Remarks on Some Recursive Estimators of a Probability Density

by

Edward J. Wegman[+]

Department of Statistics
University of North Carolina at Chapel Hill

and

H. I. Davies

Department of Mathematics
University of New England
Armidale, Australia

Institute of Statistics Mimeo Series No. 1021

July, 1975

+ The work of this author was supported by the Air Force Office of Scientific Research under Grant No. AFOSR-75-2840.

ABSTRACT

The density estimator of Yamato (1971), $f_n^*(x) = n^{-1} \sum_{j=1}^{n} h_j^{-1} K\big(x-X_j)/h_j\big)$, as well as a closely related one $f_n^\dagger(x) = n^{-1} h_n^{-\frac{1}{2}} \sum_{j=1}^{n} h_j^{-\frac{1}{2}} K\big((x-X_j)/h_j\big)$ are considered. Expressions for asymptotic bias, and variance are developed and weak consistency is shown. Using the almost sure invariance principle, laws of the iterated logarithm are developed. Finally application of these results to sequential estimation procedures are made.

Keywords: recursive estimators, asymptotic bias, asymptotic variance, weak consistency, almost sure invariance principle, law of the iterated logarithm, strong consistency, asymptotic distribution, sequential procedure.

A.M.S. Classification Nos. 62G05, 62L12, 60F20, 60G50.

I. Introduction. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. observations drawn according to a probability density, f. Rosenblatt (1956) introduced the kernel estimator of the density, $f(x)$,

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

and, in a now classic paper, Parzen (1962) developed many of the important properties of these estimators. A closely related estimator

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

was introduced by Yamato (1971). This latter estimator has the very useful property that it can be calculated recursively, i.e.

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

This property is particularly useful for fairly large sample sizes, since addition of a few extra observations means $\hat{f}_n(x)$ must be entirely re-computed - 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

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

This latter estimator can also be recursively formulated;

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

In addition, we will give a law of the iterated logarithm for $f_n^\dagger$, a rate of convergence for f_n^* and some properties of both when used as sequential density estimators.

Throughout this paper, we shall deal with univariate estimators. The extension to the multivariate case is straightforward. We shall assume throughout that

$$
(1) \qquad
\begin{aligned}
&K \text{ is symmetric about } 0, \\
&K(u) > 0 , \\
&\int_{-\infty}^{\infty} K(u)\,du < \infty , \\
&\sup_u K(u) < \infty , \\
&|u| K(u) \to 0 \quad \text{as} \quad u \to \pm\infty ,
\end{aligned}
$$

and that $\{h_n\}$ is a sequence of numbers such that

$$
(2) \qquad
\begin{aligned}
&h_n \to 0 \\
&nh_n \to \infty .
\end{aligned}
$$

Other assumptions on K and $\{h_n\}$ will be made as needed.

2. <u>Asymptotic Bias, Variance and Consistency</u>: Throughout this paper, it will be convenient to deal with the sum

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

We recall a useful lemma from Parzen (1962).

<u>Lemma 1</u>: Suppose $K(u)$ is a Borel function satisfying (1). Let $g(u)$ satisfy

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

and let $\{h_n\}$ satisfy (2). Then

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

Theorem 1: (a) Let K and $\{h_n\}$ satisfy (1) and (2). If f is a bounded density,

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

(b) Let us suppose K has Fourier transform K^* so that $K^*(u) = \int_{-\infty}^{\infty} e^{-iuy} K(y)\,dy$. Suppose further that for some r, $\lim_{u\to 0} \{[1-K^*(u)]/|u|^r\}$ is finite and that $f^{(r)}(x)$ exists. Then

$$|E\ f_n^*(x) - f(x)| \le 0\Big(\frac{1}{n} \sum_{j=1}^{n} h_n^r\Big) .$$

(c) Under the assumptions of (b), and choosing $h_n = bn^{-\gamma}$,

$$Ef_n^{\dagger}(x) \to \frac{f(x)}{1-\frac{\gamma}{2}} .$$

Proof: Now

$$nh_n \text{ var } f_n^{\dagger}(x) = nh_n \left[\frac{1}{n^2 h_n} \sum_{j=1}^{n} \frac{1}{h_j} \left[E\ K^2\left(\frac{x-X_j}{h_j}\right) - \left(E\ K\left(\frac{x-X_j}{h_j}\right)\right)^2 \right] \right]$$

