# ON SEQUENTIAL DENSITY ESTIMATION

by

Raymond J. Carroll

*Department of Statistics*

*University of North Carolina at Chapel Hill*

Institute of Statistics Mimeo Series #1025

August, 1975

## SUMMARY

We consider the problem of sequential estimation of a density function $f$ at a point $x_0$ which may be known or unknown. Let $\{T_n\}$ be a sequence of estimators of $x_0$. For two classes of density estimators $\hat{f}_n$, namely the kernel estimates and a recursive modification of these, we show that if $N(d)$ is a sequence of integer-valued random variables and $n(d)$ a sequence of constants with $N(d)/n(d) \to 1$ in probability as $d \to 0$, then $\hat{f}_{N(d)}\left(T_{N(d)}\right) - f(x_0)$ is asymptotically normally distributed (when properly normed). We also propose two new classes of stopping rules based on the ideas of fixed-width interval estimation and show that for these rules, $N(d)/n(d) \to 1$ almost surely and $EN(d)/n(d) \to 1$ as $d \to 0$. One of the stopping rules is itself asymptotically normally distributed when properly normed and yields a confidence interval for $f(x_0)$ of fixed-width and prescribed coverage probability.

## Introduction

While there have been many papers on the estimation of a probability density function $f(x)$ (see Wegman (1972)), the literature on sequential density estimation is relatively small. Srivastava (1973) notes that one often takes as many observations as possible in a certain time period, so that the number of observations is random. Davies and Wegman (1975) were interested in developing sequential rules which satisfy a certain error control. In this paper we provide a treatment of the asymptotic distributions of two types of density estimators when the number of observations is random; the estimators considered are the kernel estimators (Rosenblatt (1956), Parzen (1962)) and a variant of these due to Yamato (1971). We then develop two new classes of sequential stopping rules for estimating $f(x)$ and obtain their precise asymptotic behavior; one of the stopping rules actually yields a confidence interval for $f(x)$ of fixed-width and prescribed coverage probability.

Specifically, we focus our attention on the following problem. Suppose we are interested in estimating $f(x_0)$, where $x_0$ may be known or unknown; an example of the latter is the case where $x_0$ is the population median or mode. Typically, there will be a sequence of estimators of $x_0$, say $\{T_n\}$. Suppose $N(d)$ is a sequence of integer-valued random variables (i.e., stopping rules) and $n(d)$ is a sequence of constants for which $N(d)/n(d) \to 1$ in probability as $d \to 0$. It is known that in many cases, if the density estimators are denoted by $\hat{f}_n(x)$, that for some sequence $\{\varepsilon_n\}$ decreasing to zero, as $n \to \infty$

(1.1a) $$(n\varepsilon_n)^{\frac{1}{2}}\left(\hat{f}_n(x_0)-E\hat{f}_n(x_0)\right)$$

(1.1b) $$(n\varepsilon_n)^{\frac{1}{2}}\left(\hat{f}_n(x_0)-f(x_0)\right)$$

converge in distribution to a normal random variable. The first two sections of this paper are concerned with finding conditions under which both

$$(1.2a)\qquad \left(N(d)\varepsilon_{N(d)}\right)^{\frac{1}{2}}\left(\hat{f}_{N(d)}(T_{N(d)})-E\hat{f}_{N(d)}(T_{N(d)})\right)$$

$$(1.2b)\qquad \left(N(d)\varepsilon_{N(d)}\right)^{\frac{1}{2}}\left(\hat{f}_{N(d)}(T_{N(d)})-f(x_0)\right)$$

still converge in distribution to a normal random variable. Srivastava (1973) attempts to show for the kernel estimators that

$$(1.3)\qquad \left(N(d)\varepsilon_{N(d)}\right)^{\frac{1}{2}}\left(\hat{f}_{N(d)}(x_0)-f(x_0)\right)$$

is asymptotically normally distributed, but the proof of his Theorem 4.1 contains a number of errors; for example, the statement after equation (4.9) is not correct (take the kernel to be uniform on the interval $[-\frac{1}{2},\frac{1}{2}]$ ). Wegman and Davies (1975) have shown the asymptotic normality of (1.3) for the Yamato (1971) recursive estimators indirectly by use of an almost sure invariance principle.

Hence, finding the asymptotic distributions of (1.2a) and (1.2b) are still unsolved problems, while that of (1.3) is unresolved for the kernel estimators. In Sections 2 and 3 we obtain the necessary results by means of the theory of weak convergence (Billingsley (1968)) and a random change of time argument. Incidentally, the approach yields the asymptotic normality of (1.3) for both types of estimators in a general and reasonably straightforward manner, although it is much harder to obtain asymptotic normality for (1.2a) and (1.2b).

In the final two sections of the paper, we propose two new classes of stopping rules $N(d)$, both based on the ideas of fixed-width interval estimation. Davies and Wegman (1975) have proposed one class of stopping rules, shown that they stop with probability one, and investigated the existence of moments, but the exact large sample behavior of their rules is unknown. For both classes we propose, we find sequences of constants $n(d)$ for which

(1.4) $$N(d)/n(d) \to 1 \quad \text{almost surely as} \quad d \to 0$$

(1.5) $$EN(d)/n(d) \to 1 \quad \text{as} \quad d \to 0 .$$

This is precise information about the stopping rules and gives the user an idea of the approximate number of observations to be taken. Interestingly enough, one of the stopping rules yields a confidence interval for $f(x_0)$ of fixed-width and prescribed coverage probability, and we have been able to show the asymptotic normality of this stopping rule itself.

## Kernel Estimators

In this section we investigate the asymptotic normality under random sample sizes of estimates of the density using the kernel estimates due to Rosenblatt (1956) and Parzen (1962). The kernel estimator of the density $f(x)$ is given by

$$f_n(x) = (n\varepsilon_n)^{-1} \sum_1^n K\left(\varepsilon_n^{-1}(x-X_i)\right) ,$$

where the kernel $K$ is a bounded density function and $\{\varepsilon_n\}$ is a sequence of constants decreasing to zero. We wish to estimate $f(x_0)$, where $x_0$ is an unknown point; we will thus assume the existence of a sequence of estimators $\{T_n\}$ of $x_0$. For example, if $x_0$ were the population median, $T_n$ would be taken as the sample median.

Although the details are fairly complicated, our method is contained in a number of basic steps. Initially, we consider the asymptotic normality of $f_n(x_0)$ under random sample sizes by giving a weak convergence argument applied to a process closely related to $f_{[nt]}(x_0)$ (where $[\cdot]$ is the greatest integer function) and then using a random change of time argument (Billingsley (1968)). Then, $f_n(x_0)$ and $f_n(T_n)$ are shown to be sufficiently close by modifying the maximal deviation argument of Woodroofe (1967).

The two basic processes with which we will work in this section are given below.

<u>DEFINITION 2.1.</u> Let $0<\alpha<1$ be fixed, let $\{a_n\}$ converge to zero, and define for $0\leq s,t\geq 1$,

$$V_n(s) = (n\varepsilon_n)^{-\frac{1}{2}} \sum_{1}^{[ns]} \left\{K\left(\varepsilon_{[ns]}^{-1}(x_0-X_i)\right) - EK\left(\varepsilon_{[ns]}^{-1}(x_0-X)\right)\right\} \quad \text{if} \quad s\geq[n\alpha]/n$$

$$= (n\varepsilon_n)^{-\frac{1}{2}} \sum_{1}^{[ns]} \left\{K\left(\varepsilon_{[n\alpha]}^{-1}(x_0-X_i)\right) - EK\left(\varepsilon_{[n\alpha]}^{-1}(x_0-X)\right)\right\} \quad \text{if} \quad s\leq[n\alpha]/n .$$

$$V_n^*(s,t) = (n\varepsilon_n)^{-\frac{1}{2}} \sum_{1}^{[ns]} \left\{K\left(\varepsilon_{[ns]}^{-1}(x_0+ta_n-X_i)\right) - EK\left(\varepsilon_{[ns]}^{-1}(x_0+ta_n-X)\right)\right\} \quad \text{if} \quad s\geq[n\alpha]/n$$

$$= (n\varepsilon_n)^{-\frac{1}{2}} \sum_{1}^{[ns]} \left\{K\left(\varepsilon_{[n\alpha]}^{-1}(x_0+ta_n-X_i)\right) - EK\left(\varepsilon_{[n\alpha]}^{-1}(x_0+ta_n-X)\right)\right\} \quad \text{if} \quad s\leq[n\alpha]/n .$$

