SEQUENTIAL NONPARAMETRIC DENSITY ESTIMATION

H.I. Davies & Edward J. Wegman

Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 884

August, 1973

SEQUENTIAL NONPARAMETRIC DENSITY ESTIMATION

by

H.I. Davies[1]

and

Edward J. Wegman

1. Introduction: In this paper, we shall discuss a sequential approach to probability density estimation. For the most part we shall confine our attention to estimators of the form

$$\hat{f}_n(x) = \frac{1}{n}\sum_{j=1}^{n} \frac{1}{h_n} K\left(\frac{x-X_j}{h_n}\right) \tag{1.1}$$

first introduced by Rosenblatt (1956) and discussed in greater detail by Parzen (1962). Here, of course, $X_1, X_2, \ldots, X_n$ are i.i.d. random variables chosen according to some density, f. In this paper, the function K, the so-called kernel, is assumed to be a bounded density on the real line satisfying

$$\lim_{u\to\pm\infty} |u|K(u) = 0 . \tag{1.2}$$

Moreover, the sequence, h_n, is assumed to be a sequence of positive real numbers satisfying

$$\lim_{n\to\infty} h_n = 0, \qquad \lim_{n\to\infty} nh_n = \infty \quad \text{and} \quad \lim_{n\to\infty} \frac{h_{n+1}}{h_n} = 1 . \tag{1.3}$$

---

[1] The work of this author was supported by a C.S.I.R.O. postgraduate studentship.

We shall principally focus our attention on a naive stopping rule defined by the following procedure:

Choose successive random samples of size M and form the differences

$$V_n(x) = \hat{f}_{nM}(x) - \hat{f}_{(n-1)M}(x) \tag{1.4}$$

where $\hat{f}_{nM}(x)$ and $\hat{f}_{(n-1)M}(x)$ are the density estimators based on sample sizes nM and $(n-1)M$ respectively. The stopping rule is

$$N(\varepsilon,M) = \begin{cases} \text{First } n \text{ such that } |V_n(x)| < \varepsilon \text{ for fixed } \varepsilon > 0 \\ \infty \text{ if no such } n \text{ exists.} \end{cases} \tag{1.5}$$

In section 2, we investigate the asymptotic structure of $V_n(x)$. In section 3, we investigate properties of the stopping variable, $N(\varepsilon,M)$. Finally section 4 is a concluding section.

2. Asymptotic Structure of $V_n(x)$:

Theorem 2.1: If K and h_n satisfy (1.2) and (1.3) respectively then

i. $|V_n(x)| \to 0$ in probability for every $x \in C(f)$, the continuity points of f,

and ii. $\sup_x |V_n(x)| \to 0$ in probability if f is uniformly continuous.

If, in addition, for some $\alpha > 0$

iii. $\sup_{|U| \geq a} |U|^m \{K(cu) - K(u)\}^2$ is locally Lipschitz of order α at $c=1$ for some $a>0$,

iv. $\int_{-\infty}^{\infty} \{K(cu) - K(u)\}^2 du$ is locally Lipschitz of order α at $c=1$,

and finally

v. $\sum_{n=1}^{\infty} \frac{1}{nh_n^{1-\beta}} \left| \frac{1}{h_{n+1}} - \frac{1}{h_n} \right|^{\beta} < \infty$, where $\beta = \min\{\frac{1}{2}\alpha, 1\}$

then

vi. $|V_n(x)| \to 0$ with probability one for every $x \in C(f)$.

Proof: i). Under the stated conditions, $\hat{f}_n(x) \to f(x)$ in probability, Parzen, (1962). Hence $\{\hat{f}_n(x)\}$ is a Cauchy sequence in probability, so that in probability $|V_n(x)| = |\hat{f}_{nM}(x) - \hat{f}_{(n-1)M}(x)| \to 0$. Results ii and vi follow in a similar way. Conditions iii, iv and v are those of Van Ryzin, (1969) for strong consistency. Alternate sufficient conditions were given by Nadaraya, (1965) which may be used as replacements for iii, iv, and v. □

The next sequence of results concerns the asymptotic variance structure.

Lemma 2.2: Let $K(y)$ be a piecewise continuous Borel function satisfying (1.2) and let g be a real function in L_1. Then if $\{h_n\}$ satisfies (1.3)

$$g_n(x) = \int_{-\infty}^{\infty} \frac{1}{h_n} K\left(\frac{y}{h_n}\right) K\left(\frac{y}{h_{n-1}}\right) g(x-y)dy$$

converges to $g(x)\int_{-\infty}^{\infty} K^2(y)dy$ for every $x \in C(g)$.