$$= \frac{1}{n} \sum_{j=1}^{n} \left[\int_{-\infty}^{\infty} \frac{1}{h_j} K^2\left(\frac{x-u}{h_j}\right) f(u)\,du - \frac{1}{h_j}\left(\int_{-\infty}^{\infty} K\left(\frac{x-u}{h_j}\right) f(u)\,du\right)^2 \right] .$$

But making a simple change of variable

$$\frac{1}{h_j}\left(\int_{-\infty}^{\infty} K\left(\frac{x-u}{h_j}\right) f(u)\,du\right)^2 = h_j\left(\int_{-\infty}^{\infty} K(u) f(x-h_j u)\,du\right)^2 \to 0 \quad \text{as} \quad j\to\infty .$$

It follows that the Cesàro sum

$$\frac{1}{n} \sum_{j=1}^{n} h_j \left(\int_{-\infty}^{\infty} K(u) f(x-h_j u)\,du\right)^2 \to 0 .$$

Similarly,

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

Since K is a bounded function, $\int_{-\infty}^{\infty} K^2(u)\,du < \infty$. By Lemma 1,

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

hence the Cesàro sum,

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

The conclusion of (a) follows. Parzen (1962) shows

$$\frac{\int_{-\infty}^{\infty} \frac{1}{h_n} K\left(\frac{x-y}{h_n}\right) f(y)\,dy - f(x)}{h_n^r} \to k_r f^{(r)}(x)$$

where $k_r = \lim_{u\to 0} \{[1-K^*(u)]/|u|^r\}$. Clearly, there exists c_r such that

$$\left| \int_{-\infty}^{\infty} \frac{1}{h_n} K\left(\frac{x-y}{h_n}\right) f(y)\,dy - f(x) \right| \le c_r h_n^r \quad \text{for all} \quad n.$$

But

$$|E\ f_n^*(x) - f(x)| \le \frac{1}{n} \sum_{j=1}^{n} \left| \int_{-\infty}^{\infty} \frac{1}{h_j} K\left(\frac{x-y}{h_j}\right) f(y)\,dy - f(x) \right|$$

$$\le \frac{1}{n} \sum_{j=1}^{n} c_r h_n^r \ .$$

To observe the result for $f_n^\dagger(x)$,

$$f(x) - c_r h_n \le \int_{-\infty}^{\infty} \frac{1}{h_j} K\left(\frac{x-y}{h_j}\right) f(y)\,dy \le f(x) + c_r h_n^r \ .$$

Multiplying by $\sqrt{\frac{h_j}{h_n}}$, dividing by n and summing yields

$$\frac{1}{n} \sum_{j=1}^{n} \sqrt{\frac{h_j}{h_n}}\, f(x) - \frac{1}{\sqrt{h_n}\, n} c_r \sum_{j=1}^{n} h_j^{r+\frac{1}{2}} \le E\ f_n^\dagger(x) \le \frac{1}{n} \sum_{j=1}^{n} \sqrt{\frac{h_j}{h_n}}\, f(x) + \frac{1}{\sqrt{h_n}\, n} c_r \sum_{j=1}^{n} h_j^{r+\frac{1}{2}} \ .$$

Under the assumptions of (c),

$$\frac{1}{n}\sum_{j=1}^{n}\sqrt{\frac{h_j}{h_n}} = \frac{\frac{1}{n}\sum_{j=1}^{n} j^{-\gamma/2}}{n^{-\gamma/2}} .$$

