The Annals of Statistics
1979, Vol. 7, No. 2, 316–327

REMARKS ON SOME RECURSIVE ESTIMATORS OF A PROBABILITY DENSITY

BY EDWARD J. WEGMAN[1] AND H. I. DAVIES

University of North Carolina at Chapel Hill and Manchester University, England; University of New England, Armidale, Australia

The density estimator, $f_n^*(x) = n^{-1}\Sigma_{j=1}^n h_j^{-1}K((x - X_j)/h_j)$, as well as the closely related one $f_n^\dagger(x) = n^{-1}h_n^{-\frac{1}{2}}\Sigma_{j=1}^n h_j^{-\frac{1}{2}}K((x - X_j)/h_j)$ are considered. Expressions for asymptotic bias and variance are developed. Using the almost sure invariance principle, laws of the iterated logarithm are developed. Finally, illustration of these results with sequential estimation procedures are made.

1. Introduction. Let $X_1, X_2, \cdots$ be a sequence of i.i.d. observations drawn according to a probability density, f. Rosenblatt (1965) introduced the kernel estimator of the density, $f(x)$,

$$\hat{f}_n(x) = \frac{1}{nh_n}\Sigma_{j=1}^n K\left(\frac{x - X_j}{h_n}\right),$$

and Parzen (1962) developed many of the important properties of these estimators. A closely related estimator

$$f_n^*(x) = \frac{1}{n}\Sigma_{j=1}^n \frac{1}{h_j} K\left(\frac{x - X_j}{h_j}\right)$$

was introduced by Wolverton and Wagner (1969) and apparently independently by Yamato (1971). This second estimator has the very useful property that it can be calculated recursively, i.e.,

$$f_n^*(x) = \frac{n-1}{n} f_{n-1}^*(x) + \frac{1}{nh_n} K_n\left(\frac{x - X_n}{h_n}\right).$$

This property is particularly useful for large sample sizes, since addition of a few extra observations means that $\hat{f}_n(x)$ must be entirely recomputed—a tedious chore even with a computer.

In this paper we shall explore some properties of f_n^* as well as a related estimator $f_n^\dagger$, defined by

$$f_n^\dagger(x) = (nh_n)^{-\frac{1}{2}}\Sigma_{j=1}^n h_j^{-\frac{1}{2}} K\left(\frac{x - X_j}{h_j}\right).$$

Received August 1975; revised March 1978.

[1]The work of this author was supported by the Air Force Office of Scientific Research under Grant No. AFOSR75-2840.

AMS 1970 *subject classifications.* 62G05, 62L12, 60F20, 60G50.

Key words and phrases. Recursive estimators, asymptotic bias, asymptotic variance, weak consistency, almost sure invariance principle, law of the iterated logarithm, strong consistency, asymptotic distribution, sequential procedure.

This latter estimator can also be recursively formulated;

$$f_n^\dagger(x) = \frac{n-1}{n}\left(\frac{h_{n-1}}{h_n}\right)^{\frac{1}{2}} f_{n-1}^\dagger(x) + \frac{1}{nh_n}K\left(\frac{x-X_n}{h_n}\right).$$

Yamato (1971) considers f_n^* in some detail developing results similar to those of Parzen (1962). Carroll (1976) considers both $\hat{f}_n$ and f_n^* and, using weak convergence methods, establishes asymptotic distribution properties for both. Davies (1973) and Deheuvals (1974) independently develop conditions for almost sure and uniformly almost sure convergence of f_n^*.

Davies and Wegman (1975) and Carroll (1976) discuss sequential procedures for density estimation using $\hat{f}_n$ and f_n^*. Finally we note Eddy (1976) also introduces weak convergence methods to the consideration of $\hat{f}_n$ for the purpose of establishing results about the mode of $\hat{f}_n$.

In this paper we introduce the almost sure invariance principle into the consideration of f_n^* and $f_n^\dagger$. Through it we are able to establish a law of the iterated logarithm for density estimators as well as the asymptotic distribution results. We also illustrate the use of f_n^* as a sequential density estimator.

Throughout this paper we shall deal with univariate estimators. The extension to the multivariate case is straightforward. We shall assume that K is a Borel function satisfying

(1)
$$\sup_{-\infty<y<\infty}|K(y)| < \infty$$
$$\int_{-\infty}^{\infty}|K(y)|\, dy < \infty$$
$$\lim_{y\to\pm\infty}|yK(y)| = 0,$$

and also

(2)
$$h_n \to 0$$
$$nh_n \to \infty.$$

Other assumptions about K and $\{h_n\}$ will be made as needed. A word of comment about the assumptions is in order. K and $\{h_n\}$ are chosen by the statistician and, hence, restrictions on these quantities should not be viewed as an undue handicap unless, of course, the technical requirements force a slower rate of convergence. Eddy (1976) and Carroll (1976) develop alternate sets of requirements on K and $\{h_n\}$ which are restrictive in different ways. In our subsequent results we could exchange our requirements for theirs to develop other sets of sufficient conditions. However, neither of them deal with a bound on the bias, a result which requires the relatively most unpleasant conditions on K. Indeed, weak convergence theory seems unsuited for dealing with the question of bias. For this reason we have chosen to develop asymptotic variance and bias by elementary methods rather than appealing to the theory of weak convergence.