Proof: The proof is in two stages. First we show

(A) $$\lim_{n\to\infty} g_n(x) = \lim_{n\to\infty} \frac{g(x)}{h_n} \int_{-\infty}^{\infty} K\left(\frac{y}{h_n}\right) K\left(\frac{y}{h_{n-1}}\right) dy$$

and then

$$\text{(B)} \qquad \lim_{n\to\infty} \frac{1}{h_n} \int_{-\infty}^{\infty} K\left(\frac{y}{h_n}\right) K\left(\frac{y}{h_{n-1}}\right) dy = \int_{\infty}^{\infty} K^2(y)dy \ .$$

Proof of (A): Consider λ_n defined by

$$\lambda_n = \left| g_n(x) - g(x) \frac{1}{h_n} \int_{-\infty}^{\infty} K\left(\frac{y}{h_n}\right) K\left(\frac{y}{h_{n-1}}\right) dy \right|$$

$$= \left| \int_{-\infty}^{\infty} \frac{1}{h_n}[g(x-y)-g(x)] K\left(\frac{y}{h_n}\right) K\left(\frac{y}{h_{n-1}}\right) dy \right| \ .$$

Letting $\delta > 0$,

$$\lambda_n \le \max_{|y|\le\delta} |g(x-y)-g(x)| \int_{|y|\le\delta} K\left(\frac{y}{h_n}\right) K\left(\frac{y}{h_{n-1}}\right) \frac{dy}{h_n} +$$

$$\frac{1}{h_n} \int_{|y|>\delta} |g(x-y)| K\left(\frac{y}{h_n}\right) K\left(\frac{y}{h_{n-1}}\right) dy +$$

$$\int_{|y|>\delta} |g(x)| K\left(\frac{y}{h_n}\right) K\left(\frac{y}{h_{n-1}}\right) \frac{dy}{h_n} .$$

In the first and third terms of the R.H.S., we make the transformation $z = y/h_n$, so that

$$\lambda_n \le \max_{|y|\le\delta} |g(x-y)-g(x)| \int_{|z|\le\delta/h_n} K(z) K\left(\frac{h_n}{h_{n-1}} z\right) dz +$$

$$\int_{|y|>\delta} \frac{|g(x-y)|}{y} \circ \frac{y}{h_n} K\left(\frac{y}{h_n}\right) K\left(\frac{y}{h_{n-1}}\right) dy +$$

$$|g(x)| \int_{|z|\ge\delta/h_n} K(z) \ K\left(\frac{h_n}{h_{n-1}} z\right) dz \ .$$

Since K is bounded, the first term can be made arbitrarily small by choosing δ arbitrarily small. The second term is bounded by

$$\frac{1}{\delta} \sup_{|z|\geq\delta/h_n} \left|zK(z)\ K\left(\frac{h_n}{h_{n-1}}z\right)\right| \int_{-\infty}^{\infty} |g(y)|dy$$

which along with the third term can be made arbitrarily small for choice of n sufficiently large.

Proof of (B):

Letting $z = y/h_n$,

$$\left| \int_{-\infty}^{\infty} \frac{1}{h_n}\left[K\left(\frac{y}{h_{n-1}}\right)-K\left(\frac{y}{h_n}\right)\right]K\left(\frac{y}{h_n}\right)dy\right| = \left|\int_{-\infty}^{\infty}\left[K\left(\frac{h_n}{h_{n-1}}z\right)-K(z)\right]K(z)dz\right|$$

$$\leq \int_{-\infty}^{\infty}\left|K\left(\frac{h_n}{h_{n-1}}z\right)-K(z)\right|K(z)dz \ .$$

Clearly this last integrand is bounded by $[2 \sup_y K(y)] \cdot K(z)$ and hence appealing to the Lebesgue Dominated Convergence theorem completes the result. □

Theorem 2.3: Let K and h_n satisfy (1.2) and (1.3) respectively and also let

$$\lim_{n\to\infty} n\left\{\frac{h_{nM}}{h_{(n-1)M}} - 1\right\} = 1 - \nu, \quad \nu < \infty$$

then $\lim_{n\to\infty} n^2Mh_{nM}\ \mathrm{var}(V_n(x)) = \nu f(x) \int_{\infty}^{\infty} K^2(u)du$.

Proof: By definition,

$$V_n(x) = \frac{1}{nM}\sum_{j=1}^{nM}\frac{1}{h_{nM}}K\left(\frac{x-X_j}{h_{nM}}\right) - \frac{1}{(n-1)M}\sum_{j=1}^{(n-1)M}\frac{1}{h_{(n-1)M}}K\left(\frac{x-X_j}{h_{(n-1)M}}\right) .$$

Since $X_1,\ldots,X_{nM}$ are i.i.d.

$$(2.1)\qquad \operatorname{var}(V_n(x)) = \frac{1}{nMh_{nM}^2}\operatorname{var}\left(K\left(\frac{x-X_1}{h_{nM}}\right)\right) + \frac{1}{(n-1)Mh_{(n-1)M}^2}\operatorname{var}\left(K\left(\frac{x-X_1}{h_{(n-1)M}}\right)\right)$$

$$- \frac{2}{nMh_{nM}h_{(n-1)M}}\operatorname{cov}\left[K\left(\frac{x-X_1}{h_{nM}}\right), K\left(\frac{x-X_1}{h_{(n-1)M}}\right)\right]$$