Using integral approximations,

$$\frac{1}{n}\sum_{j=1}^{n}\sqrt{\frac{h_j}{h_n}} \to \frac{1}{1-\gamma/2} .$$

Similarly, using integral approximations

$$\frac{1}{\sqrt{h_n}\, n}\sum_{j=1}^{n} h_j^{r+\frac{1}{2}} \to 0 .$$

Thus

$$\lim_{n\to\infty} E\, f_n^{\dagger}(x) = \frac{f(x)}{1-\gamma/2} . \qquad \square$$

Using the integral approximation, it can be shown that if $h_n = bn^{-\gamma}$, then $\frac{1}{n}\sum_{j=1}^{n} h_j^r = 0(n^{-\gamma r}) = 0(h_n^r)$. Thus for $h_n = bn^{-\gamma}$, the Yamato estimator has the same rate of convergence for the bias term as the Parzen estimator. However this need not always be the case. For example, if $r=1$ and $h_n = b\left(\frac{\log n}{n}\right)$, then $\frac{1}{n}\sum_{j=1}^{n} h_j = 0\left(\frac{\log\log n \circ \log n}{n}\right) = 0(\log\log n \circ h_n)$. Thus the Yamato estimator may have worse bias properties than the standard kernel estimators.

3. <u>An Almost Sure Invariance Principle</u>. 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,

$$\sum_{j=1}^{n} \frac{1}{\sqrt{h_j}} \left(K\left(\frac{x-X_j}{h_j}\right) - E\, K\left(\frac{x-X_j}{h_j}\right)\right)$$

is with probability one close to Brownian motion in a sense made precise below. The asymptotic fluctuation behavior 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 $Y_1,\ldots,Y_n,\ldots$ be a sequence of zero mean random variables with finite second moments. Let $S_n = \sum_{j=1}^{n} Y_j$ and $V_n = \sum_{j=1}^{n} E[Y_j^2]$, $S_0=0=V_0$.

Theorem 2: For a fixed $\alpha\geq 0$, assume

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

and

$$\sum_{k=1}^{\infty} \frac{(\log_2 V_k)^{\alpha}}{V_k} E\left\{Y_k^2\, I_{\left[Y_k^2 > \frac{V_k}{\log V_k}(\log_2 V_k)^{2(\alpha+1)}\right]}\right\} < \infty \quad \text{a.s.} \tag{4}$$

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 $\{S(t),\ t\geq 0\}$, if necessary, on a new probability space, there exists a Brownian motion ξ such that

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

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

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

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

according as

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

Let us identify $Y_j = \frac{1}{\sqrt{h_j}} \left(K\left(\frac{x-X_j}{h_j}\right) - E\, K\left(\frac{x-X_j}{h_j}\right)\right)$. Thus we have,

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

$$= h_n\, E \sum_{j=1}^n \frac{1}{h_n h_j} \left[K\left(\frac{x-X_j}{h_j}\right) - E\, K\left(\frac{x-X_j}{h_j}\right) \right]^2$$

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

But under the assumptions of Theorem 1

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

so that $\frac{V_n}{n} = h_n n \operatorname{var} f_n^\dagger(x) \to f(x) \int_{-\infty}^\infty K^2(u)du$. Thus for $\varepsilon>0$ and for n sufficiently large,

$$c_1 n = \left(f(x) \int_{-\infty}^\infty K^2(u)du - \varepsilon\right)n \le V_n \le \left(f(x) \int_{-\infty}^\infty K^2(u)du + \varepsilon\right)n = c_2 n .$$

<u>Theorem 3</u>: (a) Let K satisfy (1) and $\{h_n\}$ satisfy (2). Let f satisfy the conditions of Theorem 1. If in addition,

(6) $$\frac{nh_n}{\log n(\log_2 n)^{2(\alpha+1)}} \text{ diverges to } \infty,$$

then (5) holds for S_n defined above.

