A PhD dissertation under the direction of EDWARD J. WEGMAN.

ON THE SEQUENTIAL ESTIMATION OF A PROBABILITY DENSITY FUNCTION

by

H.I. Davies

Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 918
April, 1974

H. I. DAVIES On the sequential estimation of a probability density function. (Under the direction of EDWARD J. WEGMAN.)

ABSTRACT

Using kernel estimators of both the Parzen- and Yamato-types, a "naive" sequential procedure for estimating an unknown probability density function is developed. It is based on the function

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

through the inequality $|V_n(x)| \leq \varepsilon$, where $\varepsilon \geq 0$ is given. The asymptotic structure of $V_n(x)$ is examined for both the Parzen and Yamato estimators.

Properties of the procedure are investigated and in particular it is shown that the stopping rule obtained has finite expectation and variance and that it is closed. It is shown that the mean square error tends to zero as $\varepsilon \to 0$ for the estimate $\hat{f}_N(x)$ where N is the stopping variable defined by the "naive" stopping rule. Some properties of the stopping variable are examined and it is shown that under certain conditions $N \to \infty$ with probability one as $\varepsilon \to 0$.

Properties of an estimator $f_{N_\varepsilon}(x)$, where N_ε is a stopping variable, are investigated and in particular it is shown that under appropriate conditions $f_{N_\varepsilon}(x)$ is both mean square consistent and asymptotically normal.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Firstly, I am greatly indebted to my adviser, Professor E.J. Wegman who proposed this investigation and who made so many valuable comments during its course. His encouragement and stimulating ideas were of great inspiration to me.

My sincere thanks go also to Professors N.L. Johnson and Professor W. Hoeffding for reading the manuscript and for making so many helpful suggestions. A special thanks is due to Professor G.D. Simons for his reading of the manuscript at such short notice.

I also wish to thank Professor R.L. Davis of the mathematics department for his help and for his contribution to my education.

In addition I thank Professor M.R. Leadbetter for his advice during my first three years as a graduate student in the Department of Statistics. I also extend my thanks to all those members of the faculty in the Department of Statistics and Mathematics who have contributed to my education at the University of North Carolina.

I am also grateful for the help extended to my by the Departmental Secretaries, especially Mrs.June Maxwell for her help and suggestions as to the typing of this dissertation.

I gratefully thank Mrs. Chas. Burnett who allowed my wife and I to live in her home for the last three months of our stay in Chapel Hill, a circumstance that contributed greatly to the completion of this dissertation.

I also gratefully acknowledge the Commonwealth Scientific

Industrial Research Organisation under whose studentship I have been studying. Without this aid this work could not have been carried out.

Finally, there are no words to thank my wife Sue enough. Not only did she painstakingly and accurately type the manuscript but she also encouraged and supported me during the years leading to its completion.

CHAPTER 1

INTRODUCTION

1.1. Introduction and Literature Review:

The estimation of a probability density function, $f(x)$, using non-parametric methods has been considered by several authors. Three principal types of estimator have been extensively discussed in the literature, Kernel, Orthogonal series and Maximum likelihood although a fourth method based on the use of spline functions has just recently been investigated (Wahba 1971, Hill 1973). Each of the four approaches is based on obtaining a random sample, $X_1, X_2, \cdots X_n$, of size n, of independently and identically distributed random variables from a population with density function $f(x)$ and using this sample to calculate an estimate, $\hat{f}_n(x)$, of $f(x)$. Depending on which of the four estimators we use this estimate we obtain may be for a fixed point x only or it may be a global estimate true for all x. In this dissertation we will be concerned only with fixed point estimators.

A natural question that arises is how to develop a rule for determining the sample size n so that one may expect to satisfy (or better) some predetermined error criterion. This leads us to consider a sequential method. However in order to develop a sequential method it is necessary to know the fixed sample properties of the estimator we are to use. Since kernel estimators have received the most discussion in the literature we will thus confine our attention in this dissertation to that type of estimator. The kernel estimator has the form

$$(1.1.1.) \qquad f_n(x) = \int_{-\infty}^{\infty} K_n(x,y)\, dF_n(y) = \frac{1}{n} \sum_{j=1}^{n} K_n(x,X_j)$$

where F_n is the empirical distribution and K_n is the kernel function.