Parzen (1962) shows

$$\lim_{n\to\infty} \frac{1}{h_{nM}}\operatorname{var}\left(K\left(\frac{x-X_1}{h_{nM}}\right)\right) = f(x)\int_{-\infty}^{\infty} K^2(u)du \;.$$

In a similar manner, using Lemma 2.2, one may also show

$$\lim_{n\to\infty} \frac{1}{h_{nM}}\operatorname{cov}\left[K\left(\frac{x-X_1}{h_{nM}}\right), K\left(\frac{x-X_1}{h_{(n-1)M}}\right)\right] = f(x)\int_{-\infty}^{\infty} K^2(u)du \;,$$

so that

$$\lim_{n\to\infty} n^2Mh_{nM}\operatorname{var}(V_n(x)) = \lim_{n\to\infty}\left\{n\,\frac{1}{h_{nM}}\operatorname{var}\left(K\left(\frac{x-X_1}{h_{nM}}\right)\right) + \frac{n^2}{(n-1)}\,\frac{h_{nM}}{h^2_{(n-1)M}}\operatorname{var}\left(K\left(\frac{x-X_1}{h_{(n-1)M}}\right)\right)\right.$$

$$\left. - 2n\frac{h_{nM}}{h_{(n-1)M}}\cdot\frac{1}{h_{nM}}\operatorname{cov}\left[K\left(\frac{x-X_1}{h_{nM}}\right), K\left(\frac{x-X_1}{h_{(n-1)M}}\right)\right]\right\}$$

$$= \lim_{n\to\infty}\left\{n+\frac{n^2}{(n-1)}\,\frac{h_{nM}}{h_{(n-1)M}} - 2n\frac{h_{nM}}{h_{(n-1)M}}\right\}f(x)\int_{-\infty}^{\infty} K^2(u)du$$

$$= \lim_{n\to\infty}\left\{n+n\left[\frac{n}{n-1}-2\right] + \left[\frac{n}{n-1}-2\right]n\left[\frac{h_{nM}}{h_{(n-1)M}}-1\right]\right\} f(x)\int_{-\infty}^{\infty} K^2(u)du \;.$$

After suitable simplification,

$$\lim_{n\to\infty} n^2Mh_{nM}\operatorname{var}(V_n(x)) = \{1-(1-\nu)\}f(x)\int_{-\infty}^{\infty} K^2(u)du$$

$$= \nu f(x)\int_{-\infty}^{\infty} K^2(u)du \;.$$

The results of Theorem 2.3 may be refined by the decomposition of Theorem 2.4 .

Theorem 2.4: $V_n(x)$ can be decomposed into the sum of two independent random variables, $A_n(x)$ and $B_n(x)$, such that under the conditions of Theorem 2.3

$$\lim_{n\to\infty} n^2 M h_{nM} \operatorname{var}(A_n(x)) = (\nu-1) f(x) \int_{-\infty}^{\infty} K^2(u)du$$

and

$$\lim_{n\to\infty} n^2 M h_{nM} \operatorname{var}(B_n(x)) = f(x) \int_{-\infty}^{\infty} K^2(u)du \ .$$

Proof:

$$V_n(x) = \sum_{j=1}^{(n-1)M} \left[\frac{1}{nMh_{nM}} K\left(\frac{x-X_j}{h_{nM}}\right) - \frac{1}{(n-1)Mh_{(n-1)M}} K\left(\frac{x-X_j}{h_{(n-1)M}}\right)\right]$$

$$+ \frac{1}{nMh_{nM}} \sum_{j=(n-1)M+1}^{nM} K\left(\frac{x-X_j}{h_{nM}}\right)$$

$$= A_n(x) + B_n(x).$$

Since $A_n(x)$ depends only on $X_1,\ldots,X_{(n-1)M}$ and $B_n(x)$ only on $X_{(n-1)M+1},\ldots,X_{nM}$, $A_n(x)$ and $B_n(x)$ are independent. Now $\operatorname{var} B_n(x) = \frac{1}{n^2M^2h_{nM}^2} M \operatorname{var} K((x-X_1)/h_{nM})$, so that

$$\lim_{n\to\infty} n^2 M h_{nM} \operatorname{var} B_n(x) = \lim_{n\to\infty} \frac{1}{h_{nM}} \operatorname{var} K\left(\frac{x-X_1}{h_{nM}}\right) = f(x) \int_{-\infty}^{\infty} K^2(u)du \ .$$

The result for $A_n(x)$ follows from the fact that $\operatorname{var} A_n(x) = \operatorname{var} V_n(x) - \operatorname{var} B_n(x)$. □