What we eventually will assume is that $T_n$ converges to $x_0$ faster than $a_n$ converges to zero, so that we can replace the parameter $t$ in $V_n^*$ by $a_n^{-1}(T_n-x_0)$. It is clear that $V_n$ is a random element of $D[0,1]$, while if $K$ is right continuous, with limits from the left, $V_n^*$ is a random element of $D_2$ (see Billingsley (1968) and Bickel and Wichura (1971) respectively for definitions). The idea is now reasonably clear; we will first consider the weak convergence of the process $V_n$, then show that $V_n$ and $V_n^*$ are "close" in probability. Then, since the normed random sample size version of the estimate of $f(x_0)$ can be basically obtained from $V^*_{n(d)}\left(N(d)/n(d),\ a^{-1}_{n(d)}(T_{N(d)}-x_0)\right)$, a random change of time argument and a few computations will yield the final result.

It will be assumed throughout the rest of this paper that for some function $h$, $\varepsilon_{[ns]}/\varepsilon_n \to h(s)$ as $n\to\infty$, and $A(f,K)$ will be defined by

$$A(f,K) = f(x_0)\int K^2(y)dy\ . \tag{2.1}$$

The proofs of all results will be delayed to the end of the section.

LEMMA 2.1. Suppose that for any $c_0>0$ there is a constant $M(c_0)$ for which

$$E\{K(\varepsilon^{-1}_{[ns_1]}(x_0-X)) - K(\varepsilon^{-1}_{[ns_2]}(x_0-X))\}^2 \le \varepsilon_n M(c_0)|s_1-s_2| \quad \text{if} \quad s_1,s_2\ge c_0\ . \tag{2.2}$$

Then, the sequence $\{V_n\}$ is tight and there exists a process $V$ for which

(2.3a) The finite dimensional distributions of $V_n$ converge to those of $V$.

(2.3b) For $0\le s\le 1$, $V(s)$ is normally distributed with mean zero and variance

$$\begin{array}{ll} sh(s)A(F,K) & \text{if} \quad s\ge\alpha \\ sh(\alpha)A(F,K) & \text{if} \quad s\le\alpha\ . \end{array}$$

REMARK 2.1. Assumption (2.2) is satisfied in many cases. Suppose that $\varepsilon_n=n^{-\beta}$ for some $0<\beta<\frac{1}{2}$. Then, (2.2) will hold if $K$ vanishes off some interval $[a,b]$ and is Lipschitz on this interval. It will also hold if $f(x)$ is bounded, $K$ is continuously differentiable, and if for any sequence of constants $\{a_n\}$ converging to zero,

$$\sup_n \int y^2[K'(y(1+a_n))]^2\, dy < \infty\ .$$

LEMMA 2.2. Suppose that

(2.4a) $a_n \varepsilon_n^{-1} = O(a_n^{\beta})$ for some $0<\beta<1$, where $O$ is the standard "big oh".

(2.4b) $n\varepsilon_n$ increases in $n$

(2.4c) $n a_n^{-2+\beta} \exp\{-c a_n^{-\beta/2}\} \to 0$ for all $c>0$

(2.4d) $K$ is Lipschitz of order one and satisfies Woodroofe's (1967) conditions.

Then

$$\sup\{|V_n(s)-V_n^*(s,t)|: \ 0\le s,t\le 1\} \xrightarrow{p} 0 ,$$

where $\xrightarrow{p}$ is convergence in probability.

REMARK 2.2. Lemma 2.2 is the key result, since as will be seen it says that estimation of $x_0$ by $T_n$ does not change the asymptotic distribution.

THEOREM 2.1. Suppose that

(2.5a) $T_n = x_0 + o(a_n^{-1})$ almost surely as $n\to\infty$, where "$o$" is the "little oh" notation.

(2.5b) $\{a_n\}$, $\{\varepsilon_n\}$, $K$ satisfy the conditions of Lemmas 2.1 and 2.2.

(2.5c) For some sequence of constants $n(d)\to\infty$, the integer-valued random variables $N(d)$ satisfy $N(d)/n(d) \to 1$ in probability as $d\to 0$.

Then

(2.6) $\left(N(d)\varepsilon_{N(d)}\right) \left[f_{N(d)}(T_{N(d)}) - \varepsilon_{N(d)}^{-1} \int K\left(\varepsilon_{N(d)}^{-1}(T_{N(d)}-y)\right) f(y)\,dy\right]$

converges in distribution to a normal random variable with mean zero and variance $A(f,K)$.

REMARK 2.3. The conditions (2.5a) and (2.5b) are satisfied in many cases. For example, if $x_0$ is the $p^{th}$ population quantile and $T_n$ is the $p^{th}$ sample quantile, Bahadur (1966) has shown that $T_n - x_0 = O\left(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}\right)$. If $x_0$ is the population mode, Venter (1967) and Sager (1975) have given rates of convergence of $T_n$ to $x_0$. Note that if one wanted merely to estimate $f(x_0)$ for a known $x_0$, (2.5a) and (2.5b) clearly hold by choosing $T_n \equiv x_0$. Thus Theorem 2.1 is a generalization of the problem considered by Srivastava (1973).

While Theorem 2.1 shows the asymptotic normality of a normed version of $f_{N(d)}(T_{N(d)})$, it is useful to ask when the integral appearing in (2.6) can be replaced by $x_0$. This is the gist of the following Corollary. The results here are comparable to those given by Cacoullos (1966).

COROLLARY 2.1. Suppose that the conditions of Theorem 2.1 hold and that on the support of $K$ the density $f$ is twice boundedly continuously differentiable. Suppose further that

(2.7) $$\int yK(y)dy = 0 \;, \quad \int y^2K(y)dy < \infty$$

(2.8) $$(n\varepsilon_n)^{\frac{1}{2}}a_n \to 0 \;, \quad n\varepsilon_n^5 \to 0 \;.$$

Then

(2.9) $$\left(N(d)\varepsilon_{N(d)}\right)^{\frac{1}{2}}\left(f_{N(d)}(T_{N(d)} - f(x_0)\right)$$

is asymptotically normally distributed with mean zero and variance $A(f,K)$.

PROOF OF LEMMA 2.1. By using the Cramer-Wold device and the method given in Parzen (1962, page 1069), we see that the finite dimensional distributions of $V_n$ converge, so it suffices to show that the sequence $\{V_n\}$ is tight. From, for example, the extension of Theorem 3 of Bickel and Wichura (1971) given in that paper, it suffices to show that there exist $\beta>\frac{1}{2}$, $M>0$ such that if $s_1=j/n$, $s_2=k/n$, then

$$E|V_n(s_1)-V_n(s_2)|^2 \le M|s_2-s_1|^\beta .$$

If $s_1<s_2\le[n\alpha]/n$, this is clear. If $s_2>s_1\ge[n\alpha]/n$, we obtain

$$\begin{aligned} E|V_n(s_1)-V_n(s_2)|^2 &\le Ms_1\varepsilon_n^{-1} \operatorname{Var}\{K\big(\varepsilon_{[ns_1]}^{-1}(x_0-X)\big) - K\big(\varepsilon_{[ns_2]}^{-1}(x_0-X)\big)\}^2 \\ &\quad + M(s_2-s_1)\varepsilon_n^{-1} \operatorname{Var}\{K\big(\varepsilon_{[ns_2]}^{-1}(x_0-X)\big)\} \\ &\le M^*|s_2-s_1| \end{aligned}$$

by assumption (2.2). The final case $(s_1\le[n\alpha]/n\le s_2)$ follows in a similar manner, so that $\beta=1$ suffices.

