frac{1}{4}$, then there is an interval $J_2 \subset J_1$ on which $(1 + B'(\theta))^2 \geqq \frac{1}{8}$ and thus (8) implies $E_\theta(\theta - \hat{\theta}_{n_1}) \geqq 1/8n_1M$ for $\theta \in J_2$. On the other hand, if $S^2 \leqq \frac{1}{4}$, we can argue as follows: $B(b) - B(a) = B'(c)(b - a)$ for some c with $a < c < b$. But $1 - |B'(c)| \leqq 1 + B'(c) \leqq S$ so $|B'(c)| \geqq 1 - S \geqq \frac{1}{2}$ and thus

$$\max(|B(a)|, |B(b)|) \geqq |B(b) - B(a)|/2 \geqq (b - a)/4\,. \tag{9}$$

Let $\theta_1 = a$ or b satisfy $|B(\theta_1)| = \max(|B(a)|, |B(b)|)$. Then (8), (9) and the choice of n_1 imply

$$E_{\theta_1}(\theta_1 - \hat{\theta}_{n_1})^2 \geqq B(\theta_1)^2 \geqq 4/n_1\,. \tag{10}$$

By the continuity of $B(\theta)$, there is a closed interval $J_2 \subset J_1$ such that for $\theta \in J_2$ one has $E_\theta(\theta - \hat{\theta}_{n_1})^2 \geqq 1/n_1$. Repeating the argument, we obtain a sequence of nested closed intervals $J_1 \supset J_2 \supset \cdots$ and a sequence of integers $n_1 < n_2 < \cdots$ such that $E_\theta(\theta - \hat{\theta}_{n_k})^2 \geqq c/n_k$ for $\theta \in J_k$. Since the intersection $\bigcap_{k=1}^{\infty} J_k$ is non-empty, there is thus a θ^* for which (7) holds for $n = n_1, n_2, \cdots$. This completes the proof.

3. Further considerations. The application of Schwarz' inequality in (5) can be replaced by an application of Hölder's inequality to yield a priori lower bounds on the mean integrated pth power error for $p \geqq 1$. We omit these for the sake of brevity.

For certain specific classes of estimators one may obtain more precise lower bounds. For example, Rosenblatt [2] shows that estimators of kernel type have MISE (n) no better than $O(n^{-\frac{4}{5}})$. Fryer [1] has made an empirical study of such estimators for small n, when f is a normal density, and these indicate that the rate predicted by Rosenblatt is attainable.

We wish to emphasize the fact that our result does not depend on any assumptions about the density f or the estimator $\hat{f}_n$. Moreover, the proof shows that, even if f were known to be normal, with known mean, no improvement on the rate $O(n^{-1})$ would be possible.

REFERENCES

[1] FRYER, M. J. (1976). Some errors associated with the non-parametric estimation of density functions. *J. Inst. Math. Appl.* **18** 371–380.

[2] ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837.

[3] TARTAR, M. E. and KRONMAL, R. A. (1976). An introduction to the implementation and theory of nonparametric density estimation. *Amer. Statist.* **30** 105–112.

[4] WEGMAN, E. J. (1972). Nonparametric probability density estimation: I. A summary of available methods. *Technometrics* **14** 513–546.

[5] WEGMAN, E. J. (1972). Nonparametric probability density estimation: II. A comparison of estimation methods. *J. Statist. Comp. and Simulation* **1** 225–245.

[6] ZACKS, S. (1971). *The Theory of Statistical Inference*. Wiley, Toronto.

DEPARTMENT OF MATHEMATICS
THE UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER, BRITISH COLUMBIA
CANADA V6T 1W5