Notice that $B_n(x)$ is a finite sum of M terms, each identically distributed. If the density were known, then for suitable conditions on K the exact distribution could be found. It would be an M-fold convolution of densities of random variables of the form

$$Z_k = \frac{1}{nMh_{nM}} K\left(\frac{x-X_k}{h_{nM}}\right) .$$

Also, it is not difficult to show that $B_n(x)$ is bounded by $\frac{1}{nh_{nM}} \sup_u K(u)$. We also note here that $h_n = Bn^{-\alpha}$ with $0 < \alpha < 1$ and B a non-negative constant, satisfies the hypotheses of Theorem 2.3 with $\nu = 1 + \alpha$.

We close this section by demonstrating the asymptotic normality of $A_n(x)$. Lemma 2.5 below follows in a manner similar to Theorem 2.3 so the proof is omitted.

Lemma 2.5: If K and h_n satisfy (1.2) and (1.3) respectively and if

$$\lim_{n\to\infty} n\left\{\frac{h_{nM}}{h_{(n-1)M}} -1\right\} = 1 - \nu, \quad \nu, \text{ a constant}$$

then $A_{n1}(x)$ defined by

$$A_{n1}(x) = \frac{(n-1)}{nh_{nM}} K\left(\frac{x-X_1}{h_{nM}}\right) - \frac{1}{h_{(n-1)M}} K\left(\frac{x-X_1}{h_{(n-1)M}}\right)$$

satisfies

$$\lim_{n\to\infty} nMh_{nM} \operatorname{var}(A_{n1}(x)) = (\nu-1)f(x)\int_{-\infty}^{\infty} K^2(u)du \quad .$$

Lemma 2.6: If K and h_n satisfy (1.2) and (1.3) respectively and if

i) $$\lim_{n\to\infty} n\left\{\frac{h_{nM}}{h_{(n-1)M}} -1\right\} = 1 -\nu, \quad \nu, \text{ a constant}$$

and finally

ii) if for every sequence of real numbers $c_n \to 1$, $K(c_n u) \to K(u)$ uniformly in u,

then

$$\lim_{n\to\infty} n^2 h_{nM}^2 E|A_{n1}(x)|^3 = 0 .$$

<u>Proof</u>: $E|A_{n1}(x)|^3 = \int_{\infty}^{\infty} |\frac{(n-1)}{nh_{nM}} K(\frac{x-y}{h_{nM}}) - \frac{1}{h_{(n-1)M}} K(\frac{x-y}{h_{(n-1)M}})|^3 f(y)dy$

$$= \frac{1}{n^3 h_{nM}^3} \int_{-\infty}^{\infty} |(n-1)K(\frac{x-y}{h_{nM}}) - n \frac{h_{nM}}{h_{(n-1)M}} K(\frac{x-y}{h_{(n-1)M}})|^3 f(y)dy.$$

Multiplying both sides by $n^2h_{nM}^2$ and making the transformation $u = \frac{x-y}{h_{nM}}$,

$$n^2h_{nM}^2 \ E|A_{n1}(x)|^2 = \frac{1}{n} \int_{\infty}^{\infty} |(n-1)K(u) - \frac{nh_{nM}}{h_{(n-1)M}} K\left(\frac{h_{nM}u}{h_{(n-1)M}}\right)|^3 f(x-h_{nM}u)du \ .$$

But by ii uniformly in u

$$\lim_{n\to\infty} \left|(n-1)K(u) - \frac{nh_{nM}}{h_{(n-1)M}} K\left(\frac{h_{nM}u}{h_{(n-1)M}}\right)\right| = \lim_{n\to\infty} \left|(n-1) - \frac{nh_{nM}}{h_{(n-1)M}}\right| K(u) \ .$$

Now $$\lim_{n\to\infty} \left|(n-1) - \frac{nh_{nM}}{h_{(n-1)M}}\right| = \lim_{n\to\infty} \left|-1+n[1 - \frac{h_{nM}}{h_{(n-1)M}}]\right| = |-1+\nu-1| = |\nu-2|.$$

Hence given $\delta > 0$, there is n_0 such that for $n > n_0$,

$$0 \le n^2h_{nM}^2 \ E|A_{n1}(x)|^3 \le \frac{|\nu-2|^3}{n}\int_{-\infty}^{\infty} K^3(u)f(x-h_{nM}u)du + \delta/n\int_{-\infty}^{\infty} f(x-h_{nM}u)du \ .$$

Since $\frac{1}{n}\int_{-\infty}^{\infty} f(x-h_{nM}u)du = \frac{1}{n} \to 0$ as $n \to \infty$, we have

$$n^2h_{nM}^2 \ E|A_{n1}(x)|^3 \to 0 \qquad \square$$

We may now apply the Normal Convergence Criterion (N.C.C.) found in Loéve, (1963, p. 316), to complete this section.