Rosenblatt (1956) was the first to introduce estimators of the form (1.1.1.) and used the kernel

$$(1.1.2.) \qquad K_n(x,y) = \frac{1}{h_n} K\left(\frac{x-y}{h_n}\right)$$

where h_n is a sequence of positive numbers satisfying $\lim_{n\to\infty} h_n = 0$ and $\lim_{n\to\infty} nh_n = \infty$. Estimators of this form were more fully discussed by Parzen (1962). They will be denoted by $\hat{f}_n(x)$ throughout this dissertation. Since the results of Parzen will be important in later chapters we summarize his results in the following theorem.

Theorem 1.1.1.

If $K(y)$ and $\{h_n\}$ satisfy appropriate conditions [1] then the kernel estimate, $\hat{f}_n(x)$, based on (1.1.2.)

(i) is an asymptotically unbiased estimator of $f(x)$ at all points x at which $f(x)$ is continuous.

(ii) If $\sigma^2\left(\hat{f}_n(x)\right)$ denotes the variance of $\hat{f}_n(x)$ then

$$\lim_{n\to\infty} nh_n\, \sigma^2\left(\hat{f}_n(x)\right) = f(x) \int_{-\infty}^{\infty} K^2(y)\, dy,$$

at all points of continuity of $f(x)$.

(iii) If $nh_n \to \infty$ as $n\to\infty$, then $\lim_{n\to\infty} E[\hat{f}_n(x) - f(x)]^2 = 0$

1.Chapter 2, equations (2.1.3.), and (2.1.4.).

form

$$f_n(x) = \int_{-\infty}^{\infty} K_n(x,y)\, dF_n(y) = \frac{1}{n} \sum_{j=1}^{n} K_n(x,X_j) \tag{1.1.1.}$$

where F_n is the empirical distribution and K_n is the kernel function.

Rosenblatt (1956) was the first to introduce estimators of the form (1.1.1.) and used the kernel

$$K_n(x,y) = \frac{1}{h_n} K\left(\frac{x-y}{h_n}\right) \tag{1.1.2.}$$

where h_n is a sequence of positive numbers satisfying $\lim_{n\to\infty} h_n = 0$ and $\lim_{n\to\infty} nh_n = \infty$. Estimators of this form were more fully discussed by Parzen (1962). They will be denoted by $\hat{f}_n(x)$ throughout this dissertation. Since the results of Parzen will be important in later chapters we summarize his results in the following theorem.

Theorem 1.1.1.

If $K(y)$ and $\{h_n\}$ satisfy appropriate conditions [1] then the kernel estimate, $\hat{f}_n(x)$, based on (1.1.2.)

(i) is an asymptotically unbiased estimator of $f(x)$ at all points x at which $f(x)$ is continuous.

(ii) If $\sigma^2\left(\hat{f}_n(x)\right)$ denotes the variance of $\hat{f}_n(x)$ then

$$\lim_{n\to\infty} nh_n\, \sigma^2\left(\hat{f}_n(x)\right) = f(x) \int_{-\infty}^{\infty} K^2(y)\, dy,$$

at all points of continuity of $f(x)$.

(iii) If $nh_n \to\infty$ as $n\to\infty$, then $\lim_{n\to\infty} E[\hat{f}_n(x) - f(x)]^2 = 0$

1.Chapter 2, equations (2.1.3.), and (2.1.4.).

and we say $\hat{f}_n(x)$ is <u>weakly</u> consistent.

(iv) If f(x) is uniformly continuous then $\hat{f}_n(x)$ is uniformly weakly consistent. That is, given any $\varepsilon>0$,

$$P\{\sup_x |\hat{f}_n(x) - f(x)| < \varepsilon\} \to 1 \text{ as } n\to\infty.$$

(v) $$\lim_{n\to\infty} P\left(\frac{\hat{f}_n(x) - f(x)}{\sigma(\hat{f}_n(x))} \le c\right) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{c} e^{-\frac{1}{2}u^2} du = \Phi(c).$$

Van Ryzin (1969) also considers the estimate $\hat{f}_n(x)$, showing that if x is a continuity point of f(x) then provided K(u) and h_n satisfy certain conditions[2], $\hat{f}_n(x) \to f(x)$ as $n\to\infty$ with probability one. He then strengthens the conditions to show that provided f(x)is uniformly continuous $\sup_x |\hat{f}_n(x) - f(x)| \to 0$ as $n\to\infty$, with probability one. We call the first of these properties strong consistency and the latter uniform strong consistency. Van Ryzin's results are proven in the multivariate case in that he assumes x is an m - dimensional vector and X_i is an m - dimensional random vector.