(b) In particular, if

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

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

according as

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

(c) For $\alpha \geq 0$

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

(d) For $\alpha > 1$,

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

<u>Proof</u>: Consider first

$$E\, Y_k^2\, I_{\left[Y_k^2 > \frac{V_k}{\log V_k (\log_2 V_k)^{2(\alpha+1)}}\right]} .$$

Now

$$\frac{V_k}{\log V_k (\log_2 V_k)^{2(\alpha+1)}} \geq \frac{c_1 k}{\log c_2 k (\log_2 c_2 k)^{2(\alpha+1)}}$$
$$\geq c^* \frac{k}{\log k (\log_2 k)^{2(\alpha+1)}} ,$$

where c^* is some constant. Thus

$$E\, Y_k^2\, I_{\left[Y_k^2 > \frac{V_k}{\log V_k (\log_2 V_k)^{2(\alpha+1)}}\right]} \leq E\, Y_k^2\, I_{\left[Y_k^2 \geq c^* \frac{k}{\log k (\log_2 k)^{2(\alpha+1)}}\right]} .$$

But

$$Y_k^2 \geq c^* \frac{k}{\log k(\log_2 k)^{2(\alpha+1)}}$$

if and only if

$$\left[K\left(\frac{x-X_k}{h_k}\right) - E\,K\left(\frac{x-X_k}{h_k}\right) \right]^2 \geq c^* \frac{h_k k}{\log k(\log_2 k)^{2(\alpha+1)}} .$$

Since K is bounded and (6) holds,

$$\left[\left[K\left(\frac{x-X_k}{h_k}\right) - E\,K\left(\frac{x-X_k}{h_k}\right) \right] \geq c^* \frac{h_k k}{\log k(\log_2 k)^{2(\alpha+1)}} \right]$$

is an impossible event for k sufficiently large. Thus

$E\, Y_k^2\, I_{\left[Y_k^2 > \frac{V_k}{\log V_k (\log_2 V_k)^{2(\alpha+1)}}\right]} = 0$ for k sufficiently large. It follows that

$$\sum_{k=1}^{\infty} \frac{(\log_2 V_k)^{\alpha}}{V_k} E\left\{Y_k^2\, I_{\left[Y_k^2 > \frac{V_k}{\log V_k (\log_2 V_k)^{2(\alpha+1)}}\right]}\right\} < \infty \text{ a.s.}$$

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

$$\left|\frac{S(t)}{(2\circ\log_2 t\circ t)^{\frac{1}{2}}} - \frac{\xi(t)}{(2\circ\log_2 t\circ t)^{\frac{1}{2}}}\right| \leq \frac{0\left(t^{\frac{1}{2}}(\log_2 t)^{\frac{1-\alpha}{2}}\right)}{(2\circ\log_2 t\circ t)^{\frac{1}{2}}}$$

$$= 0\left((\log_2 t)^{-\alpha/2}\right) .$$

But $\frac{\xi(t)}{(2\circ\log_2 t\circ t)^{\frac{1}{2}}} \to 1$ a.s. as $t\to\infty$, hence for $\alpha\geq 0$,

$$\frac{S(t)}{(2\circ\log_2 t\circ t)^{\frac{1}{2}}} \to 1 \text{ a.s. as } t\to\infty .$$

Thus

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

But

$$\left(\frac{\log_2 n\cdot n}{\log_2 V_n\cdot V_n}\right)^{\frac{1}{2}} \to \left(\frac{1}{f(x)\int_{-\infty}^{\infty}K^2(u)du}\right)^{\frac{1}{2}} .$$

Hence

$$\frac{S_n}{(\log_2 n\cdot n)^{\frac{1}{2}}} \to \left(2f(x)\int_{-\infty}^{\infty} K^2(u)du\right)^{\frac{1}{2}} \text{ a.s. as } n\to\infty . \tag{7}$$

Finally noting that $n\sqrt{h_n}\left(f_n^{\dagger}(x) - E\, f_n^{\dagger}(x)\right) = S_n$, we have

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

Hence part (c). For part (d), we observe that since $\xi(t)$ is Brownian motion $\xi(t)/\sqrt{t}$ is normal mean zero variance $1\left(n(0,1)\right)$. But