Davis (1975) presents a discussion of bias for the usual kernel estimate in which the assumption in Theorem 1, that the kernel, K, has a Fourier transform, is unneeded. However, her discussion is limited to the case $r = 2$. Her sufficient conditions may be exchanged for ours in the more specialized situation where $r = 2$.

2. Asymptotic bias, variance and consistency. In this section we shall base our results largely on the methods of Parzen (1962); therefore, we shall only sketch proofs. It will be convenient, throughout this paper, to deal with the sum

$$n(h_n)^{\frac{1}{2}} f_n^\dagger(x) = \Sigma_{j=1}^n h_j^{-\frac{1}{2}} K\left(\frac{x - X_j}{h_j}\right).$$

We recall a useful lemma from Parzen (1962).

LEMMA 1. *Suppose K is a Borel function satisfying* (1) *and* $\{h_n\}$ *satisfies* (2). *Let g satisfy*

$$\int_{-\infty}^{\infty} |g(u)|\, du < \infty.$$

Then

$$\frac{1}{h_n} \int_{-\infty}^{\infty} K\left(\frac{u}{h_n}\right) g(x - u)\, du \to g(x) \int_{-\infty}^{\infty} K(u)\, du \quad \text{as} \quad n \to \infty$$

at each continuity point, x, of g.

THEOREM 1. (a) *Let K and* $\{h_n\}$ *satisfy* (1) *and* (2). *If f is continuous at x,*

$$nh_n \operatorname{Var} f_n^\dagger(x) \to f(x) \int_{-\infty}^{\infty} K^2(u)\, du.$$

(b) *Let us further suppose K has Fourier transform K^* so that $K^*(u) = \int_{-\infty}^{\infty} e^{-iuy} K(y)\, dy$. Suppose further that for some r, $\lim_{u \to 0}\{[1 - K^*(u)]/|u|^r\} = k_r$ is finite and that $f^{(r)}(x)$ exists. Suppose finally that $nh_n^r \to \infty$ and $1/nh_n^r \Sigma_{j=1}^n h_j^r$ converges to γ_r. Then $(Ef_n^*(x) - f(x))/h_n^r \to \gamma_r \cdot k_r \cdot f^{(r)}(x)$.*

(c) *Let K satisfy the condition of part* (b) *and assume* $\{h_n\}$ *satisfies* (2) *and that $nh_n^{r+\frac{1}{2}} \to \infty$ and $1/nh_n^{r+\frac{1}{2}} \Sigma_{j=1}^n h_j^{r+\frac{1}{2}}$ converges to $\gamma_{r+\frac{1}{2}}$. Then*

$$\frac{Ef_n^\dagger(x) - f(x)(nh_n)^{-\frac{1}{2}} \Sigma_{j=1}^n h_j^{\frac{1}{2}}}{h_n^r} \to \gamma_{r+\frac{1}{2}} \cdot k_r \cdot f^{(r)}(x).$$

COMMENT. A popular choice for the sequence h_n is $bn^{-\gamma}$. In this case γ_r becomes $1/(1 - \gamma r)$ so that $Ef_n^*(x) - f(x) = 0(n^{-\gamma r})$. Similarly for $f_n^\dagger$. A slight variation in the argument yields the slightly more general result that $Ef_n^*(x) - f(x) = 0(n^{-1} \Sigma_{j=1}^n h_j^r)$ in the case of (b). For the case $h_n = bn^{-\gamma}$, these orders of convergence are the same. However, this need not always be the case. For example, if $r = 1$ and $h_n = b(\log n/n)$, then $n^{-1} \Sigma_{j=1}^n h_j = 0(\log \log n \cdot \log n/n) = 0(\log \log n \cdot h_n)$. Similar results hold for (c). We note that if $h_n = bn^{-\gamma}$, then $f_n^\dagger(x)$ is an asymptotically unbiased estimator of $f(x)/(1 - \gamma/2)$.