PROOF OF LEMMA 2.2. Here we make use of the results of Woodroofe (1967). First define

$$Z_n(s) = ([ns]\varepsilon_{[ng(s)]})^{-\frac{1}{2}} (n\varepsilon_n)^{\frac{1}{2}} V_n(s)/\sigma_{[ng(s)]}(x_0)$$

$$Z_n^*(s,t) = ([ns]\varepsilon_{[ng(s)]})^{-\frac{1}{2}} (n\varepsilon_n)^{\frac{1}{2}} V_n^*(s,t)/\sigma_{[ng(s)]}(x_0+ta_n) ,$$

where

$$\begin{aligned} g(s) = g_n(s) &= s && \text{if } s\ge[n\alpha]/n \\ &= [n\alpha]/n && \text{if } s<[n\alpha]/n \end{aligned}$$

$$\sigma_n^2(x) = \varepsilon_n^{-1} \operatorname{Var}\left(K\left(\varepsilon_n^{-1}(x-X_1)\right)\right) .$$

If we show that

$$\sup\{|Z_n(s)-Z_n^*(s,t)|: 0\le s,t\le 1\} = \sup\{|Z_n(i/n)-Z_n^*(i/n,t)|: 0\le t\le 1,\ 0\le i\le n\} \xrightarrow{p} 0 ,$$

this will yield the result because of Lemma 2.1 and since

$$\sup\{|\sigma_{[ng(s)]}(x_0+ta_n)-\sigma_{[ng(s)]}(x_0)|: 0\le s,t\le 1\} \to 0 .$$

Now, fix $i$ and let, for $p=0,1,\dots,[\varepsilon_i^{-1}]$ ,

$$x_{np} = x_0+p\varepsilon_i$$

$$\sigma_{np}(x) = \sigma_i(x_{np}+x\varepsilon_i)/\sigma_i(x_{np}) \qquad 0\le x\le 1$$

$$Z_{np}(x) = \sigma_{np}(x)Z_n^*(i/n,\ p\varepsilon_i a_n^{-1}+x\varepsilon_i a_n^{-1}) \quad 0\le x\le 1$$

Next, define $Z_{np}^{(k)}(x) = Z_{np}(j2^{-k})$ if $x=j2^{-k}$, $j=0,1,\dots,2^k$, with $Z_{np}^{(k)}(x)$ defined by linear interpolation otherwise. Choose $k$ such that $s^{k/2}\sim a_n^{-1}$. From Lemma 3.2 of Woodroofe (1967), there exists $D_1d$, $n_1$ such that for $i\ge n_1$,

$$\Pr\{\sup_{0\le x\le 1} |Z_{no}(x)-Z_{no}^{(k)}(x)|>\varepsilon\} \le D\exp\{-\varepsilon d a_n^{-1}\} ,$$

so that as $n_1,n\to\infty$,

$$\sup\{|Z_{no}(x)-Z_{no}^{(k)}(x)|: 0\le x\le 1,\ n_1\le i\le n\} \xrightarrow{p} 0 .$$

Now, since $a_n\varepsilon_n^{-1} = O(a_n^{\beta})$ and since $Z_{no}^{(k)}(x)$ is piecewise linear for $0\le x\le 1$, we have

$$\sup\{|Z_{no}(o)-Z_{no}^{(k)}(ta_n\varepsilon_n^{-1})|: 0\le t\le 1\} \le \sup\{|Z_{no}(o)-Z_{no}^{(k)}(j2^{-k})|: j\le M2^{k(1-\frac{\beta}{2})}\} ,$$

because $a_n\varepsilon_i^{-1} \le a_n\varepsilon_n^{-1} \le Ma_n^{\beta} \sim M2^{-k\beta/2}$. Hence, by Lemma 3.1 of Woodroofe (1967),

$$\Pr\{\sup_{0\le t\le 1} |Z_{no}(o) - Z_{no}^{(k)}(ta_n\varepsilon_i^{-1})| > \varepsilon\} \le M^* a_n^{-2(1-\frac{\beta}{2})} \exp\{-\varepsilon M 2^{k\beta/4}\}$$
$$\sim M^* a_n^{-2(1-\frac{\beta}{2})} \exp\{-\varepsilon M a_n^{-\beta/2}\} .$$

Thus, as $n_1, n \to \infty$,

$$\sup\{|Z_{no}(x) - Z_{no}(o)| : 0 \le x \le a_n\varepsilon_i^{-1},\ n_1 \le i \le n\} \xrightarrow{p} 0 .$$

Since $Z_{no}(o) = Z_n(i/n)$ and $Z_{no}(ta_n\varepsilon_i^{-1}) = Z_n^*(i/n,t)\sigma_{no}(ta_n\varepsilon_i^{-1})$, we have that

$$\sup\{|Z_n(i/n) - Z_n^*(i/n,t)| : 0 \le t \le 1,\ n_1 \le i \le n\} \xrightarrow{p} 0 ,$$

so that

$$\sup\{|V_n(i/n) - V_n^*(i/n,t)| : 0 \le t \le 1,\ n_1 \le i \le n\} \xrightarrow{p} 0 .$$

Since $K$ is Lipschitz, it is continuous so that the result now follows.

<u>PROOF OF THEOREM 2.1.</u> Define $m(d) = 2n(d)$ and $W_n(s) = V_n(s) - V_n(\frac{1}{2})$. Then $W_{m(d)}$ is tight and

$$\sup_{\frac{1}{2}\le s\le \frac{1}{2}+\eta} |W_{M(d)}(s)|$$
$$\le |W_{m(d)}(\tfrac{1}{2}+\eta)| + \sup_{\frac{1}{2}\le s\le\frac{1}{2}+\eta} \min\{|W_{m(d)}(s) - W_{m(d)}(\tfrac{1}{2})|,\ |W_{m(d)}(s) - W_{m(d)}(\tfrac{1}{2}+\eta)|\}.$$

Since the first term on the right hand side of the above equation converges in probability to zero as $d, \eta \to 0$ (because of (2.2) and Chebychev's inequality) and the second term is bounded by the modulus of continuity, we have (if "$\xrightarrow{L}$" denotes convergence in distribution)

$$V_{M(d)}\left(N(d)/m(d)\right) \xrightarrow{L} V(\tfrac{1}{2}) .$$

Because of Lemma 2.2 and assumption (2.5a), we thus obtain

$$V^*_{m(d)}\left(N(d)/m(d),\ a^{-1}_{m(d)}(T_{N(d)}-x_0)\right) \xrightarrow{L} V(\tfrac{1}{2}) . \tag{2.10}$$

But, by choosing $\alpha<\frac{1}{2}$ in Definition 1.1, we obtain

$$\left(m(d)\varepsilon_{m(d)}\right)^{-1}\left(N(d)\varepsilon_{N(d)}\right)\left[f_{N(d)}(T_{N(d)})-\varepsilon^{-1}_{N(d)}\int K\left(\varepsilon^{-1}_{N(d)}(T_{N(d)}-y)\right)f(y)\,dy\right] \xrightarrow{L} V(\tfrac{1}{2}) , \tag{2.11}$$

with $\left(N(d)\varepsilon_{N(d)}\right)^{\frac{1}{2}}\left(m(d)\varepsilon_{m(d)}\right)^{-\frac{1}{2}} \xrightarrow{P} \frac{1}{2}h(\frac{1}{2})$ . Since $V(\frac{1}{2})$ has a normal distribution with mean zero and variance $A(f,K)\left(\frac{1}{2}h(\frac{1}{2})\right)$, the proof is complete.

PROOF OF COROLLARY 2.1. Because of Theorem 2.1, it suffices to show that

$$\sup_{|t|\le 1} (n/\varepsilon_n)^{\frac{1}{2}} \left| \int K\left(\varepsilon_n^{-1}(x_0+ta_n-y)\right)f(y)\,dy - \varepsilon_n f(x_0)\right| \to 0 .$$

Now, since $K$ is a density function, this term is bounded above by

$$\sup_{|t|\le 1} (n\varepsilon_n)^{\frac{1}{2}} \int K(y)\ |f(x_0-\varepsilon_n y-ta_n)|-f(x_0)\,dy .$$

A Taylor's expansion now completes the proof.