Both Parzen and Van Ryzin discuss estimation of the mode. However we will not consider that problem.

The uniform consistency of $\hat{f}_n(x)$ with probability one is also proved by Nadaraya (1965). He however uses a different set of conditions to Van Ryzin.

We will also discuss kernel estimators of the form introduced by Yamato (1972). The kernel in this case has the form

---

2. See Chapter 2 conditions for lemma 2.2.2.

$$K_n(x, X_j) = \frac{1}{h_j^m} K\left(\frac{x-X_j}{h_j}\right) \tag{1.1.3.}$$

where X_j is an m - dimensional random vector and x is an m - dimensional vector. The Yamato-type estimator will be denoted by $\tilde{f}_n(x)$. When m=1, Yamato's results parallel those of Parzen summarized in Theorem 1.1.1, the main difference being that if $\sigma^2(\tilde{f}_n(x))$ denotes the variance of $\tilde{f}_n(x)$ then

$$\lim_{n\to\infty} nh_n \sigma^2(\tilde{f}_n(x)) = \nu_0 f(x) \int_{-\infty}^{\infty} K^2(y)\, dy$$

where ν_0 is a constant such that $0<\nu_0 \leq 1$. This means that $\tilde{f}_n(x)$ is asymptotically at least as good as $\hat{f}_n(x)$ in the sense that the asymptotic variance of $\tilde{f}_n(x)$ is less than or equal to the asymptotic variance of $\hat{f}_n(x)$.

Yamato also notes that $\tilde{f}_n(x)$ can be written as

$$\tilde{f}_n(x) = \frac{n-1}{n} \tilde{f}_{n-1}(x) + \frac{1}{nh_n^m} K\left(\frac{x-X_n}{h_n}\right)$$

so that he claims it is more suitable than $\hat{f}_n(x)$ in order to correct the estimate successively in the case where a sequence of random vectors or variables is observed. Yamato in fact refers to his estimator as sequential though he makes no attempt to use it in a sequential context. By this we mean that he is concerned only with fixed sample properties and asymptotic properties of $\tilde{f}_n(x)$ and makes no attempt to suggest a stopping rule for use with $\tilde{f}_n(x)$.

A more general kernel function is discussed by Watson and Leadbetter (1963). They consider

$$K_n(x,y) = \delta_n(x-y) \tag{1.1.4}$$

where $\{\delta_n(x)\}$ satisfies (i) $\int |\delta_n(x)|\, dx < A$, all n, some fixed A, (ii) $\int \delta_n(x)\, dx = 1$, all n, (iii) $\delta_n(x) \to 0$ uniformly in $|x| \geq \lambda$, for any fixed $\lambda > 0$, (iv) $\int_{|x|\geq\lambda} |\delta_n(x)|\, dx \to 0$ as $n\to\infty$ for any fixed $\lambda>0$.

We will not discuss estimators using kernels of the form (1.1.4) but will only note that using mean integrated square error (M.I.S.E.) as a basis, that is

$$E \int_{-\infty}^{\infty} \left(f_n(x) - f(x)\right)^2 dx,$$

Watson and Leadbetter find the optimal δ_n which minimizes that quantity. They show that the form of the optimum estimate is in general a complicated expression dependent on the explicit form of the density function and depends heavily on the form of the characteristic function of f. Special forms of the characteristic function are considered and consistency in quadratic mean is shown to be attained in each of the special cases.

Leadbetter (1963) discusses pointwise consistency results for L_2 densities using mean square error as a basis for his discussion.

Srivastava (1973) treats the problem of the estimation of a probability density function when the sample size is a random variable. He shows that if N_t is the number of observations that occur in a time interval (o,t], then as $t\to\infty$, $\hat{f}_{N_t}(x)$ is both a consistent and uniformly consistent estimator of f(x) at the point x. Further, he proves that,

$$\lim_{t\to\infty} P\left[\left(N_t\, h_{N_t}\right)^{1/2} \left(\hat{f}_{N_t}(x) - f(x)\right) \leq c\right] = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{c} e^{-\frac{u^2}{2}}\, du.$$

Srivastava discusses both the case when N_t is independent and also when it is dependent on the observations. To my knowledge this paper is the only one that considers density estimation when the sample size is a random variable.