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

For $\alpha>1$,

$$\frac{S(t)}{\sqrt{t}} \text{ is asymptotically } n(0,1).$$

Letting $t=V_n$

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

But $V_n = n^2 h_n \text{var}\, f_n^{\dagger}(x) = \text{var}\, n\sqrt{h_n}\, f_n^{\dagger}(x)$. Also $S_n = \sum_{j=1}^{n} Y_j = n\sqrt{h_n}\left(f_n^{\dagger}(x) - E\, f_n^{\dagger}(x)\right)$

$$\frac{f_n^{\dagger}(x)-E\, f_n^{\dagger}(x)}{\sqrt{\text{var}\, f_n^{\dagger}(x)}} \text{ is asymptotically } n(0,1). \qquad \square$$

While we know the exact order of $f_n^{\dagger}(x) - E\, f_n^{\dagger}(x)$, the fact that $f_n^{\dagger}(x)$ is a biased estimator and the fact that we do not have any rate on the bias term limits the usefulness of $f_n^{\dagger}$. Of course, γ is a parameter of h_n and hence known to us. We could therefore consider $(1-\gamma)f_n^{\dagger}(x)$ which would be asymptotically unbiased.

Combining this result with Theorem 1 part (a) and Theorem 3 part (c) yields a weakly and a strongly consistent estimator respectively.

Results for $f_n^\dagger$ can be translated to results for f_n^* by the next two very useful lemmas. These were suggested by the Toeplitz Lemma and the Kronecker Lemma. See Loève (1963).

<u>Lemma 2</u>: Let $b_n \uparrow \infty$, $c_n \uparrow \infty$ and s_n be sequences such that $s_n/c_n \to s$. Let $a_j = b_j - b_{j-1}$, $j \ge 2$ with $a_1 = b_1$, then $\frac{1}{b_n c_n} \sum_{j=1}^{n-1} a_j s_j \to s$.

<u>Proof</u>: Note that $b_n = \sum_{j=1}^{n} a_j$. Let $\varepsilon > 0$. There is N_ε such that $n > N_\varepsilon$ implies

$$s-\varepsilon \le \frac{s_n}{c_n} \le s+\varepsilon .$$

Let $s_n' = \frac{1}{b_n c_n} \sum_{j=1}^{n-1} a_j s_j$. Then

$$\frac{1}{b_n c_n} \sum_{j=1}^{N_\varepsilon} a_j s_j + \frac{1}{b_n} \sum_{j=N_\varepsilon+1}^{n-1} a_j (s-\varepsilon) \le s_n' \le \frac{1}{b_n c_n} \sum_{j=1}^{N_\varepsilon} a_j s_j + \frac{1}{b_n} \sum_{j=N_\varepsilon+1}^{n-1} a_j (s+\varepsilon) .$$

Taking lim inf and lim sup,

$$s-\varepsilon \le \liminf s_n' \le \limsup s_n' \le s+\varepsilon . \qquad \square$$

<u>Lemma 3</u>: If $\frac{1}{c_n} \sum_{j=1}^{n} y_j \to s$ and $b_n \uparrow \infty$, then

$$\frac{1}{b_n c_n} \sum_{j=1}^{n} b_j y_j \to 0 .$$

<u>Proof</u>: Let $s_n = \sum_{j=1}^{n} y_j$, $s_0 = 0$ and $a_j = b_j - b_{j-1}$ with $a_1 = b_1$. Then

$$\begin{aligned} \frac{1}{b_n c_n} \sum_{j=1}^{n} b_j y_j &= \frac{1}{b_n c_n} \sum_{j=1}^{n} b_j (s_j - s_{j-1}) \\ &= \frac{s_n}{c_n} - \frac{1}{b_n c_n} \sum_{j=1}^{n-1} (b_j - b_{j-1}) s_j \\ &= \frac{s_n}{c_n} - \frac{1}{b_n c_n} \sum_{j=1}^{n-1} a_j s_j . \end{aligned}$$