PROOF. (a) By Lemma 1,

$$\int_{-\infty}^{\infty} \frac{1}{h_j} K^2\left(\frac{x - u}{h_j}\right) f(u)\, du \to f(x) \int_{-\infty}^{\infty} K^2(u)\, du.$$

But, using Cesaro sums,

$$\lim_{n\to\infty} nh_n \operatorname{Var} f_n^\dagger(x) = \lim_{n\to\infty} \frac{1}{n}\Sigma_{j=1}^n \frac{1}{h_j}\int_{-\infty}^{\infty} K^2\left(\frac{x-u}{h_j}\right) f(u)\,du$$
$$= f(x)\int_{-\infty}^{\infty} K^2(u)\,du.$$

(b) Let $k_r = \lim_{u\to 0}[1 - K^*(u)]/|u|^r$, and define

$$E_j = \int_{-\infty}^{\infty} h_j^{-1} K\left(\frac{x-y}{h_j}\right) f(y)\,dy$$
$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iux} K^*(h_j u)\phi(u)\,du,$$

where, of course, $\phi(u)$ is the characteristic function associated with f. Then

$$\frac{E_j - f(x)}{h_j^r} = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iux}\frac{K^*(h_j u) - 1}{|h_j u|^r}|u|^r\phi(u)\,du$$
$$\to k_r f^{(r)}(x) \quad \text{as } j\to\infty.$$

But

$$\frac{1}{n}\Sigma_{j=1}^n \frac{E_j - f(x)}{h_n^r} = \frac{1}{nh_n^r}\Sigma_{j=1}^n h_j^r \frac{E_j - f(x)}{h_j^r} \to \gamma_r \cdot k_r f^{(r)}(x).$$

Finally $1/n\Sigma_{j=1}^n E_j = Ef_n^*(x)$ and (b) follows.

(c) As in part (b), consider

$$\frac{1}{nh_n^{r+\frac{1}{2}}}\Sigma_{j=1}^n h_j^{\frac{1}{2}}(E_j - f(x)) = \frac{1}{nh_n^{r+\frac{1}{2}}}\Sigma_{j=1}^n h_j^{r+\frac{1}{2}}\cdot\frac{E_j - f(x)}{h_j^r}$$
$$\to \gamma_{r+\frac{1}{2}}\cdot k_r\cdot f^{(r)}(x)$$

But

$$\frac{1}{nh_n^{\frac{1}{2}}}\Sigma_{j=1}^n h_j^{\frac{1}{2}} E_j = Ef_n^\dagger(x)$$

and (c) follows.

3. An almost sure invariance principle. Strassen (1964, 1965) introduced the idea of an almost sure invariance principle and this notion has been developed by Jain, Jogdeo and Stout (1975). Briefly put, we will use the almost sure invariance principle by showing that the sum,

$$\Sigma_{j=1}^n h_j^{-\frac{1}{2}}\left(K\left(\frac{x-X_j}{h_j}\right) - EK\left(\frac{x-X_j}{h_j}\right)\right)$$

is, with probability one, close to Brownian motion in a sense made precise below. The asymptotic fluctuation behaviour of Brownian motion has been investigated and, by use of the almost sure invariance principle, we may transfer results about Brownian motion to our density estimates.

We first shall reproduce some relevant results from Jain, Jogdeo and Stout (1975). Theorem 2 represents a less general version of Theorems 3.2 and 5.1 of Jain, Jogdeo and Stout (1975). Let $Y_1, \cdots, Y_n, \cdots$ be a sequence of zero mean random variables with finite second moments. Let $S_n = \Sigma_{j=1}^n Y_j$ and $V_n = \Sigma_{j=1}^n E[Y_j^2]$, $S_0 = 0 = V_0$.

THEOREM 2. *For a fixed* $\alpha \geqslant 0$, *assume*

$$V_n \to \infty \tag{3}$$

and

$$\Sigma_{k=1}^{\infty} \frac{(\log_2 V_k)^\alpha}{V_k} E\left\{ Y_k^2 I\left[Y_k^2 > \frac{V_k}{\log V_k} (\log_2 V_k)^{2(\alpha+1)} \right] \right\} < \infty. \tag{4}$$

Let S *be a random function defined on* $[0, \infty)$ *obtained by setting* $S(t) = S_n$ *for* $t \in [V_n, V_{n+1})$. *Then, redefining* $\{S(t),\ t \geqslant 0\}$, *if necessary, on a new probability space, there exists a Brownian motion* ξ *such that*

$$|S(t) - \xi(t)| = 0\left(t^{\frac{1}{2}} (\log_2 t)^{(1-\alpha)/2}\right) \text{ a.s.} \tag{5}$$

Here $\log_2 t = \log\log t$.