Two questions now arise quite naturally in the discussion of kernel estimators.

(i) How should we choose the kernel function, K(u), and

(ii) how should we select the sequence h_n?

Since these questions are not only relevant to the fixed sample problem but also to the sequential problem we will briefly review the information that is available in the literature. First we consider the choice of K(u).

1.2. Choice of K(u):

Parzen (1962) requires that K(u) be a non-negative, even, Borel function satisfying (i) $\sup_u K(u) < \infty$, (ii) $\int_{-\infty}^{\infty} K(u)\, du < \infty$ and (iii) $\lim_{|u|\to\infty} |u| K(u) = 0$. In addition he requires

$\int_{-\infty}^{\infty} K(y)\, dy = 1$ so that $\hat{f}_n(x)$ will be a density function. He suggests several possible weighting functions (see Appendix 1) and evaluates $\int_{-\infty}^{\infty} K^2(y)\, dy$ for these kernels. From the form of the asymptotic variance of $\hat{f}_n(x)$ (Theorem 1.1.1.) we see that it may be minimized by minimizing $\int_{-\infty}^{\infty} K^2(y)\, dy$ so that Parzen's calculations do give some idea which kernel is to be preferred in this sense. No attempt is made to find the "optimal" kernel which would minimize $\int_{-\infty}^{\infty} K^2(y)\, dy$ subject to the given conditions.

This problem is discussed by Epanechnikov (1969). Epanechnikov defines "relative global approximation error" as

$$\hat{u}^2 = \frac{1}{Q} \int E \left(\hat{f}_n(x) - f(x)\right)^2 dx$$

where x is a m - dimensional vector and

$$Q = \int f^2(x)\ dx.$$

We note that for m=1, $\hat{f}_n(x)$ corresponds to Parzen's estimate.

Letting $n\to\infty$, Epanechnikov obtains a "relative global approximation error"

$$\hat{u}^2 \sim \frac{n^{-1}\ h_n^{-m}\ L^m + (1/4)\ h_n^4\ M}{Q}$$

where

$$L = \int_{-\infty}^{\infty} K^2(y)\ dy$$

and

$$M = \int \cdots \int \left(\sum_{\ell=1}^{m} \frac{\partial^2}{\partial x_1^2} f\left(x_1, \cdots, x_m\right)\right)^2 dx_1 \cdots dx_m.$$

The "optimal" kernel, $K_0(y)$, is obtained by minimizing "relative global error" which amounts to minimizing L subject to certain constraints. He obtains

$$K_0(y) = \begin{cases} \dfrac{3}{4\sqrt{5}} - \dfrac{3y^2}{20\sqrt{5}} & \text{for } |y| \leq \sqrt{5} \\ 0 & \text{for } |y| > \sqrt{5} \end{cases}$$

Epanechnikov shows that $K_0(y)$ is independent of the true probability density, the sample size and the dimensionality of the space. He calculates L and $r = L / \int_{-\infty}^{\infty} K_0^2(y)\ dy$ for several kernels (see Appendix 1, table 2). The calculations indicate that relative global error is not necessarily greatly altered by using a non-optimal kernel and for this reason we feel justified in using a non-optimal kernel in later chapters, especially if calculations are made simpler by so doing.

Anderson (1969) carried out a Monte - Carlo study and showed that if the optimum h_n is used then M.I.S.E. is also relatively independent of the kernel. The choice of kernel seems far less critical than the choice of h_n which we will now discuss.