## A Second Class of Estimators

In this section we investigate the asymptotic normality under random sample sizes of the recursive density estimators introduced by Yamato (1971) and defined by

$$\begin{aligned} f^*_n(x) &= n^{-1}\sum_{i=1}^{n} \varepsilon_i^{-1}K\left(\varepsilon_i^{-1}(x-X_i)\right) \\ &= \left(\frac{n-1}{n}\right)f^*_{n-1}(x) + (n\varepsilon_n)^{-1}K\left(\varepsilon_n^{-1}(x-X_n)\right) . \end{aligned}$$

The recursive property of this class of density estimators is clearly useful in sequential investigations and also for fairly large sample sizes since addition of a few extra observations means the kernel estimate $f_n(x)$ must be entirely recomputed. Wegman and Davies (1975) have recently shown that for a given $x_0$, $f_n^*(x_0)$ satisfies an almost sure invariance principle; this method is closely related to that of Jain, Jogdeo and Stout (1975). We are still left with the problem of estimating $f(x_0)$ where $x_0$ is unknown. The outline of our results is very much like that of Section 2, but the methods are different (especially the analogue to Lemma 2.2, where we use a weak convergence argument) and we are able to relax the Lipschitz condition on $K$. One can obtain an analogue of Lemma 2.1 using the results of Wegman and Davies (1975); however, their assumptions and methods are different from those used here.

We again start with two processes.

DEFINITION 3.1. Define

$$V_n = V_n(s) = (n/\varepsilon_n)^{-\frac{1}{2}} \sum_{i=1}^{[ns]} \{K(\varepsilon_i^{-1}(x_o-X_i)) - EK(\varepsilon_i^{-1}(x_0-X))\}/\varepsilon_i$$

$$V_n^* = V_n(s,t) = (n/\varepsilon_n)^{-\frac{1}{2}} \sum_{i=1}^{[ns]} \{K(\varepsilon_i^{-1}(x_0+ta_n-X_i)) - EK(\varepsilon_i^{-1}(x_0+ta_n-X))\}/\varepsilon_i .$$

LEMMA 3.1. Suppose that

$$\lim_n \sup \int K^2(y)f(x_0-y\varepsilon_n)dy < \infty$$

$$\int K^2(y)dy < \infty ,$$

that $K,f$ satisfy the conditions of Theorem 5 of Yamato (1971), and that there exists a number $\alpha$ $(0\le\alpha\le1)$ for which

$$n^{-1}\sum_{i=1}^{n}\varepsilon_n/\varepsilon_i \to \alpha .$$

Then $V_n$ is tight and there is a process $V$ in $D[0,1]$ for which the finite dimensional distributions of $V_n$ converge to those of $V$ and $V(s)$ is normally distributed with mean zero and variance $\alpha sh(s)A(f,K)$.

The following two results form an analogue to Lemma 2.2. Note that while the Lipschitz condition on $K$ is somewhat reduced in Lemma 3.3, the price being paid is a stronger relationship between the sequences $\{a_n\}$ and $\{\varepsilon_n\}$.

<u>LEMMA 3.2.</u> Suppose one of the following hold:

(3.1a) $K$ is Lipschitz of order one and $a_n^2/\varepsilon_n^3 \to 0$.

(3.1b) $K$ is continuously differentiable, $a_n^4/\varepsilon_n^5 \to 0$, and for any sequence $\eta_n\to0$,

$$\lim\sup\int \{K'\big(y(1+\eta_n)\big)\}^{2r} f(x_0-y\varepsilon_n)dy < \infty \quad (r=1,2).$$

Then,

$$\sup\{|V_n(s)-V_n^*(s,t)|: 0\le s,t\le1\} \xrightarrow{p} 0 .$$

<u>LEMMA 3.3.</u> Suppose $K$ vanishes off a closed interval $[a,b]$ and is Lipschitz on $[a,b]$. Suppose further that

(3.2) $a_n/\varepsilon_n\to0$, $a_n/n\varepsilon_n^2 = O(n^{-\beta})$ for some $\beta>0$, and $\varepsilon_n = O(n^{-\gamma})$ for some $\gamma>0$.

Then,

$$\sup\{|V_n(s)-V_n^*(s,t)|:\ 0\le s,t\le 1\} \xrightarrow{P} 0 .$$

Theorem 3.1 and Corollary 3.1 (which are given below) follow in a manner similar to Theorem 2.1 and Corollary 2.1.

THEOREM 3.1. Assume (2.5a), (2.5c), the conditions of Lemma 3.1, and that the conditions of either Lemma 3.2 or Lemma 3.3 hold. Then

$$(3.3)\quad \left(N(d)\varepsilon_{N(d)}\right)^{\frac{1}{2}}\left[f^*_{N(d)}(T_{N(d)}) - \frac{1}{N(d)}\sum_{i=1}^{N(d)} \varepsilon_i^{-1}\int K\left(\varepsilon_i^{-1}(T_{N(d)}-y)\right)f(y)\,dy\right]$$

converges in distribution to a normal random variable with mean zero and variance $\alpha A(f,K)$.

COROLLARY 3.1. Under the conditions of Theorem 3.1 and Corollary 2.1,

$$\left(N(d)\varepsilon_{N(d)}\right)^{\frac{1}{2}}\left(f_{N(d)}(T_{N(d)})-f(x_0)\right)$$

is asymptotically normally distributed with mean zero and variance $\alpha A(f,K)$.

PROOF OF LEMMA 3.1. As in Lemma 2.1, it is straightforward to verify the convergence of the finite dimensional distributions of $V_n$. Thus, it is again sufficient to show the existence of $M>0$, $\beta>\frac{1}{2}$ such that for $s_1=j/n$, $s_2=k/n$, $j\le k$,

$$(3.4)\qquad E|V_n(s_1)-V_n(s_2)|^2 \le M|s_1-s_2|^\beta .$$

Now, the left hand side of (3.4) is bounded by

$$(\varepsilon_n/n) \sum_{i=[ns_1]}^{(ns_2]} \varepsilon_i^{-2} \left\{ \int K^2\left(\varepsilon_i^{-1}(x_0-y)\right) f(y)dy - \left( \int K\left(\varepsilon_i^{-1}(x_0-y)\right) f(y)dy\right)^2 \right\}$$

$$= (\varepsilon_n/n) \sum_{i=[ns_1]}^{[ns_2]} \left\{ \varepsilon_i^{-1} \int K^2(y) f(x_0-y\varepsilon_i)dy - \left( \int K(y) f(x_0-y\varepsilon_i)dy\right)^2 \right\} ,$$

which completes the proof with $\beta=1$.