In particular, if (4) *holds with* $\alpha = 2$ *and* $\phi > 0$ *is a nondecreasing function, then*

$$P\left[S_n > V_n^{\frac{1}{2}} \phi(V_n) \text{ i.o.} \right] = 0 \qquad \text{or} \quad 1$$

according as

$$\int_1^\infty \frac{\phi(t)}{t} \exp(-\phi^2(t)/2)\, dt < \infty \qquad \text{or} \quad = \infty.$$

Let us identify $Y_j = h_j^{-\frac{1}{2}}(K((x - X_j)/h_j) - EK((x - X_j)/h_j))$, so that $S_n = nh_n^{\frac{1}{2}}(f_n^\dagger(x) - Ef_n^\dagger(x))$ and

$$V_n = \Sigma_{j=1}^n EY_j^2 = E\Sigma_{j=1}^n \frac{1}{h_j}\left[K\left(\frac{x - X_j}{h_j}\right) - EK\left(\frac{x - X_j}{h_j}\right)\right]^2$$

$$= h_n \operatorname{Var} nf_n^\dagger(x).$$

But under the assumptions of Theorem 1

$$nh_h \operatorname{Var} f_n^\dagger(x) \to f(x) \int_{-\infty}^{\infty} K^2(u)\, du,$$

so that $V_n/n = h_n n \operatorname{Var} f_n^\dagger(x) \to f(x)\int_{-\infty}^{\infty} K^2(u)\, du$. Thus $V_n = 0(n)$.

THEOREM 3. (a) *Let* K *satisfy* (1) *and* $\{h_n\}$ *satisfy* (2). *Let* f *satisfy the conditions of Theorem* 1. *If, in addition,*

$$\frac{nh_n}{\log n (\log_2 n)^{(\alpha+1)}} \quad \textit{diverges to} \quad \infty, \tag{6}$$

then (5) *holds for* S_n *defined above.*

(b) *In particular, if*

$$\frac{nh_n}{\log n(\log_2 n)^6} \quad \textit{diverges to} \quad \infty,$$

then

$$P\Big[S_n > V_n^{\frac{1}{2}}\phi(V_n)\text{ i.o.}\Big] = 0 \qquad \textit{or} \quad 1$$

according as

$$\int_1^\infty \frac{\phi(t)}{t}\exp(-\phi^2(t)/2)\,dt < \infty \qquad \textit{or} \quad = \infty.$$

(c) *For* $\alpha \geqslant 0$

$$\left(\frac{nh_n}{\log_2 n}\right)^{\frac{1}{2}}(f_n^\dagger(x) - Ef_n^\dagger(x)) \to \left(2f(x)\int_{-\infty}^{\infty}K^2(u)\,du\right)^{\frac{1}{2}} \text{ a.s.} \qquad \textit{as} \quad n \to \infty.$$

(d) *For* $\alpha \geqslant 1$,

$$\lim_{n\to\infty} P\left[\frac{f_n^\dagger(x) - Ef_n^\dagger(x)}{(\operatorname{Var} f_n^\dagger(x))^{\frac{1}{2}}} \leqslant w\right] = (2\pi)^{-\frac{1}{2}}\int_{-\infty}^{w} e^{-\frac{1}{2}y^2}\,dy.$$

PROOF. (a) Since K is bounded and (6) holds, the event $[Y_k^2 \geqslant c^*(k/\log k(\log_2 k)^{2(\alpha+1)})]$ is an impossible event for k sufficiently large. It follows that

$$E\left\{Y_k^2 I\left[Y_k^2 > \frac{V_k}{\log V_k(\log_2 V_k)^{2(\alpha+1)}}\right]\right\} < \infty$$

and the conclusion of (a) holds. Part (b) is immediate. To see part (c), divide (5) by $(2\log_2 t \cdot t)^{\frac{1}{2}}$, so that

$$\left|\frac{S(t) - \xi(t)}{(2\cdot\log_2 t\cdot t)^{\frac{1}{2}}}\right| \leqslant 0\big((\log_2 t)^{-\alpha/2}\big).$$

Since $\xi(t)$ satisfies the law of the iterated logarithm, so does $S(t)$ for $\alpha \geqslant 0$. It follows that

$$\frac{S_n}{(2\cdot\log_2 n\cdot n)^{\frac{1}{2}}}\cdot\frac{(\log_2 n\cdot n)^{\frac{1}{2}}}{(\log_2 V_n\cdot V_n)} \to 1 \text{ a.s.} \qquad \text{as} \quad n\to\infty.$$

But $(\log_2 n \cdot n)^{\frac{1}{2}}/(\log_2 V_n \cdot V_n)^{\frac{1}{2}} \to (1/f(x)\int_{-\infty}^{\infty}K^2(u)\,du)^{\frac{1}{2}}$ so that, recalling $n(h_n)^{\frac{1}{2}}(f_n^\dagger(x) - Ef_n^\dagger(x)) = S_n$, we have

$$\left(\frac{nh_n}{\log_2 n}\right)^{\frac{1}{2}}(f_n^\dagger(x) - Ef_n^\dagger(x)) \to \left(2f(x)\int_{-\infty}^{\infty}K^2(u)\,du\right)^{\frac{1}{2}}.$$

For part (d), we observe that $\xi(t)/t^{\frac{1}{2}}$ is normal mean zero variance one ($n(0, 1)$).