1.3. Choice of h_n:

Rosenblatt (1956) shows that for large n the mean square error can be written approximately as

$$E\,|\hat{f}_n(x) - f(x)|^2 \sim \frac{f(x)}{2nh_n} + \frac{h_n^4}{36}\,|f''(x)|^2 + 0\left(\frac{1}{nh_n} + h_n^4\right)$$

If $h_n = B\,n^{-\alpha}$, $\alpha>0$ and B a constant independent of n, then the optimum choice of h_n that minimizes mean square error is attained for $\alpha = {}^1/_5$

and
$$B = \left(\frac{9}{2}\;\frac{f(x)}{|f''(x)|^2}\right)^{1/5}$$

Similarly, if M.I.S.E. is to be minimized then again $\alpha = {}^1/_5$

$$B = \left(\frac{9}{2\int_{-\infty}^{\infty} |f''(x)|^2\,dx}\right)^{1/5}$$

Unfortuneately, the unknown density function f appears in the choice of B. However, we can still obtain both pointwise consistency and consistency in quadratic mean for a suitable choice of B. Rosenblatt shows that if the first three derivatives of f exist at x, then the optimum choice of h_n gives a mean square error no smaller than $0(n^{-4/5})$.

Parzen (1962) obtains the optimum sequence h_n that minimizes mean square error as

$$h_n = \left(f(x) \int_{-\infty}^{\infty} K^2(y)\ dy\right) \left(n^{2r} |k_r\ f^{(r)}(x)|^2\right)^{\frac{1}{(2r+1)}}, \qquad r>0$$

where $k_r = \lim_{r \to 0} \frac{[1 - k(u)]}{|u|^r}$ is not zero and

$f^{(r)}(x) = -(2\pi)^{-1} \int_{-\infty}^{\infty} e^{iux}\ |u|^r\ \phi(u)\ du$. Note k(u) is the Fourier transform of K(x) and $\phi(u)$ is the charasteristic function of f.

Woodroofe (1970) uses a two stage procedure to estimate h_n. The first step is to estimate f and its derivatives using $\hat{f}_n(x)$ with a h_n sequence satisfying the usual conditions. An estimate of h_n, $\hat{h}_n$, is then obtained by finding the optimal $\hat{h}_n$ sequence with respect to the density $\hat{f}_n(x)$. We can then estimate f by replacing h_n with $\hat{h}_n$ in $\hat{f}_n(x)$. Under appropriate regularity conditions, Woodroofe then proves that if h_n is the optimal sequence with respect to the true density f(x) then

(1.3.1.) $$E\left(\hat{f}_n(x\ ;\ \hat{h}_n) - f(x)\right)^2 \sim E\left(\hat{f}_n(x\ ;\ h_n) - f(x)\right)^2$$

as $n \to \infty$. Unfortuneately the estimates $\hat{h}_n$ still depend on arbitrary sequences "b_n" and "$t_{n,i}$" introduced by Woodroofe. Of these he says "... the determination of the b_n and $t_{n,i}$ sequences ... is not as crucial as that of the h_n sequence. The former affects only the rate of corvergence in (1.3.1) while the latter affects the rate of mean square consistency".

From the discussion above we conclude that the choice of the sequence h_n using only the observations is still an open problem. We will conjecture a possible solution in chapter 5.

1.4. Summary of Results:

Epanechnikov discusses the determination of the sample size to assure a prescribed level for the minimum relative global error. However, this discussion entails assuming the form of f(x), an assumption that will usually not be justified. For this reason we felt a new approach to be necessary. One approach that is often employed in determining a sample size necessary to satisfy a given criterion, is that of sequential methods. Thus in order to obtain the sample size required to use the estimator $\hat{f}_n(x)$ and satisfy some given criterion, we will use a sequential method. We feel that the results obtained in this dissertation bear out the appropriateness of the approach.

In chapter two we use the estimator $\hat{f}_n(x)$ to define a function $V_n(x) = \hat{f}_{nM}(x) - \hat{f}_{(n-1)M}(x)$. We define a "naive" sequential procedure based on this function $V_n(x)$ and constant ε. The asymptotic structure of the function $V_n(x)$ is studied and properties of the stopping rule, N, are obtained. We show that under appropriate conditions, $P[N > nM]$ has an upper bound which decreases exponentially to zero as $n\to\infty$ and that EN^r is finite for r finite.