<u>Theorem 2.7</u>: If K and h_n satisfy the hypotheses of Lemma 2.6, then

$$(2.2)\qquad \lim_{n\to\infty} P\left[\frac{A_n(x) - EA_n(x)}{(\text{var } A_n(x))^{\frac{1}{2}}} \le c\right] = \int_{\infty}^{c} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du = \Phi(u).$$

Proof: By the N.C.C., a necessary and sufficient condition for (2.2) to hold is that for $\varepsilon > 0$,

$$(2.3)\qquad (n-1)M\ P\left[\left|\frac{A_{n1}(x) - EA_{n1}(x)}{[\text{var } A_{n1}(x)]^{\frac{1}{2}}}\right| \ge \varepsilon[(n-1)M]^{\frac{1}{2}}\right] \to 0.$$

A sufficient condition (Liapounov's condition) for 2.3 is that for some $\delta > 0$,

$$\frac{E\left|A_{n1}(x) - E[A_{n1}(x)]\right|^{2+\delta}}{(nM)^{\delta/2}\sigma^{2+\delta}[A_{n1}(x)]} \to 0 \quad \text{as } n \to \infty$$

where $\sigma^2[A_{n1}(x)] = \text{var}[A_{n1}(x)]$. We let $\delta = 1$. Then using the inequality $(a+b)^3 \le 4(a^3+b^3)$, we obtain

$$E|A_{n1}(x) - E[A_{n1}(x)]|^3 \le 8E|A_{n1}(x)|^3,$$

so that

$$\frac{E|A_{n1}(x) - E[A_{n1}(x)]|^3}{(nM)^{\frac{1}{2}}\sigma^3[A_{n1}(x)]} \le \frac{8E|A_{n1}(x)|^3}{(nM)^{\frac{1}{2}}\sigma^3[A_{n1}(x)]}.$$

By Lemmas 2.4 and 2.6, $E|A_{n1}(x)|^3 = 0(\frac{1}{n^2h_{nM}^2})$ and $\sigma^2[A_{n1}(x)] = 0(\frac{1}{nh_{nM}})$, so that

$$\frac{8E|A_{n1}(x)|^3}{(nM)^{\frac{1}{2}}\sigma^3[A_{n1}(x)]} = 0\left(\frac{1}{n^{\frac{1}{2}}}\frac{1}{(nh_{nM})^{\frac{1}{2}}}\right),$$

which completes the result. □

3. The Stopping Variable $N(\varepsilon,M)$: In this section, we shall generally suppress in the notation the explicit dependence of $N(\varepsilon,M)$ on ε and M. Hence we write $N(\varepsilon,M)$ simply as N. Noting that $[N \leq n] = [V_n(x) \leq \varepsilon]$, it is clear that the probabilistic structure of N is closely related to that of $V_n(x)$. Inasmuch as the structure of $B_n(x)$ depends on $f(x)$, we will be, in general, unable to give the exact asymptotic structure of N. In this section, we demonstrate the finiteness of the moments of N, the closure of N, and the divergence of N as $\varepsilon \to 0$.

Lemma 3.1: For arbitrary $t > 0$ and given $\varepsilon > 0$,

$$P[V_n(x) > \varepsilon] \leq e^{-nMh_{nM}\varepsilon t} E\, e^{S_n(x)t} \tag{3.1}$$

and

$$P[V_n(x) < -\varepsilon] \leq e^{-nMh_{nM}\varepsilon t} E\, e^{-S_n(x)t} \tag{3.2}$$

where

$$S_n(x) = \sum_{j=1}^{hM} K\left(\frac{x-X_j}{h_{nM}}\right) - \sum_{j=1}^{(n-1)M} \frac{n}{n-1} \frac{h_{nM}}{h_{(n-1)M}} K\left(\frac{x-X_j}{h_{(n-1)M}}\right) .$$

Proof: Define $T(x)$ to be the indicator of $[S_n(x) > nMh_{nM}\varepsilon]$. Then for arbitrary $t > 0$, $T(x) \leq e^{t(S_n(x)-nMh_{nM}\varepsilon)}$, so that

$$P[S_n(x) > nMh_{nM}\ ] \leq E\, e^{t(S_n(x)-nMh_{nM}\varepsilon)} .$$

Noting that $[V_n(x) > \varepsilon] = [S_n(x) > nMh_{nM}\varepsilon]$ completes the proof of (3.1). Equation (3.2) follows by similar arguments. □