Using Lemma 2, $\frac{1}{b_n c_n} \sum_{j=1}^{n} b_j y_j \to s-s = 0.$ □

Theorem 4: Let K satisfy (1) and $\{h_n\}$ satisfy (2). Let f satisfy the conditions of Theorem 1. If in addition, $\frac{nh_n}{\log n(\log_2 n)^2}$ diverges to ∞, then

$$\left(\frac{nh_n}{\log_2 n}\right)^{\frac{1}{2}} \left(f_n^*(x) - E\, f_n^*(x)\right) \to 0 \text{ a.s. as } n\to\infty .$$

Moreover if the conditions of Theorem 1 part (b) hold and $h_n = bn^{-\gamma}$ with $\gamma \geq \frac{1}{2r+1}$, then

$$\left(\frac{n^{1-\gamma}}{\log_2 n}\right)^{\frac{1}{2}} \left(f_n^*(x) - f(x)\right) \to 0 \text{ a.s. as } n\to\infty .$$

Proof: We observe that

$$n\sqrt{h_n}\left(f_n^*(x) - E\, f_n^*(x)\right) = \sum_{j=1}^{n} \frac{\sqrt{h_n}}{h_j}\left[K\left(\frac{x-X_j}{h_j}\right) - E\, K\left(\frac{x-X_j}{h_j}\right)\right] .$$

Identify $c_n = (n \log_2 n)^{\frac{1}{2}}$, $b_n = \frac{1}{\sqrt{h_n}}$, $y_n = \frac{1}{\sqrt{h_n}}\left[K\left(\frac{x-X_n}{h_n}\right) - E\, K\left(\frac{x-X_n}{h_n}\right)\right]$ in Lemma 3. The result follows from (7) and Lemma 3.

Notice that $E\, f_n^*(x) - f(x) = 0(n^{-\gamma r})$, hence,

$$\left(\frac{n^{1-\gamma}}{\log_2 n}\right)^{\frac{1}{2}} \left|E\, f_n^*(x) - f(x)\right| \leq 0\left(\left(\frac{n^{1-\gamma(2r+1)}}{\log_2 n}\right)^{\frac{1}{2}}\right) .$$

Since $\gamma \geq \frac{1}{2r+1}$, $1-\gamma(2r+1) \leq 0$, so that

$$\left(\frac{n^{1-\gamma}}{\log_2 n}\right)^{\frac{1}{2}} \left|E\, f_n^*(x) - f(x)\right| \to 0 \text{ as } n\to\infty .$$ □

It follows that the Yamato estimator, $f_n^*(x)$, while possibly somewhat worse in terms of bias is better in terms of variance. In fact, Yamato (1971, p. 6) concludes under suitable conditions that if $\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \frac{h_n}{h_j} = \alpha$, then $\lim_{n\to\infty} nh_n \operatorname{var}[f_n^*(x)] = \alpha f(x) \int_{-\infty}^{\infty} K^2(y)dy$.

In fact, we can apply Lemma 3 again to the expression $\frac{1}{n}\sum_{j=1}^{n}\frac{h_n}{h_j}$. Let $c_n=n$, $b_n=\frac{1}{h_n}$, and $y_n\equiv 1$. Clearly, $\frac{1}{c_n}\sum_{j=1}^{n} y_j = \frac{1}{n}\cdot n = 1 \to 1$. Hence $\frac{1}{n}\sum_{j=1}^{n}\frac{h_n}{h_j} \to 0$, so that $\alpha\equiv 0$. Hence for the Yamato estimator,

$$\lim_{n\to\infty} nh_n \operatorname{var}[f_n^*(x)] = 0.$$

5. <u>A Sequential Procedure</u>. 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:

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

In this section we shall discuss briefly a rule suggested by the recursive estimator itself. For both the Yamato estimator, $f_n^*(x)$, and the estimator introduced in this paper, $f_n^\dagger(x)$, the correction term due to observation, X_n, is $\frac{1}{nh_n}K\left(\frac{x-X_n}{h_n}\right)$. A reasonable stopping rule might be to stop when $\frac{1}{nh_n}K\left(\frac{x-X_n}{h_n}\right)$ gets "too small". Unfortunately, since $nh_n\to\infty$ and K is bounded, $\frac{1}{nh_n}K\left(\frac{x-X_n}{h_n}\right)$ gets "too small" independent of the observations. Thus we choose a stopping variable N_ε such that

$$N_\varepsilon = \begin{cases} \text{First } n \text{ such that } \frac{1}{h_n}K\left(\frac{x-X_n}{h_n}\right) < \varepsilon \\ \infty \text{ if no such } n \text{ exists.} \end{cases}$$