In chapter 3 we consider the properties on an estimator $f_{N_\varepsilon}(x)$, where N_ε is a stopping variable. As few assumptions as possible are made as to the form of either $f_n(x)$ or the stopping rule N_ε. We show that if $N_\varepsilon \to \infty$ in probability as $\varepsilon \to 0$ then provided $f_n(x)$ satisfies appropriate conditions, $Ef_{N_\varepsilon}(x) \to f(x)$ as $\varepsilon \to 0$. It is also shown that under certain conditions $E[f_{N_\varepsilon}(x) - Ef_{N_\varepsilon}(x)]^2 \to 0$ as $\varepsilon \to 0$ and $E[f_{N_\varepsilon}(x) - f(x)]^2 \to 0$ as $\varepsilon \to 0$. We also show that if in fact $N_\varepsilon \to \infty$ a.s. as $\varepsilon \to 0$ and $f_n(x) \to f(x)$ a.s. as $n \to \infty$, then $f_{N_\varepsilon}(x) \to f(x)$ a.s. as $\varepsilon \to 0$.

In chapter four we show that $\tilde{f}_n(x)$ is both pointwise and uniformly strongly consistent. We use these results to parallel the results obtained in chapter 2, using $\tilde{f}_n(x)$ instead of $\hat{f}_n(x)$.

CHAPTER 2

A SEQUENTIAL PROCEDURE.

2.1. Introduction.

Under appropriate conditions, we have seen that the density estimate

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

is asymptotically consistent. We now show how this property may be utilised to define a "naive" sequential procedure. Basically the procedure will consist of taking successive samples of size M consisting of M mutually independent and identically distributed random variables and defining the difference as

$$V_n(x) = \hat{f}_{nM}(x) - \hat{f}_{(n-1)M}(x) \tag{2.1.1}$$

where $\hat{f}_{nM}(x)$ and $\hat{f}_{(n-1)M}(x)$ are the density estimates based on sample sizes of nM and (n-1)M respectively. The stopping rule we will consider is then of the form

$$N(\varepsilon,M) = \begin{cases} \text{First } n \text{ such that } |V_n(x)|<\varepsilon \text{ for given } \varepsilon>0 \\ \infty \text{ if no such } n \text{ exists.} \end{cases} \tag{2.1.2}$$

In this chapter we discuss the properties of the stopping rule and some of the properties of the function $V_n(x)$ upon which the stopping rule depends. We examine the asymptotic structure of the function $V_n(x)$ and prove that both the expected sample size and the variance of N are finite. In fact we prove that under appropriate conditions all moments of the distribution of N are finite and that $P[N>nM]$ has an exponential upper bound which tends to zero as $n\to\infty$. Further, we will show that $N(\varepsilon,M) \to \infty$ as $\varepsilon\to 0$ in both probability and with probability one and that $N(\varepsilon,M)$ as defined by (2.1.2) is a "closed" stopping rule, by which we will mean that

$$P[N<\infty] = 1.$$

Finally, we will redefine $N(\varepsilon,M)$ so that the class of kernels for which $N(\varepsilon,M) \to \infty$ as $\varepsilon\to 0$ is considerably enlarged.

In order to obtain the results in this chapter we will make the following assumptions concerning the kernel function and the h_n sequence. The kernel, $K(u)$, will be assumed to satisfy:

(2.1.3)
(i) $K(u)$ is a density on R
(ii) $\sup_{u\in R} K(u) < \infty$
(iii) $\lim_{u\to\infty} |u|\, K(u) \to 0$

The sequence $\{h_n\}$, $n = 1,2,\cdots$ satisfies

(2.1.4)
(i) $h_n > 0$ for all n
(ii) $\lim_{n\to\infty} h_n = 0$
(iii) $\lim_{n\to\infty} nh_n = \infty$

and

$$(2.1.5)\qquad \lim_{n\to\infty} \frac{h_{(n+1)M}}{h_{nM}} = 1$$

We note that conditions (2.1.3) and conditions (2.1.4) (i) and (ii) are necessary for obtaining the asymptotic unbiasedness of $\hat{f}_n(x)$, and these together with (2.1.4) (iii) are necessary for obtaining consistency.

2.2. Properties.

Lemma 2.2.1 gives some elementary properties of $V_n(x)$ which allow its use in the stopping rule (2.1.2).

Lemma 2.2.1.

As $n\to\infty$, if K(u) and $\{h_n\}$, $n = 1,2,\cdots$ satisfy (2.1.3) and (2.1.4) then

(a) $|V_n(x)| \to 0$ in probability if x is a continuity point of f(x), and

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

Proof: (a) By definition

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

Under the conditions stated $\{\hat{f}_n(x)\}$ is a Cauchy sequence in probability since $\hat{f}_n(x) \to f(x)$ in probability and so the result follows.