Now,

$$P[N > nM] = P\left[\bigcap_{k=2}^{n} \{|V_k(x)| > \varepsilon\}\right]$$

$$\leq P[|V_n(x)| > \varepsilon] .$$

By Lemma 3.1, for arbitrary 3.1, for arbitrary $t > 0$,

$$P[N > nM] \le e^{-nMh_{nM} t} [E e^{S_n(x)t} + E e^{-S_n(x)t}] .$$

We next examine $E e^{S_n(x)t}$ and $E e^{-S_n(x)t}$. Let us decompose

$$S_n(x) = \sum_{k=1}^{(n-1)M} A^*_{nk}(x) + \sum_{k=(n-1)M+1}^{nM} B^*_{nk}(x) \quad \text{where}$$

$$A^*_{nk}(x) = K\left(\frac{x-X_k}{h_{nM}}\right) - \frac{n}{n-1} \frac{h_{nM}}{h_{(n-1)M}} K\left(\frac{x-X_k}{h_{(n-1)M}}\right) , \quad k=1,2,\ldots,(n-1)M$$

and

$$B^*_{nk}(x) = K\left(\frac{x-X_k}{h_{nM}}\right) \quad k = (n-1)M+1,\ldots,nM.$$

Notice that $A^*_{n1}(x),\ldots,A^*_{n,(n-1)M}(x)$ are i.i.d., that $B^*_{n,(n-1)M+1}(x),\ldots,$ $B^*_{n,nM}(x)$ are i.i.d. and that *all* of these random variables are mutually independent. Thus

$$E e^{S_n(x)} = \left[E e^{A^*_{n1}(x)}\right]^{(n-1)M} \left[E e^{B^*_{n,nM}(x)}\right]^{M} .$$

Buf if $L = \sup_x K(x)$,

$$E\left[e^{B_{n,nM}(x)}\right] = \int_{-\infty}^{\infty} e^{K(\frac{x-u}{h_{nM}})} f(u)du$$

$$\le e^{L} .$$

<u>Lemma 3.2</u>: Let $a_n = E[|A^*_{n1}(x)|e^{|A^*_{n1}(x)|}]$. If a_n satisfies

$\overline{\lim_{n \to \infty}} \frac{(n-1)}{\log n} a_n \le \gamma$ for some $\gamma \ge 0$, then

$$E\left[e^{\pm S_n(x)}\right] \le n^{M\gamma}e^{ML}$$

Proof: For n sufficiently large, since $a_n \to 0$, $a_n > \log(1+a_n)$. Combining this with the inequality on a_n, we have for n sufficiently large,

$$(n-1)M \log(1+a_n) \le M\gamma \log n.$$

Exponentiating both sides,

$$(1+a_n)^{(n-1)M} \le n^{M\gamma}.$$

Now

$$E\, e^{S_n(x)} \le [E\, e^{A^*_{n1}(x)}]^{(n-1)M}\, e^{ML} .$$

This, together with the observation that

$$e^{A^*_{n1}(x)} \le e^{|A^*_{n1}(x)|} = 1 + |A^*_{n1}(x)| + \frac{|A^*_{n1}(x)|^2}{2!} + \ldots$$

$$\le 1 + |A^*_{n1}(x)|e^{|A^*_{n1}(x)|} ,$$

so that $E\, e^{A^*_{n1}(x)} \le 1 + a_n$ completes the proof. □

Under the hypotheses of Lemmas 3.1 and 3.2 we have for arbitrary $t > 0$,

(3.3) $$P[N > nM] \le 2e^{ML}n^{M\gamma}e^{-nMh_{nM}\varepsilon t} .$$

Notice in general $nh_{nM} \to \infty$, so that $e^{-nMh_{nM}\varepsilon t} \to 0$. Since $n^{M\gamma} \to \infty$, however, we will usually want to choose h_n in such a way that for any $\delta > 0$ and for n sufficiently large

(3.4) $$e^{-nMh_{nM}\varepsilon} \le n^{-\delta} .$$

We note here that the usual choice $h_n = Bn^{-\alpha}$, $0 < \alpha < 1$ is sufficient to guarantee (3.4).

Theorem 3.3: Under the hypotheses of Lemmas 3.1 and 3.2 and assuming (3.4), we have $EN^r < \infty$ for every $r \geq 0$.

Proof:
$$EN^r = \sum_{n=0}^{\infty} n^r P[N = nM]$$
$$\leq \sum_{n=1}^{\infty} (n+1)^r P[N \geq nM] .$$

Using (3.3) with $t = 1$,

$$EN^r \leq 2e^{ML} \sum_{n=1}^{\infty} (n+1)^{r+M\gamma} e^{-(n+1)Mh (n+1)M^{\varepsilon}} .$$

Reindexing
$$EN^r \leq 2e^{ML} \sum_{n=2}^{\infty} n^{r+M\gamma} e^{-nMh\, nM^{\varepsilon}} .$$