PROOF OF LEMMA 3.2. We will use a weak convergence argument, so it is first necessary to verify the convergence of the finite dimensional distributions. Fix $x$ and $t$. Then

$$E|V_n(s)-V_n^*(s,t)|^2 \le (\varepsilon_n/n) \sum_{i=1}^{n} \varepsilon_i^{-2} \int \{K\left(\varepsilon_i^{-1}(x_0-y)\right) - K\left(\varepsilon_i^{-1}(x_0+ta_n-y)\right)\}^2 f(y)dy$$

$$= (\varepsilon_n/n) \sum_{i=1}^{n} \varepsilon_i^{-1} \int \{K(y) - K(y+ta_n/\varepsilon_i)\}^2 f(x_0-\varepsilon_i y)dy .$$

If (3.1a) holds, the first integral expression is bounded by $Ma_n^2/\varepsilon_n^3$. If (3.1b) holds, a Taylor's expansion shows that the last integral expression is bounded by $M(a_n/\varepsilon_n)^2$. Hence, by Chebychev's inequality, the finite dimensional distributions each converge to zero in probability. To verify tightness, we again use the extension given by Bickel and Wichura (1971) of their Theorem 3; to do so, first define the process $V_n^{**}(s,t) = V_n^*(s,[nt]/n)$. It is clear that $V_n^*-V_n^{**} \xrightarrow{p} 0$, so that we may work with $V_n^{**}$. It is thus sufficient to verify the moments condition given by equation (3) of Bickel and Wichura (1971). We will adapt their notation and let $B,C$ be neighboring blocks. By the Schwartz inequality, it suffices to show that there exists $M>0$, $\beta>1$ for which if $i,p,q$ are integers,

$$\text{(3.5)} \qquad E|V_n(i/n,p/n) - V_n(i/n,q/n)|^4 \le M|i/n|^\beta \left|\frac{p-q}{n}\right|^\beta .$$

Letting

$$Z_{in} = \varepsilon_i^{-1}\{K(\varepsilon_i^{-1}(x_0+pa_n/n-X_i)) - K(\varepsilon_i^{-1}(x_0+qa_n/n-X_i))\}$$

$$- \varepsilon_i^{-1}E\left[K(\varepsilon_i^{-1}(x_0+pa_n/n-X)) - K(\varepsilon_i^{-1}(x_0+qa_n/n-X))\right] ,$$

we see that the term on the left hand side of (3.5) is

$$(\varepsilon_n/n)^2 \sum_{j=1}^{i} E\, Z_{jn}^4 + (\varepsilon_n/n)^2 \sum_{\substack{j\neq \ell \\ j,\ell\leq i}} E\, Z_{jn}^2\, E\, Z_{\ell n}^2 .$$

If (3.1a) holds, we have thus bounded (3.5) by (for some $M>0$ ),

$$M\left\{(\varepsilon_n/n)^2 i\left(\frac{p-q}{n}\right)^2(a_n/\varepsilon_n^2)^4 + (\varepsilon_n/n)^2 i^2\left(\frac{p-q}{n}\right)^2(a_n/\varepsilon_n^2)^4\right\} \leq M(i/n)^2\left(\frac{p-q}{n}\right)^2(a_n^2/\varepsilon_n^3)^2 ,$$

so that tightness holds in this case. If (3.1b) holds, we have

$$Z_{in} = \varepsilon_i^{-1}\{K(\varepsilon_i^{-1}(x_0+pa_n/n-X_i)) - K(\varepsilon_i^{-1}(x_0+qa_n/n-X_i))\} + O((p-q)a_n/n\varepsilon_n) ,$$

so that

$$E\, Z_{in}^4 \leq M_1\varepsilon_n^{-3}((p-q)/n)^2(a_n/\varepsilon_n)^4$$

and hence (3.5) is bounded by

$$M_2(i/n)^2\left(\frac{p-q}{n}\right)^2(a_n^4/\varepsilon_n^5) ,$$

which completes the proof.

<u>PROOF OF LEMMA 3.3.</u> We first show that each of the finite dimensional distributions converge to zero. Fix $s,t$, and assume (without loss of generality) that $a=-1$, $b=1$, so that

$$E|V_n(s)-V_n^*(s,t)|^2$$

$$\le (\varepsilon_n/n) \sum_{i=1}^{n} \varepsilon_i^{-2} \int_{x_0-\varepsilon_i+ta_n}^{x_0+\varepsilon_i} \{K(\varepsilon_i^{-1}(x_0-y)) - K(\varepsilon_i^{-1}(x_0-ta_n-y))\}^2 f(y)dy$$

$$+ (\varepsilon_n/n) \sum_{i=1}^{n} \varepsilon_i^{-2}\left\{\int_{x_0-\varepsilon_i}^{x_0-\varepsilon_i+ta_n} K^2(\varepsilon_i^{-1}(x_0-y)) f(y)dy + \int_{x_0+\varepsilon_i}^{x_0+\varepsilon_i+ta_n} K^2(\varepsilon_i^{-1}(x_0-ta_n-y)) f(y)dy\right\}$$

$$\le (\varepsilon_n/n) \sum_{i=1}^{n} \varepsilon_i^{-1} \int_{-1+ta_n/\varepsilon_i}^{1} [K(z)-K(z-ta_n/\varepsilon_i)]^2 f(x_0-\varepsilon_i z)dz + M_1(\varepsilon_n/n) \sum_{i=1}^{n} a_n/\varepsilon_n^2$$

$$\le M_2(\varepsilon_n/n) \sum_{i=1}^{n} \{a_n^2\varepsilon_i^{-3} + a_n\varepsilon_i^{-2}\} \to 0 \ .$$

Thus, the finite dimensional distributions converge, so that it remains only to verify tightness. As before, we first try to verify

(3.6) $$V_n^*-V_n^{**} \xrightarrow{p} 0 \ .$$

Let $I_A$ denote the indicator function of the event $A$, and define $A_{in}$ as the union of the intervals

$$[\varepsilon_i,\ \varepsilon_i+a_n+n^{-1}],\ [-\varepsilon_i,\ -\varepsilon_i+a_n+n^{-1}] \ .$$

Since $K$ is bounded and Lipschitz on its support, there is a constant $M$ for which

(3.7) $$|V_n^*(s,t)-V_n^{**}(s,t)| \le M\{(a_n/\varepsilon_n)^3 + (\varepsilon_n/n)^{\frac{1}{2}} \sum_{i=1}^{n} I_{A_{in}}\} \ .$$

By using Chebychev's inequality for fourth moments, we see that each term on the right hand side of (3.7) converges almost surely to zero. Hence we need only verify the tightness of $V_n - V_n^{**}$. Defining $Z_{in}$ as in the proof of Lemma 3.2, we see that for $p<q$, $p, q$ integers,

$$EZ_{in}^2 \le M\varepsilon_i^{-2}\left((p-q)/n)^2 (a_n/\varepsilon_i)^2\right) + \varepsilon_i^{-2} \int_{x_0-\varepsilon_i}^{x_0-\varepsilon_i+(q-p)a_n/n} K^2\left(\varepsilon_i^{-1}(x_0-y)\right) f(y-pa_n/n)\,dy$$

$$+ \varepsilon_i^{-2} \int_{x_0+\varepsilon_i}^{x_0+\varepsilon_i+(q-p)a_n/n} K^2\left[\varepsilon_i^{-1}\left(x_0-(q-p)a_n/n-y\right)\right] f(y-pa_n/n)\,dy$$

$$\le M_1\varepsilon_i^{-2} \left|\frac{p-q}{n}\right| a_n\{(a_n/\varepsilon_i)^2 \left|\frac{p-q}{n}\right| + 1\}$$

$$\le M_2\varepsilon_n^{-2} \left|\frac{p-q}{n}\right| a_n .$$

Also, one shows by similar computations that

$$EZ_{in}^4 \le M_3\varepsilon_i^{-4} \left|\frac{p-q}{n}\right| a_n .$$

Thus (3.5) is bounded by

$$\text{(3.8)} \qquad M\{(i/n)\varepsilon_n^{-2} \left|\frac{p-q}{n}\right| (a_n/n) + (i/n)^2 \left|\frac{p-q}{n}\right|^2 (a_n/\varepsilon_n)^2\} .$$

Since $a_n/n\varepsilon_n^2 = O(n^{-\beta})$ for some $\beta>0$, and since $\left|\frac{p-q}{n}\right| \ge 1/n$, (3.8) is bounded by

$$\text{(3.9)} \qquad M^* (i/n)^{1+\beta} \left|\frac{p-q}{n}\right|^{1+\beta} ,$$

which completes the proof.

## First Class of Stopping Rules

In the previous sections we have assumed (in (2.5c)) the existence of a stopping rule with the property that $N(d)/n(d)\to 1$ in probability as $d\to\infty$. While it is easy to write down various stopping rules, it is not clear how to develop rules which are based on reasonable criteria for density estimation.