(b)
$$\sup_x |V_n(x)| = \sup_x |\hat{f}_{nM}(x) - \hat{f}_{(n-1)M}(x)|$$
$$\leq \sup_x|\hat{f}_{nM}(x) - f(x)| + \sup_x|\hat{f}_{(n-1)M}(x) - f(x)|$$

Under the conditions stated each term on the right hand side can be made arbitrarily small by making n large enough and so the result follows.

Lemma 2.2.2

If K(u) and $\{h_n\}$, $n = 1,2,\cdots$ are as in Lemma 2.2.1, and in addition condition 2.1.5. and

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

(ii) $\int_{-\infty}^{\infty} \{K(cu) - K(u)\}^2 \, du$ is locally Lipschitz of order α at $c = 1$ for some $\alpha>0$,

(iii) $\sum_{n=1}^{\infty} \frac{1}{nh_n^{1-\beta}} \left| \frac{1}{h_{n+1}} - \frac{1}{h_n} \right|^{\beta} < \infty$, where $\beta = \min \{^1/_2\alpha, 1\}$

then $|V_n(x)| \to 0$ with probability one if x is a continuity point of f(x).

Proof:

The conditions stated are sufficient to ensure $\{\hat{f}_n(x)\}$ is a Cauchy sequence (a.s) and the result follows immediately.

Parzen (1962) shows his estimator is asymptotically normal. While we cannot show $V_n(x)$ is also asymptotically normal we will show that it may be written as the sum of two independent random variables, one of which does have an asymptotic normal distribution. We develop this result with the sequence of lemmas 2.2.3 and 2.2.4.

Lemma 2.2.3

If K(y) is a piecewise continuous Borel function satisfying the conditions (2.1.3) and g is a real function satisfying

$$\int_{-\infty}^{\infty} |g(y)| \, dy < \infty$$

and if $\{h_n\}$ is a sequence of positive constants satisfying the conditions (2.1.4) then

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

converges to

$$g(x) \int_{-\infty}^{\infty} K^2(y)\, dy$$

at every point of continuity of $g(\cdot)$.

Proof:

The proof is in two stages. First we will show that

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

and then that

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

Proof of (A): Using the definition of $g_n(x)$,

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

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

We now split the range of integration into $|y|>\delta$ and $|y|\le\delta$ where δ is an arbitrary, positive number.

Thus, the last quantity is less than

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

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

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

We now make the transformation, $z = \frac{y}{h_n}$ in the first and third terms so that we can write (2.2.1) as

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

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

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

which in turn is less than

(2.2.2)
$$\max_{|y|\le\delta} |g(x-y) - g(x)| \int_{-\infty}^{\infty} K(z)\, K\left(\frac{h_n}{h_{n-1}} z\right) dz$$

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

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

But, by Schwarz's inequality we obtain

$$\left(\int_{-\infty}^{\infty} \left|K(z)\, K\left(\frac{h_n}{h_{n-1}} z\right)\right| dz\right)^2 \le \int_{-\infty}^{\infty} K^2(z)dz \int_{-\infty}^{\infty} K^2\left(\frac{h_n}{h_{n-1}} z\right) dz$$

$$= \frac{h_{n-1}}{h_n} \left(\int_{-\infty}^{\infty} K^2(z)dz\right)^2$$

$< \infty$. by assumption.

Since δ is arbitrary, the first term can be made arbitrarily small by making δ arbitrarily small. Then, letting $n\to\infty$, the second and third terms in (2.2.2) tend to zero.

Thus the proof of (A) is complete.

Proof of (B): Consider

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

On letting $z = \frac{y}{h_n}$ this becomes

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

Let

$$|\ell_n(z)| = \left| K\left(\frac{h_n}{h_{n-1}} = z\right) - K(z)\right| K(z)$$

$\leq \left(2 \sup_z | K(z)|\right) K(z)$ which by condition (2.1.3) (ii)

is finite since $\sup_z K(z) < \infty$. Thus

$|\ell_n(y)| \leq C.\, K(z)$ where C is a constant and so $\ell_n(y) \in L_1$

since

$\int_{-\infty}^{\infty} |K(y)|\, dy < \infty$. But, since K(y) is assumed piecewise

continuous,

$|\ell_n(y)| \to 0$ as $n \to \infty$.