Now for $\delta = r+M\gamma+2$, there is n_0 such that for $n \geq n_0$, (3.4) holds. Hence

$$EN^r \leq 2e^{ML}\left[\sum_{n=2}^{n_0-1} n^{r+M\gamma} e^{-nMh\, nM^{\varepsilon}} + \sum_{n=n_0}^{\infty} \frac{1}{n^2}\right] < \infty ,$$

which completes the proof. □

The definition of a_n in Lemma 3.2 involves the density, f, so is unsatisfactory from a statisticians point of view. Let us suppose $\limsup_{n\to\infty} \sup_x \frac{n}{\log n} |A^*_{n1}(x)| = c < \infty$. It is clear that $\lim_{n\to\infty} \sup_x |A^*_{n1}(x)| = 0$. Hence for n sufficiently large, $e^{|A^*_{n1}(x)|} < 2$ for every x. Clearly then the condition $\lim_{n\to\infty} \frac{(n-1)}{\log n} a_n \leq \gamma$ holds. The normal and double exponential kernels satisfy this latter sufficient condition. The uniform kernel does not, but it does satisfy the condition on a_n for every density, f.

Theorem 3.4: Under the hypothesis of Lemmas 3.1 and 3.2 and assuming (3.4), $P[N < \infty] = 1$, hence N is a closed stopping variable.

Proof: $P[N = \infty] = \lim_{n\to\infty} P[N \geq nM] = \lim_{n\to\infty} 2e^{ML}n^{M\gamma}e^{-nMh_{nM}\varepsilon} = 0.$ □

Now let us consider the behavior of N as a function of ε. Let $\Omega_n=[V_j(x) = 0$ for some $j\leq n]$ and let $\Omega_n^{\varepsilon} = [|V_j(x)| \leq \varepsilon$ for some $j\leq n]$. Since $\Omega_n^{\varepsilon} \to \Omega_n$ as $\varepsilon \to 0$, it follows that $P[N\leq nM] = P(\Omega_n^{\varepsilon})$ converges to $P(\Omega_n)$. If $P(\Omega_n) = 0$, then it follows that $N \to \infty$ in probability as $\varepsilon \to 0$. In general, $P(\Omega_n)$ may not be zero. Consider

$$K(u) = \begin{cases} \frac{1}{2} & |u| < 1 \\ 0 & |u| \geq 1 \,. \end{cases}$$

Then $V_n(x) = 0$ if $K\left(\frac{x-X_k}{h_{nM}}\right) = 0$ for $k = 1,2,\ldots,nM$. Thus

$$P\left[K\left(\frac{x-X_k}{h_{nM}}\right) = 0\right] = P\left[\left|\frac{x-X_k}{h_{nM}}\right| \geq 1\right] = 1-P\left[x-h_{nM}<X_k<x+h_{nM}\right]$$
$$= 1 - \int_{x-h_{nM}}^{x+h_{nM}} f(u)du \,.$$

This last quantity will in general be strictly positive, so that

$$P(\Omega_n) \geq P(V_n(x) = 0) \geq [1 - \int_{x-h_{nM}}^{x+h_{nM}} f(u)du]^{nM} > 0 \,.$$

The significant point of this example is that the uniform kernel may miss all the observations and hence both $\hat{f}_{nM}(x)$ and $\hat{f}_{(n-1)M}(x)$ could be 0. Clearly, were we to consider a normal or double exponential kernel, this could not happen. Let $\mathcal{K}$ be the class of kernels satisfying (1.2) and for which $P[V_n(x) = 0] = 0$ for any n.

Lemma 3.5: If $K(u)$ is a kernel satisfying (1.2) such that

(i) $K(u)$ is differentiable at all but possibly a finite number of values of u

and (ii) $K'(u)$ is continuous and non-zero at all but a finite number of values of u

then $V_n(x)$ has an absolutely continuous distribution so that $K \epsilon \mathcal{K}$.

Proof: The random variables $Y_k = \frac{1}{nMh_{nM}} K\left(\frac{x-X_k}{h_{nM}}\right)$, $k=1,2,\ldots,nM$ have a common absolutely continuous distribution (Parzen, 1960, p. 313). But then $\hat{f}_{nM}(x) = Y_1+\ldots+Y_{nM}$ has a density given by the n-fold convolution of the density of Y_1. It follows that $V_n(x) = \hat{f}_{nM}(x) - \hat{f}_{(n-1)M}(x)$ has a density (Parzen, 1960, p. 318). □

We note the normal, double exponential and Cauchy kernels all satisfy i and ii, so that $\mathcal{K}$ is not empty.

Theorem 3.6: Let $K \epsilon \mathcal{K}$ and let h_n satisfy (1.3). Then $N \to \infty$ in probability and with probability one as $\varepsilon \to 0$.