We are aware of only one class of stopping rules for the density estimation problem; this class was suggested by Davies and Wegman (1975) who base their idea on stopping when $f_n$ and $f_{n-1}$ are close together. However, their stopping rules have not been shown to satisfy (2.5c) and the precise asymptotic behavior of the rules is not known, other than that they terminate with probability one and have certain moment properties. In this and the next section we will discuss two classes of stopping rules which are motivated in a natural manner from the ideas in the theory of fixed-width confidence intervals (Chow and Robbins (1965), Govindarajulu (1975)). We will verify (2.5c) and the related property

$$EN(d)/n(d) \to 1$$

for all the rules we propose, thus making their properties clear. One of the stopping rules, discussed in this section, will actually yield fixed-length confidence intervals for $f(x_0)$; we will also be able to discuss the asymptotic normality of this stopping rule itself.

In this and the next section we will make the following general assumptions. There will exist a sequence of density estimators $\hat{f}_n(x)$ such that

(4.1) for a sequence of statistics $\{T_n\}$, $\hat{f}_n(T_n) \to f(x_0)$ almost surely.

(4.2) if $N(d)/n(d) \xrightarrow{p} 1$, $B(K)$ is a constant depending on some known function $K$, and $N(0,\sigma^2)$ denotes a normal random variable with mean zero and variance $\sigma^2$, then

$$\left(N(d)\varepsilon_{N(d)}\right)^{\frac{1}{2}}\left(\hat{f}_{N(d)}(T_{N(d)}) - f(x_0)\right) \xrightarrow{L} N\left(0,\ f(x_0)B(K)\right) .$$

Some discussion is in order. We are not assuming $\hat{f}_n(x)$ is a special type of estimator such as discussed in Sections 2 and 3. However, as we have seen, these latter estimators often satisfy (4.2), and it is not in general very difficult to verify (4.1). For example, if $\hat{f}_n$ converges uniformly to $f$ in some neighborhood of $x_0$, $f$ is continuous, and $T_n$ converges to $x_0$ almost surely, then (4.1) holds since

$$|\hat{f}_n(T_n)-f(x_0)| \le \sup_x |\hat{f}_n(x)-f(x)| + |f(T_n)-f(x_0)| .$$

Conditions on uniform convergence are found in Schuster (1969) and Davies (1973).

Another important point to note is that many of the results in the next two sections hold if (4.2) is replaced by convergence as in Theorems 2.1 and 3.1, if $f(x_0)$ is merely replaced by the correct function. Unless specified otherwise (as in Lemma 4.4) this will be the case.

The first stopping rules we discuss arise in a manner similar to those introduced by Chow and Robbins (1965). If $\Phi$ is the normal distribution function and $\Phi^{-1}$ its inverse function, define

$$b = \left(B(K)\right)^{\frac{1}{2}}\Phi^{-1}(1-\alpha/2) . \tag{4.3}$$

<u>STOPPING RULE 4.1.</u> The stopping rule $N(d)$ stops the first time $n \ge n_0$ that

$$n\varepsilon_n \ge (b/d)^2 \hat{f}_n(T_n) .$$

<u>LEMMA 4.1.</u> Suppose that (4.1) holds and that $\varepsilon_n/\varepsilon_{n-1} \to 1$. Then

$$N(d)\varepsilon_{N(d)}\left(b^2 f(x_0)/d^2\right)^{-1} \to 1 \text{ a.s. as } d \to 0 .$$

LEMMA 4.2. Suppose (4.1) holds and $\varepsilon_n/\varepsilon_{n-1} \to 1$. If (4.2) holds, then

$$\lim_{d\downarrow 0} \Pr\{|\hat{f}_{N(d)}(T_{N(d)})-f(x_0)|\le d\} = 1-\alpha .$$

The proof of Lemma 4.1 is contained in Lemma 1 of Chow and Robbins (1965). The proof of Lemma 4.2 is immediate from (4.2) and (4.3).

Because of Lemma 4.2, we say that stopping rule 4.1 yields a confidence interval of fixed width $2d$ and prescribed coverage probability $1-\alpha$.

LEMMA 4.3. Suppose that $E\hat{f}_n(T_n)\to f(x_0)$ and that there is a constant $M^*$ for which

$$\sum_{n=1}^{\infty} \Pr\{|\hat{f}_n(T_n)-E\hat{f}_n(T_n)|>M^*\} < \infty . \tag{4.4}$$

If $\varepsilon n=n^{-\alpha}$ for some $\alpha>0$ and $\nu(d) = \left(b^2 f(x_0)/d^2\right)^{1/(1-\alpha)}$,

$$EN(d)/\nu(d) \to 1 .$$

REMARK 4.3. Schuster (1969) has shown that the kernel estimators $f_n(x)$ satisfy (4.4), while under the assumption $n^{-1}\sum_{i=1}^{n} \varepsilon_n/\varepsilon_i \to \alpha$, one may use the exponential bounds (Loeve (1968), page 254) to show that $f_n^*(x)$ satisfies (4.4).

PROOF OF LEMMA 4.3. Because of Lemma 4.1 and by Bickel and Yahev (1968), it suffices to show there is a $d_0>0$ for which

$$\sum_{m=1}^{\infty} \sup_{0<d<d_0} \Pr\{N(d)>md^{-2(1-\alpha)^{-1}}\} < \infty .$$

Letting $m(d)=md^{-2(1-\alpha)^{-1}}$, we see that

(4.5a) $N(d)^{1-\alpha}-\nu(d)^{1-\alpha} \geq (\nu(d))^{1-\alpha}(f_{N(d)}(T_{N(d)})-f(x_0))/f(x_0)$

(4.5b) $N(d)^{1-\alpha}-\nu(d)^{1-\alpha} \leq (\nu(d))^{1-\alpha}(f_{N(d)-1}(T_{N(d)-1})-f(x_0))/f(x_0)+2n_0$ .

Since $N(d)/\nu(d)\to 1$ a.s. as $d\to 0$, multiplying all terms by $\nu(d)^{-(1-\alpha)/2}$ completes the proof.

Thus, taken together, Lemmas 4.1, 4.3 and 4.4 yield a great deal of information, telling us in detail how the stopping rule 4.1 behaves. Lemma 4.2 gives us a very nice property of the stopping rule, namely that it yields confidence intervals for $f(x_0)$ of fixed length and prescribed coverage probability. It should be mentioned here that Starr (1966) has shown that for this same class of stopping rules when trying to estimate the mean of a normal distribution the following approximation is almost true for all values of $d$:

$$\Pr\{|\overline{X}_{N(d)}-\mu|\leq d\} \doteq 1-\alpha .$$

This indicates that stopping rule 4.1 may well achieve its asymptotic behavior for moderate values of $d$.

A second stopping rule of this type is more global in scope and would appear more useful in situations where one is interested in estimating the density at the mode. Specifically,

<u>STOPPING RULE 4.2.</u> The stopping rule $N(d)$ stope the first time $n\geq n_0$ that

$$n\varepsilon_n \geq (b/d)^2 \sup_x \hat{f}_n(x) .$$

$$\Pr\{N(d)>m(d)\} \le \Pr\{b^2\hat{f}_{m(d)}(T_{m(d)})>d^2(m(d))^{1-\alpha}\}$$
$$\le \Pr\{|\hat{f}_{m(d)}(T_{m(d)})-E\hat{f}_{m(d)})|>m^{1-\alpha}b^{-2}-E\hat{f}_{m(d)}(T_{m(d)})\} .$$

LEMMA 4.4. Suppose (4.1) and (4.2) hold and $\varepsilon_n=n^{-\alpha}$ for some $\alpha>0$. Then if $A(F)=B(K)f(x_0)$

$$f(x_0)\nu(d)^{-(1-\alpha)/2}\left(N(d)^{1-\alpha}-\nu(d)^{1-\alpha}\right)/A(F) \xrightarrow{L} N(0,1) .$$

REMARK 4.1. This Lemma requires (4.2). It may not be true if convergence only on the order of Theorems 2.1 and 3.1 is known.