But

$$\left|\frac{S(t)}{t^{\frac{1}{2}}} - \frac{\xi(t)}{t^{\frac{1}{2}}}\right| \leqslant 0\big((\log_2 t)^{((1-\alpha)/2)}\big) \text{ a.s.}$$

For $\alpha > 1$, letting $t = V_n$

$$\frac{S_n}{(V_n)^{\frac{1}{2}}} \quad \text{is asymptotically} \quad n(0, 1).$$

But $V_n = n^2 h_n \operatorname{Var} f_n^\dagger(x)$ and $S_n = n(h_n)^{\frac{1}{2}}(f_n^\dagger(x) - Ef_n^\dagger(x))$ so that (d) follows.

Part (c) of Theorem 3 is particularly notable since it gives a law of the iterated logarithm for the density estimator, $f_n^\dagger$. This, together with part (c) of Theorem 1, gives the exact rate of convergence for $f_n^\dagger$. This result may be stated in several ways, depending on choice of h_n. We make one single statement here for the popular choice of $h_n = bn^{-\gamma}$.

THEOREM 4. *Let K and h_n be chosen in accordance with the hypotheses of Theorems* 1 *and* 3 *and, in addition, let $h_n = bn^{-\gamma}$. If* $(1/(2r + 1)) \leqslant \gamma$,

$$\left(\frac{n^{1-\gamma}}{\log_2 n}\right)^{\frac{1}{2}}\left(f_n^\dagger(x) - \frac{f(x)}{n(h_n)^{\frac{1}{2}}}\Sigma_{j=1}^n (h_j)^{\frac{1}{2}}\right) \to \left(2f(x)\textstyle\int_{-\infty}^{\infty} K^2(u)\, du\right)^{\frac{1}{2}} \text{ a.s.} \qquad as \quad n \to \infty.$$

PROOF. By Theorem 1(c),

$$\left(\frac{n^{1-\gamma}}{\log_2 n}\right)^{\frac{1}{2}}\left[Ef_n^\dagger(x) - \frac{1}{n(h_n)^{\frac{1}{2}}}\Sigma_{j=1}^n h_j f(x)\right] = 0\left(\frac{n^{1-\gamma(2r+1)}}{\log_2 n}\right)^{\frac{1}{2}}$$

which clearly converges to 0 if $\gamma \geqslant (1/(1 + 2r))$.

While $f_n^\dagger$ is an asymptotically biased estimator, of course, $(1 - \gamma/2)f_n^\dagger$ is asymptotically unbiased. However, the point of introducing $f_n^\dagger$ is to use it as a vehicle for obtaining results for f_n^* which we do by the following lemmas. Unfortunately, the results obtained from the almost sure invariance principle cannot be directly applied to f_n^*, hence the circuitous route.

Let $b_n = h_n^{-\frac{1}{2}}$ and $c_n = (n \log_2 n)^{\frac{1}{2}}$. Let

$$S_n = \Sigma_{j=1}^n h_j^{-\frac{1}{2}}\left[K\left(\frac{x - X_j}{h_j}\right) - EK\left(\frac{x - X_j}{h_j}\right)\right]$$

$$= n(h_n)^{\frac{1}{2}}\left[f_n^\dagger(x) - Ef_n^\dagger(x)\right].$$

LEMMA 2. *Let $a_j = b_j - b_{j-1}, j \geqslant 2$ with $a_1 = b_1$. If*

$$\lim_{n\to\infty}\left(\frac{h_n}{n \log_2 n}\right)^{\frac{1}{2}}\Sigma_{j=1}^{n-1}\left(h_j^{-\frac{1}{2}} - h_{j-1}^{-\frac{1}{2}}\right)(j \log_2 j)^{\frac{1}{2}} = \nu < \infty,$$

then

$$\lim_{n\to\infty} \frac{1}{b_n' c_n} \Sigma_{j=1}^{n-1} a_j S_j = \nu\left(2f(x)\int_{-\infty}^{\infty} K^2(u)\, du\right)^{\frac{1}{2}} \text{ a.s.}$$

Furthermore, if $h_n = bn^{-\gamma}$.

$$\lim_{n\to\infty} \frac{1}{b_n c_n} - \Sigma_{j=1}^{n-1} a_j S_j = \frac{\gamma}{\gamma+1}\left(2f(x)\int_{-\infty}^{\infty} K^2(u)\, du\right)^{\frac{1}{2}} \text{ a.s.}$$

PROOF. By Theorem 3 part (c), $S_n/c_n \to (2f(x)\int_{-\infty}^{\infty} K^2(u)\, du)^{\frac{1}{2}} = s$. Let $\varepsilon > 0$ and let $S_n' = (1/b_n c_n)\Sigma_{j=1}^{n-1} a_j S_j$. There is N_ε such that $n > N_\varepsilon$ implies

$$\frac{1}{b_n c_n}\Sigma_{j=1}^{N_\varepsilon} a_j S_j + \frac{1}{b_n c_n}\Sigma_{j=N_\varepsilon+1}^{n-1} a_j c_j (s - \varepsilon) \leqslant S_n'$$

$$\leqslant \frac{1}{b_n c_n}\Sigma_{j=1}^{N_\varepsilon} a_j S_j + \frac{1}{b_n c_n}\Sigma_{j=N_\varepsilon+1}^{n-1} a_j c_j (s - \varepsilon).$$

Taking lim inf and lim sup,

$$\nu(s - \varepsilon) \leqslant \liminf S_n' \leqslant \limsup S_n' \leqslant (s + \varepsilon)\nu.$$

Letting $\varepsilon \downarrow 0$ yields the desired result.