<u>Theorem 5</u>: 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 $E\,e^{tN_\varepsilon}$ 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 and if $K(x)>0$ for all x,

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

and

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

Proof: Let X have density, f. We first observe

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

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

where $p_j = P\left[\frac{1}{h_j} K\left(\frac{x-X}{h_j}\right) \geq \varepsilon\right]$.

$$P[N_\varepsilon < \infty] = \sum_{j=1}^{\infty} P[N_\varepsilon = j] = 1-p_1 + p_1(1-p_2) + \ldots + p_1 \cdots p_{n-1}(1-p_n) + \ldots$$
$$= 1 .$$

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

Hence $E\, N^k = \sum_{n=1}^{\infty} n^k P[N_\varepsilon = n] \leq \sum_{n=1}^{n_p} n^k + \sum_{n=n_p+1}^{\infty} n^k p^{n-1-n_p} < \infty$. Similarly

$$E\, e^{tN_\varepsilon} = \sum_{n=1}^{\infty} e^{tn} P[N_\varepsilon = n] \leq \sum_{n=1}^{n_p} e^{tn} + e^{t(1+n_p)} \sum_{n=n_p+1}^{\infty} (e^t p)^{n-1-n_p} .$$

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 $P[N_\varepsilon \leq n] = 1-p_n \to 0$ 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 $\frac{1}{h_n} K\left(\frac{x-X(\omega)}{h_n}\right) > 0$. Let N_0 be any positive integer. Choose $\varepsilon < \min_{1 \le j \le N_0} \frac{1}{h_j} K\left(\frac{x-X(\omega)}{h_j}\right)$ (ε may depend on ω). Thus $N_\varepsilon(\omega) > N_0$. Taking lim inf $\varepsilon \downarrow 0$,

$$\liminf_{\varepsilon\to 0} N_\varepsilon \ge N_0 \text{ a.s.}$$

But N_0 was arbitrary

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

Part (e) follows immediately. □

A slightly more general stopping rule might be

$$N = \begin{cases} \text{First } n \text{ such that } \frac{1}{g(h_n)} K\left(\frac{x-X_n}{h_n}\right) < \varepsilon \\ \infty \text{ if no such } n \text{ exists .} \end{cases}$$

Were $g(x)$ in some monotone non-decreasing function of x. To illustrate consider the rule

$$N = \begin{cases} \text{First } n \text{ such that } \frac{1}{h_n^2} K\left(\frac{X_n}{h_n}\right) < \varepsilon \\ \infty \text{ if no such } n \text{ exists .} \end{cases}$$

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 $K(x) = \frac{1}{\pi(1+x^2)}$, $-\infty < x < \infty$. We observe then

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

In this case, we notice that $p_n \uparrow 2\Phi(\sqrt{\frac{1}{\pi\varepsilon}}) - 1$. 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 the knowledge of K, $\{h_n\}$ and f.

REFERENCES

Davies, H.I. and Wegman, E.J. (1975), Sequential nonparametric density estimation, to appear *IEEE Transactions on Information Theory*, November, (1975).

Jain, N.C., Jogdeo, K., and Stout, W.F. (1975), Upper and lower functions for martingales and mixing processes, *Ann. Prob.*, 3, pp. 119-145.

Loève, 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, pp. 1065-1076.

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

Strassen, V. (1964), An invariance principle for the law of the iterated logarithm, *A. Wahrscheinlichkeitstheorie*, 3, pp. 211-226.

Strassen, V. (1965), Almost sure behavior of sums of independent random variables and martingales, *Proc. Fifth Berkeley Symp. Math. Statist. Prob.*, 2, pp. 315-343.

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