Proof: The divergence in probability follows from previous remarks. Now since $P[V_n(x) = 0$ for some $n] = 0$, $P[|V_n(x)| \to 0$ as $n \to 0$, $|V_n(x)| > 0$ for every $n] = 1$. Let $\omega \epsilon [|V_n(x)| \to 0$ as $n \to \infty$, $|V_n(x)| > 0$ for every $n]$. Let n_0 be any finite number and assume $N < n_0$ for all ε. Then for every $\varepsilon > 0$, $|V_j(x)| < \varepsilon$ for at least one $j \le n_0$. But since $|V_j(x)| > 0$ for all $j \le n_0$, let $\varepsilon^* = \frac{1}{2} \min_{1 \le j \le n_0} |V_j(x)|$ so that for ε^*, $|V_j(x)| > \varepsilon^*$ for all $j \le n_0$. This is a contradition to $N < n_0$ for all ε. Hence $N \ge n_0$ for ε sufficiently small. That is to say N is greater than any finite number for ε sufficiently small and for $\omega \epsilon [|V_n(x)| \to 0$ as $n \to \infty$, $|V_n(x)| > 0$ for every $n]$. Thus $P[N \to \infty$ as $\varepsilon \to 0] = 1$. □

We are now able to state a convergence theorem based on Theorem 3.6.

Theorem 3.7: Suppose $N \to \infty$ as $\varepsilon \to 0$ with probability one and $\hat{f}_n(x) \to f(x)$ as $n \to \infty$ with probability one, then $\hat{f}_N(x) \to f(x)$ as $\varepsilon \to 0$ with probability one.

Proof: Let A be the set of probability one for which $N \to \infty$. Let B be the set of probability one for which $\hat{f}_n(x) \to f(x)$. Clearly on $A \cap B$, $\hat{f}_N(x) \to f(x)$ and $P(A \cap B) = 1$.

Sufficient conditions for $N \to \infty$ appear in Theorem 3.6. Sufficient conditions for $\hat{f}_n(x) \to f(x)$ appear in Theorem 2.1. □

We close this section by noting that a slightly revised stopping rule N' given by

$$N'(\varepsilon,M) = \begin{cases} \text{1st } n \text{ such that } |V_n(x)| < \varepsilon \text{ but } |V_n(x)| > 0 \\ \infty \text{ if no such } n \text{ exists} \end{cases}$$

obviates the need to consider the class $\mathcal{K}$. For $K \in \mathcal{K}$, $N = N'$ a.s. The result of this section holds for N' as well as N except that the reference $K \in \mathcal{K}$ may be removed from Theorem 3.6 . Modifications needed in the proofs are obvious and left to the reader.

4. Concluding Remarks:

The problems associated with the choice of K and h_n are well-known and appreciated by users and theoreticians alike. We shall not comment except to say these problems remain in the sequential case. To these we have added those associated with the choice of ε and M. Some clue to the choice of ε is given by the following easily-proved observation:

$$(4.1)\qquad E|V_n(x)| \le \{E|\hat{f}_{nM}(x) - f(x)|^2\}^{\frac{1}{2}} + \{E|\hat{f}_{(n-1)M}(x) - f(x)|^2\}^{\frac{1}{2}} .$$

In general, we would like to choose ε so that the mean square error meets some prespecified error level, say δ. For heuristic purposes, let us suppose that M is sufficiently small and n sufficiently large so that

$$E[\hat{f}_{nM}(x) - f(x)]^2 \approx \delta \approx E[\hat{f}_{(n-1)M}(x) - f(x)]^2 .$$

Then according to (4.1), $E|V_n(x)| \lesssim 2\delta^{\frac{1}{2}}$, suggesting that $\varepsilon \lesssim 2\delta^{\frac{1}{2}}$ is a suitable choice of ε to meet error δ. Actually, this appears to be quite a conservative way of choosing ε for $\delta < 1$.

If there is no penalty for sampling items one-at-a-time rather than in blocks of M, it is clear that $M = 1$ is the best choice. If M is too large, $|V_N(x)|$ will be substantially less than ε, and hence too many items will be sampled. The optimal M must be determined by weighing the costs of sampling one-at-a-time against the cost of taking an unnecessarily large sample.

A really satisfying theory for choice of ε and M is yet to be devised. In the meantime, $M = 1$ and $\varepsilon = 2\delta^{\frac{1}{2}}$ appear to be adequate.

6. References:

1 Loéve, M. (1963), *Probability Theory*, Van Nostrand, New York.

2 Nadaraya, E.A. (1965), "On nonparametric estimates of density functions and regression curves," *Theory Prob. Appl.*, *10*, 186-190.

3 Parzen, E. (1960), *Modern Probability Theory and Its Applications*, John Wiley and Sons, New York.

4 Parzen, E. (1962), "On the estimation of a probability density function and the mode," *Ann. Math. Statist.*, *33*, 1065-1076.

5 Rosenblatt, M. (1956), "Remarks on some nonparametric estimates of a density function," *Ann. Math. Statist.*, *27*, 832-837.

6 Van Ryzin, J. (1969), "On strong consistency of density estimates," *Ann. Math. Statist.*, *40*, 1765-1772.