To demonstrate the second part we approximate

$$\frac{1}{b_n c_n}\Sigma_{j=1}^{n-1} a_j c_j = \frac{1}{n^{((1+\gamma)/2)}(\log_2 n)^{\frac{1}{2}}}\Sigma_{j=1}^{n-1}\left(j^{\gamma/2} - (j-1)^{\gamma/2}\right)(j \log_2 j)^{\frac{1}{2}}$$

by

$$\frac{\int_0^x \left(y^{\gamma/2} - (y-1)^{\gamma/2}\right)(y \log_2 y)^{\frac{1}{2}}\, dy}{x^{(1+\gamma)/2}(\log_2 x)^{\frac{1}{2}}}.$$

Using L'Hospital's rule, expanding in a Taylor series and then taking limits yields the desired result.

LEMMA 3. *If* $(1/c_n)\Sigma_{j=1}^{n} y_j \to s$, *then*

$$\frac{1}{b_n c_n}\Sigma_{j=1}^{n} b_j y_j \to (1 - \nu)s.$$

PROOF. Let $S_n = \Sigma_{j=1}^{n} y_j$, $S_0 = 0$

$$\frac{1}{b_n c_n}\Sigma_{j=1}^{n} b_j y_j = \frac{1}{b_n c_n}\Sigma_{j=1}^{n} b_j (s_j - s_{j-1})$$

$$= \frac{S_n}{c_n} - \frac{1}{b_n c_n}\Sigma_{j=1}^{n-1}(b_j - b_{j-1}) s_j$$

$$\to s - \nu s \qquad \text{by Lemma 2.}$$

THEOREM 5. *Let K satisfy* (1) *and* $\{h_n\}$ *satisfy* (2). *Let f satisfy the conditions of Theorem* 1. *If, in addition*

$$\frac{nh_n}{\log n(\log_2 n)} \quad \textit{diverges to} \quad \infty$$

and

$$\lim_{n\to\infty}\left(\frac{h_n}{n\log_2 n}\right)^{\frac{1}{2}}\Sigma_{j=1}^{n-1}\left(h_j^{-\frac{1}{2}} - h_{j+1}^{-\frac{1}{2}}\right)(j\log_2 j)^{\frac{1}{2}} = \nu < \infty,$$

then $(nh_n/\log_2 n)^{\frac{1}{2}}(f_n^*(x) - Ef_n^*(x)) \to (1-\nu)(2f(x)\int_{-\infty}^{\infty}K^2(u)\,du)^{\frac{1}{2}}$ a.s. *as* $n\to\infty$. *Moreover, if the conditions of Theorem* 1 *part* (b) *hold and* $h_n = bn^{-\gamma}$ *with* $\gamma \geqslant (1/(2r+1))$
then

$$\left(\frac{n^{1-\gamma}}{\log_2 n}\right)^{\frac{1}{2}}(f_n^*(x) - f(x)) \to \left(\frac{1}{\gamma+1}\right)(2f(x)\int_{-\infty}^{\infty}K^2(u)\,du)^{\frac{1}{2}} \text{ a.s. } \textit{as } n\to\infty.$$

PROOF. We observe that

$$nh_n^{\frac{1}{2}}(f_n^*(x) - Ef_n^*(x)) = \Sigma_{j=1}^{n}\frac{h_n^{\frac{1}{2}}}{h_j}\left[K\left(\frac{x-X_j}{h_j}\right) - EK\left(\frac{x-X_j}{h_j}\right)\right].$$

Identify $c_n = (n\log_2 n)^{\frac{1}{2}}$, $b_n = h_n^{-\frac{1}{2}}$, $y_n = (h_n)^{-\frac{1}{2}}[K((x-X_n)/h_n) - EK((x-X_n)/h_n)]$ in Lemma 3. The result follows from (7) and Lemma 3. Notice that $Ef_n^*(x) - f(x) = 0(n^{-\gamma r})$, hence,

$$\left(\frac{n^{1-\gamma}}{\log_2 n}\right)^{\frac{1}{2}}|Ef_n^*(x) - f(x)| = 0\left(\left(\frac{n^{1-\gamma(2r+1)}}{\log_2 n}\right)^{\frac{1}{2}}\right)$$

and the result follows as in Theorem 4.