PROOF OF LEMMA 4.4. The proof follows exactly along the lines of a result due to Ghosh and Mukhophadyay (1975). Since their result is as yet unpublished, we will sketch the proof. Since $N(d)\to\infty$ a.s., by (4.1) and (4.2),

$$N(d)^{(1-\alpha)/2}\left(f_{N(d)}(T_{N(d)})-f(x_0)\right)/A(F) \xrightarrow{L} N(0,1)$$
$$\left(N(d)-1\right)^{(1-\alpha)/2}\left(f_{N(d)-1}(T_{N(d)-1})-f(x_0)\right)/A(F) \xrightarrow{L} N(0,1) .$$

Now,

$$N(d)^{1-\alpha} \ge (b/d)^2 f_{N(d)}(T_{N(d)})$$
$$\left(N(d)-1\right)^{1-\alpha} \le (b/d)^2 f_{N(d)-1}(T_{N(d)-1}) + n_0$$

so that

$$N(d)^{1-\alpha} \le (b/d)^2 f_{N(d)-1}(T_{N(d)-1}) + 2n_0 .$$

Thus,

LEMMA 4.5. Suppose there is a unique $x_0$ such that

$$\max_x f(x) = f(x_0)$$

and

$$\max_x \hat{f}_n(x) \to f(x_0) \text{ a.s. as } n\to\infty .$$

If $\varepsilon_n = n^{-\alpha}$ and $\nu(d) = \left(b^2 f(x_0)/d^2\right)^{1/(1-\alpha)}$, then

$$N(d)/\nu(d)\to 1 \text{ a.s. as } d\to 0 .$$

If, in addition, there is a constant $M^*$ for which

$$(4.6) \qquad \sum_{n=1}^{\infty} \Pr\{\sup_x |\hat{f}_n(x)-f(x)|>M^*\} < \infty ,$$

we have

$$EN(d)/\nu(d)\to 1 \text{ a.s.}$$

The proof of Lemma 4.5 is the same as that of Lemmas 4.1 and 4.3. Schuster (1969) has shown that the kernel estimators $f_n(x)$ satisfy (4.6), but it is not known whether the estimators $f_n^*(x)$ satisfy (4.6). We have also been unable to obtain a result similar to Lemma 4.4.

Stopping rule 4.2 is a competitor to Stopping rule 4.1 in the case that $f$ is symmetric and unimodal and $x_0$ is the mode. $T_n$ here might be the sample median. The almost sure asymptotic properties of the two rules would be the same, but rule 4.2 would of course always take more observations than rule 4.1. However, in this particular case, both yield fixed-width confidence intervals of prescribed coverage probability for $f(x_0)$.

Finally, note that $N(d)$ defined by stopping rule 4.2 diverges with probability one, as $d\to 0$ so that

$$(4.7)\quad \sup_x |\hat{f}_n(x)-f(x)|\to 0 \text{ almost surely} \Rightarrow \sup_x |\hat{f}_{N(d)}(x)-f(x)|\to 0 \text{ almost surely.}$$

## A Second Class of Stopping Rules

The second class of stopping rules we investigate is motivated by work of Farrell (1966) and Sen and Ghosh (1971). Their idea was to obtain upper and lower bounds on the parameter of interest and to stop when the difference in these two bounds becomes at most $2d$. The first rule of this section then becomes

STOPPING RULE 5.1. Define for some sequence of constants $\{b_n\}$ decreasing to zero

$$(5.1)\qquad V_n = \left(\hat{f}_n(T_n+b_n)-\hat{f}_n(T_n)\right)/\hat{f}_n(T_n) ,$$

and let $N(d)$ be the first time $n\geq n_0$ that $|V_n|\leq 2d$.

The motivation for dividing by $\hat{f}_n(T_n)$ in (5.1) is that one will stop when there is little change in $\hat{f}_n$ (in a neighborhood of $x_0$ ) relative to $f(x_0)$. At the end of this section we briefly discuss a rule which does not divide by $\hat{f}_n(T_n)$. As a notational device we make the following:

DEFINITION 5.1. A sequence of statistics $Y_n=o^*(a_n)$ if for all $\varepsilon>0$,

$$\sum_{n=1}^{\infty} \Pr\{|Y_n|>\varepsilon a_n\} < \infty .$$

Now, in order to investigate the stopping rule $N(d)$ we want to look at $V_n$.

LEMMA 5.1. Suppose that $f$ has three continuous bounded derivatives in a neighborhood of $x_0$, $f^{(2)}(x_0) \neq 0$, and that

(5.2a) $$T_n - x_0 = o^*(b_n)$$

(5.2b) $\sup|\hat{f}_n(x)-f(x)| = o^*(b_n^2)$, where the supremum is taken in some neighbor hood of $x_0$.

Then, if $f^{(1)}(x_0) = 0$,

(5.3) $$\left(2f(x_0)/b_n^2 f^{(2)}(x_0)\right)V_n - 1 = o^*(1) .$$

If $f^{(1)}(x_0) \neq 0$,

(5.4) $$\left(f(x_0)/b_n f^{(1)}(x_0)\right)V_n - 1 = o^*(1) .$$

PROOF: Note that

$$\hat{f}_n(T_n+b_n) - \hat{f}_n(T_n) = \hat{f}_n(T_n+b_n) - f(T_n+b_n) + f(T_n) - \hat{f}_n(T_n) + f(T_n+b_n) - f(T_n)$$

$$= H_n + b_n f^{(1)}(T_n) + b_n^2 f^{(2)}(T_n)/2 + b_n^3 f^{(3)}(\xi_n)/6 ,$$

where $|H_n| \leq 2 \sup|\hat{f}_n(x)-f(x)|$. Thus,

$$\hat{f}_n(T_n+b_n) - \hat{f}_n(T_n) = b_n f^{(1)}(x_0) + \tfrac{1}{2} b_n^2 f^{(2)}(x_0) + o^*(b_n^2) .$$

Now (5.3) and (5.4) follow easily.

REMARK 5.1. For a discussion of (5.2b), see the remarks after Lemma 4.5.

We are now in a position to discuss the almost sure behavior of $N(d)$.

THEOREM 5.1. Define

$$\nu_1(d) = |4f(x_0)d/f^{(2)}(x_0)|^{-\frac{1}{2}\alpha}$$

$$\nu_2(d) = |2f(x_0)d/f^{(1)}(x_0)|^{-1/\alpha} .$$

Suppose $b_n = n^{-\alpha}$ satisfies the conditions of Lemma 5.1. Then, if $f^{(1)}(x_0)=0$,

$$N(d)/\nu_1(d) \to 1 \text{ almost surely as } d \to 0.$$

If $f^{(1)}(x_0) \neq 0$,

$$N(d)/\nu_2(d) \to 1 \text{ almost surely as } d \to 0.$$

PROOF: The proof follows a method of Sen and Ghosh (1971). We consider only the case $f^{(1)}(x_0)$. If "$\subseteq$" denotes set inclusion and "$\cup$" set union,

$$\begin{aligned}(5.5)\quad \{(N(d)-\nu_1(d)|\geq \varepsilon\nu_1(d)\} &\subseteq \{N(d)\geq(1+\varepsilon)\nu_1(d)\} \cup \{N(d)\leq(1-\varepsilon)\nu_1(d)\} \\ &\subseteq \{V_{(1+\varepsilon)\nu_1(d)} > 2d\} \\ &\qquad \cup \{V_n \leq 2d \text{ for some } n \leq (1-\varepsilon)\nu_1(d)\} .\end{aligned}$$

The first set on the right hand side of (5.5) is contained in

$$\left\{\frac{2f(x_0)}{f^{(2)}(x_0)} \left((1+\varepsilon)\nu_1(d)\right)^{2\alpha} V_{[1+\varepsilon]\nu_1(d)} - 1 > (1+\varepsilon)^{2\alpha} - 1\right\} ,$$

so that by Lemma 5.1, it suffices to consider only the last event in (5.5). Now, for any $n \leq (1-\varepsilon)\nu_1(d)$,

$$\{V_n \leq 2d\} \subseteq \left\{\left|\frac{2f(x_0)}{b_n^2 f^{(2)}(x_0)} V_n - 1\right| \geq 1 - (1-\varepsilon)^{2\alpha}\right\} ,$$

so that again by Lemma 5.1 and since $d \to 0$, the proof is now complete.