4. A sequential procedure. One particularly useful application of recursively formulated density estimators is to sequential procedures. Davies and Wegman (1975) introduce sequential density estimation, studying in some detail rules of the form:

$$\text{Stop if} \quad |\hat{f}_n(x) - \hat{f}_{n-1}(x)| < \varepsilon, \qquad \text{otherwise continue.}$$

Carroll (1976), also introduces stopping rules based on fixed width interval estimation. In this section we introduce a rule which is illustrative of sequential density estimation using recursive estimators. For both the estimator, $f_n^*(x)$, and the estimator introduced in this paper, $f_n^\dagger(x)$, the correction term due to observation, X_n, is $1/nh_nK((x-X_n)/h_n)$. A reasonable stopping rule might be to stop when the correction, $1/nh_nK((x-X_n)/h_n)$, gets "too small". Unfortunately, if $nh_n\to\infty$ and K is bounded, $1/nh_nK((x-X_n)/h_n)$ gets "too small" independent of the observations. Thus we choose a stopping variable N_ε such that

$$\begin{aligned} N_\varepsilon &= \text{first } n \text{ such that } 1/h_nK((x-X_n)/h_n) < \varepsilon;\\ &= \infty \text{ if no such } n \text{ exists.}\end{aligned}$$

THEOREM 5. *We assume* (1) *and* (2) *hold for* K *and* $\{h_n\}$ *respectively.*

(a) $P[N_\varepsilon < \infty] = 1$, i.e., N_ε *is a closed stopping rule.*

(b) $EN_\varepsilon^k < \infty$ *for every* k. *Moreover, there is a number,* p, *with* $0 < p < 1$ *such that* Ee^{tN_ε} *exists for* $t < -\log p$.

(c) *If* $K(x) > 0$ *for all* x, *then* $N_\varepsilon \to \infty$ *in probability as* $\varepsilon \downarrow 0$.

(d) *If* $K(x) > 0$ *for all* x, *then* $N_\varepsilon \to \infty$ a.s. *as* $\varepsilon \downarrow 0$.

(e) *Under the hypotheses of Theorems* 3 *and* 4 *if* $K(x) > 0$ *for all* x,

$$f^*_{N_\varepsilon}(x) \to f(x) \text{ a.s. as } \varepsilon \downarrow 0$$

and

$$(1 - \gamma/2) f^\dagger_{N_\varepsilon}(x) \to f(x) \text{ a.s. as } \varepsilon \downarrow 0.$$

PROOF. Let X have density, f. We first observe

$$P[N_\varepsilon = n] = P\left[\frac{1}{h_1} K\left(\frac{x - X}{h_1}\right) \geqslant \varepsilon\right] \cdots P\left[\frac{1}{h_{n-1}} K\left(\frac{x - X}{h_{n-1}}\right) \geqslant \varepsilon\right] P\left[\frac{1}{h_n} K\left(\frac{x - X}{h_n}\right) < \varepsilon\right].$$

$$= p_1 \cdots p_{n-1}(1 - p_n)$$

where $p_j = P[1/h_j K((x - X)/h_j) \geqslant \varepsilon]$.

$$\begin{aligned} P[N_\varepsilon < \infty] &= \Sigma_{j=1}^\infty P[N_\varepsilon = j] \\ &= 1 - p_1 + p_1(1 - p_2) + \cdots + p_1 \cdots p_{n-1}(1 - p_n) + \cdots \\ &= 1. \end{aligned}$$

Since $|u|K(u) \to 0$ as $u \to \pm\infty$, it follows that $P[1/h_j K((x - X)/h_j) \geqslant \varepsilon] \to 0$ as $j \to \infty$, i.e., $p \to 0$ as $j \to \infty$. Let $0 < p < 1$, for j sufficiently large, say $j \geqslant n_p$, $p_j < p$.

Hence $EN^k = \Sigma_{n=1}^\infty n^k p[N_\varepsilon = n] \leqslant \Sigma_{n=1}^{n_p} n^k + \Sigma_{n=n_p+1}^\infty n^k p^{n-1-n_p} < \infty$. Similarly

$$Ee^{tN_\varepsilon} = \Sigma_{n=1}^\infty e^{tn} P[N_\varepsilon = n] \leqslant \Sigma_{n=1}^{n_p} e^{tn} + e^{t(1+n_p)} \Sigma_{n=n_p+1}^\infty (e^t p)^{n-1-n_p}.$$

This latter sum will be finite provided $e^t p < 1$ or $t < -\log p$.

To show (c), we note that $p_j \uparrow 1$ as $\varepsilon \downarrow 0$. But $P[N_\varepsilon \leqslant n] = 1 - p_1 \cdots p_n \to 0$ as $\varepsilon \downarrow 0$. Thus $P[N_\varepsilon > n] \to 1$ as $\varepsilon \downarrow 0$ for fixed n. Hence $N_\varepsilon \to \infty$ in probability as $\varepsilon \downarrow 0$.