The next step is to show that an analogue to Lemma 4.3 holds. Before proceeding, a few definitions are needed.

DEFINITION 5.2. Let

$$n_1(d) = [(1-\varepsilon)\nu_1(d)],\; n_2(d) = [(1+\varepsilon)\nu_1(d)],\; n_2^*(d) = [(1+\varepsilon)\nu_2(d)] ,$$

$$H_\varepsilon(d) = \sum_{n_2(d)}^{\infty} \Pr\{N(d)>n\},\; H_\varepsilon^*(d) = \sum_{n_2^*(d)}^{\infty} \Pr\{N(d)>n\} .$$

LEMMA 5.3. Under the conditions of Lemma 5.1,

$$\lim_{d\to 0} H_\varepsilon(d) < \infty \quad \text{if} \quad f^{(1)}(x_0) = 0$$

$$\lim_{d\to 0} H_\varepsilon^*(d) < \infty \quad \text{if} \quad f^{(1)}(x_0) \neq 0 .$$

PROOF: Again, consider only the case $f^{(1)}(x_0) = 0$. Then for some $\varepsilon'>0$,

$$H_\varepsilon(d) \le \sum_{n_2(d)}^{\infty} \Pr\left\{\frac{2f(x_0)}{b_n^2 f^{(2)}(x_0)} V_n - 1 > \frac{4f(x_0)d}{b_n^2 f^{(2)}(x_0)} - 1\right\}$$

$$\le \sum_{n_2(d)}^{\infty} \Pr\left\{\left|\frac{2f(x_0)}{b_n^2 f^{(2)}(x_0)} V_n - 1\right| > \varepsilon'\right\} .$$

The last sum converges and is a decreasing function of $d$.

LEMMA 5.4. Under the conditions of Lemma 5.1,

(5.6a) $$EN(d)/\nu_1(d) \to 1 \quad \text{if} \quad f^{(1)}(x_0) = 0$$

(5.6b) $$EN(d)/\nu_2(d) \to 1 \quad \text{if} \quad f^{(1)}(x_0) \neq 0 .$$

PROOF: We will only show (5.6a). We have

$$EN(d)/\nu_1(d) = \nu_1^{-1}(d)\left(\sum^{'} + \sum^{2} + \sum^{3} nPr\{N(d)=n\}\right) ,$$

where

$$\sum^1 \text{ extends over } \{n\le n_1(d)\}$$

$$\sum^2 \text{ extends over } \{n_1(d)<n<n_2(d)\}$$

$$\sum^3 \text{ extends over } \{n\ge n_2(d)\} .$$

Now,

$$\nu_1^{-1}(d) \sum^1 nPr\{N(d)=n\} \le (1-\varepsilon)Pr\{N(d)\le n_1(d)\} \to 0$$

by Lemma 5.1. Also,

$$\nu_1^{-1}(d) \sum^3 nPr\{N(d)=n\} \le \nu_1^{-1}(d) \sum^3 Pr\{N(d)\ge n\} + \nu_1^{-1}(d)\; n_2(d)\; Pr\{N(d)>n_2(d)\} \to 0$$

by Lemma 5.2. Finally,

$$\left|\nu_1^{-1}(d) \sum^2 nPr\{N(d)=n\} - 1\right|$$

$$\le \left|\sum^2 \left(\frac{n}{\nu_1(d)}-1\right)Pr\{N(d)=n\}\right| + \left(\sum^1 + \sum^3 Pr\{N(d)=n\}\right) \to 0$$

as $\varepsilon\to 0$, $d\to 0$. This completes the proof.

The choice of $V_n$ given in (5.1) is certainly not the only possible one available. We list below more statistics $V_n$ and the sequence of constants $\nu_1(d)$, $\nu_2(d)$ that go with them:

(5.7) $$|V_n^1| = \max\left\{\left|\frac{\hat{f}_n(T_n+b_n)-\hat{f}_n(T_n)}{\hat{f}_n(T_n)}\right|,\ \left|\frac{\hat{f}_n(T_n-b_n)-\hat{f}_n(T_n)}{\hat{f}_n(T_n)}\right|\right\}$$

$\nu_1(d)$, $\nu_2(d)$ as in Lemma 5.2

(5.8) $$V_n^2 = \hat{f}_n(T_n+b_n) - \hat{f}_n(T_n)$$

$$\nu_1(d) = |4d/f^{(2)}(x_0)|^{-1/2\alpha}$$

$$\nu_2(d) = |2d/f^{(1)}(x_0)|^{-1/\alpha}$$

(5.9) $$|V_n^3| = \max\{|\hat{f}_n(T_n+b_n) - \hat{f}_n(T_n)|,\ |\hat{f}_n(T_n-b_n) - \hat{f}_n(T_n)|\}$$

$\nu_1(d)$, $\nu_2(d)$ as in (5.8).

Again, modifications of these stopping rules along the lines of Lemma 4.5 are also possible. When this is done, we again see that since $N(d)$ diverges with probability one as $d\to 0$,

$$\sup_x|\hat{f}_n(x)-f(x)|\to 0 \text{ almost surely}$$

$$\Rightarrow \sup_x|\hat{f}_{N(d)}(x)-f(x)|\to 0 \text{ almost surely .}$$

## REFERENCES


[1] Bahadur, R.R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* (37) 577-580.

[2] Bickel, P.J. and Wichura, M.J. (1971). Convergence criteria for multi-parameter stochastic processes and some applications. *Ann. Math. Statist.* (42) 1656-1670.

[3] Bickel, P.J. and Yahav, J.A. (1968). Asymptotically optimal Bayes and minimax procedures in sequential estimation. *Ann. Math. Statist.* (39) 442-456.

[4] Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[5] Cacoullos, T. (1966). Estimation of a multivariate density. *Ann. Inst. Statist. Math.* (18) 178-189.

[6] Chow, Y.S. and Robbins, H., (1965). On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *Ann. Math. Statist.* (36) 463-467.

[7] Davies, H.I. (1973). Strong consistency of a sequential estimation of a probability density function. *Bull. Math. Statist.* (15), 49-54.

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

[9] Farrell, R.H. (1966). Bounded length confidence intervals for the *p*-point of a distribution function III. *Ann. Math. Statist.* (37) 586-592.

[10] Ghosh, M. and Mukhophadyay, N. (1975). Asymptotic normality of stopping times in sequential analysis. Unpublished paper.

[11] Govindarajulu, Z. (1975). *Sequential Statistical Procedures.* Academic Press, New York.

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

[13] Loéve, M. (1968). *Probability Theory*, 3rd ed., Van Nostrand, Princeton.

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

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

[16] Sager, T.W. (1975). Consistency in nonparametric estimation of the mode. *Ann. Statist.* (3) 698-706.

[17] Schuster, E.F. (1969). Estimation of a probability density function and its derivatives. *Ann. Math. Statist.* (40) 1187-1195.

[18] Sen, P.K. and Ghosh, M. (1971). On bounded length confidence intervals based on one-sample rank order statistics. *Ann. Math. Statist.* (42) 189-203.

[19] Srivastava, R.C. (1973). Estimation of probability density function based on random number of observations with applications. *Int. Statist. Rev.* (41) 77-86.

[20] Starr, N. (1966). The performance of a sequential procedure for the fixed-width interval estimate. *Ann. Math. Statist.* (36) 36-50.

[21] Venter, J.H. (1967). On estimation of the mode. *Ann. Math. Statist.* (38) 1446-1455.

[22] Wegman, E.J. (1972). Nonparametric probability density estimation: I. A summary of available methods. *Technometrics* (14) 533-546.

[23] Wegman, E.J. and Davies, H.I. (1975). Remarks on some recursive estimators of a probability density. Institute of Statistics Mimeo Series #1021, University of North Carolina at Chapel Hill.

[24] Woodroofe, M. (1967). On the maximum deviation of the sample density. *Ann. Math. Statist.* (38) 475-481.

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