Next let ω be any point in the basic probability space. We have $1/h_n K((x - X(\omega))/h_n) > 0$. Let N_0 be any positive integer. Choose $\varepsilon < \min_{1 \leqslant j \leqslant N_0} 1/h_j K((x - X(\omega))/h_j)$ (ε may depend on ω). Thus $N_\varepsilon(\omega) > N_0$. Taking lim inf $\varepsilon \downarrow 0$,

$$\liminf_{\varepsilon \downarrow 0} N_\varepsilon \geqslant N_0 \text{ a.s.}$$

But N_0 was arbitrary

$$\liminf_{\varepsilon \downarrow 0} N_\varepsilon = \infty \text{ a.s.}$$

Part (e) follows immediately.

A slightly more general stopping rule might be

$$N = \text{first } n \text{ such that } 1/g(h_n)K((x - X_n)/h_j) < \varepsilon.$$
$$= \infty \text{ if no such } n \text{ exists}$$

where $g(x)$ is some monotone nondecreasing function of x. To illustrate consider the rule

$$N = \text{first } n \text{ such that } 1/h_n^2 K(X_n/h_n) < \varepsilon$$
$$= \infty \text{ if no such } n \text{ exists.}$$

In this example, we presume $X_1, \ldots, X_n, \ldots$ is a $n(0, 1)$ sample and we are estimating $f(0)$. Let us assume that $K(x) = 1/\pi(1 + x^2)$, $-\infty < x < \infty$. We observe then

$$p_n = P\left[\frac{1}{h_n^2}K\left(\frac{X}{h_n}\right) > \varepsilon\right]$$
$$= 2P\left[0 < X < h_n\left(\frac{1}{\pi\varepsilon h_n^2} - 1\right)^{\frac{1}{2}}\right]$$
$$= 2\left(\Phi\left(\left(\frac{1}{\pi\varepsilon} - h_n^2\right)^{\frac{1}{2}}\right) - \Phi(0)\right)$$
$$= 2\Phi\left(\left(\frac{1}{\pi\varepsilon} - h_n^2\right)^{\frac{1}{2}}\right) - 1.$$

In this case, we notice that $p_n \uparrow 2\Phi(1/\pi\varepsilon)^{\frac{1}{2}} - 1$. Thus $P[N_\varepsilon = n]$ is very close to a geometric distribution. We also note here that, in general, we can compute the exact distribution of N_ε given K, $\{h_n\}$ and f.

One may argue that the rule proposed here is unsatisfactory since we stop on the basis of one observation. A more satisfactory rule might be based on the last M observations, where $0 < M < n$. One may even choose M an increasing function of n. Our present discussion, however, is not meant to be definitive, only illustrative of what might be done.

REFERENCES

CARROLL, R. J. (1976). On sequential density estimation. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **36** 137–151.

DAVIES, H. I. (1973). Strong consistency of a sequential estimator of a probability density function. *Bull. Math. Statist.* **15** 49–54.

DAVIES, H. I. and WEGMAN, E. J. (1975). Sequential nonparametric density estimation. *IEEE Trans. Information Theory* **IT-21** 619–628.

DAVIS, K. B. (1975). Mean square error properties of density estimates. *Ann. Statist.* **3** 1025–1030.

DEHEUVALS, P. (1974). Conditions nécessaires et suffisantes de convergence ponctuelle presque sûre et uniforme presque sûre des estimateurs de la densité. *C. R. Acad. Sci. Paris Ser. (A)* **278** 1217–1220.

EDDY, W. F. (1976). *Optimum Kernel Estimators of the Mode.* Ph.D. dissertation, Yale University.

JAIN, N. C., JOGDEO, K. and STOUT, W. F. (1975). Upper and lower functions for martingales and mixing processes. *Ann. Probability* **3** 119–145.

LOEVE, M. (1963). *Probability Theory*. D. Van Nostrand, Princeton, N.J.

PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076.

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

STRASSEN, V. (1964). An invariance principle for the law of the iterated logarithm. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **3** 211–226.

STRASSEN, V. (1965). Almost sure behaviour of sums of independent random variables and martingales. *Proc. Fifth Berkeley Symp. Math. Statist. Probability* **2** 315–343.

WOLVERTON, C. T. and WAGNER, T. J. (1969). Asymptotically optimal discriminant functions for pattern classification. *IEEE Trans. Information Theory* **IT-15** 258–265.

YAMATO, H. (1971). Sequential estimation of a continuous probability density function and mode. *Bull. Math. Statist.* **14** 1–12.

DIRECTOR,
STATISTICS AND PROBABILITY PROGRAM
OFFICE OF NAVAL RESEARCH
ARLINGTON, VIRGINIA 22217

MATHEMATICS DEPARTMENT
UNIVERSITY OF NEW ENGLAND
ARMIDALE, N.S.W.
AUSTRALIA 2351