Thus, by the dominated convergence theorem,

$$\lim_{n\to\infty} \int_{\infty}^{\infty} |\ell_n(y)|\, dy = 0$$

and the proof of (B) is completed.

The conclusion of the lemma then follows.

Lemma 2.2.4.

Let $V_n(x)$ be defined by (2.1.1), then if K(u) and $\{h_n\}$ satisfy

(2.1.3), (2.1.4) and (2.1.5) and if in addition

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

then

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

Proof:

From the definition of $\hat{f}_n(x)$ we have

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

Since $X_1, X_2, \cdots, X_{nM}$ are independently and identically distributed

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

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

Parzen (1962) shows that

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

Using this with lemma 2.2.3 we have from (2.2.4)

$$\lim_{n\to\infty} n^2 M h_{nM} \operatorname{Var} V_n(x) = \lim_{n\to\infty} \left(\frac{n}{h_{nM}} \operatorname{Var} K\left(\frac{x-X_1}{h_{nM}}\right) \right.$$

$$+ \frac{n^2}{(n-1)} \frac{h_{nM}}{h_{(n-1)M}^2} \operatorname{Var} K\left(\frac{x-X_1}{h_{(n-1)M}}\right)$$

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

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

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

$= \nu.\ f(x) \int_{\infty}^{\infty} K^2(u)\, du.$ using our assumptions (2.1.5) and (2.2.3).

This proves the lemma.

Lemma 2.2.5.

$V_n(x)$ can be written as the sum of independent random variables, that is

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

such that under the conditions of lemma 2.2.4

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

and

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

Proof: We can write

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

(2.2.4A)

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

$$= A_n(x) + B_n(x) \quad \text{say.}$$

Since $A_n(x)$ depends only on $X_1, X_2, \cdots, X_{(n-1)M}$ and $B_n(x)$ depends only on $X_{(n-1)M+1}, \cdots, X_{nM}$; $A_n(x)$ and $B_n(x)$ are independent. Thus

$$\text{Var } V_n(x) = \text{Var } A_n(x) + \text{Var } B_n(x).$$

But

$$\text{Var } B_n(x) = \frac{1}{n^2 M^2 h_{nM}^2} M \ \text{Var } K\left(\frac{x-X_1}{h_{nM}}\right)$$

so that

$$\lim_{n\to\infty} n^2 M\, h_{nM} \text{ Var } B_n(x) = \lim_{n\to\infty} \frac{1}{h_{nM}} \ \text{Var } K\left(\frac{x-X_1}{h_{nM}}\right)$$

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

But, since $A_n(x)$ and $B_n(x)$ are independent we can now write

$$\text{Var } A_n(x) = \text{Var } V_n(x) - \text{Var } B_n(x)$$

so that

$$\lim_{n\to\infty} n^2 M\, h_{nM} \text{ Var } A_n(x) = \lim_{n\to\infty} n^2 M\, h_{nM} \text{ Var } V_n(x) - \lim_{n\to\infty} n^2 M\, h_{nM} \text{ Var } B_n(x)$$

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

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

and the lemma is proved.

Remarks:

(1) $\nu \geq 1$ so that $-1 + \nu \geq 0$.

(2) The quantity $B_n(x)$ is a finite sum depending only on the final sample which consists of M independent random variables. Further $B_n(x)$ is always positive since we assume the kernel, $K(u)$, is always positive.

(3) If the density function $f(x)$ were known then under appropriate conditions on the kernel function, $K(u)$, the exact distribution of $B_n(x)$ could be found. It would depend on n and would be an M - fold convolution of densities of functions of the form

$$z_k = \frac{1}{nMh_{nM}} K\left(\frac{x-X_k}{h_{nM}}\right) \qquad (n-1)M + 1 \le k \le nM$$

(4) The quantity $A_n(x)$ depends on the first $(n-1)M$ observations and we will show it has an asymptotic normal distribution.

Example 2.2.1.

We will now give an example of a $\{h_n\}$ sequence which satisfies the condition (2.2.3). Let

$$h_n = B\, n^{-\alpha} \qquad , \qquad 0 < \alpha < 1$$

where B is a constant independent of n, then (2.2.3) gives

$$\text{(2.2.5)} \qquad n\left(\frac{h_{nM}}{h_{(n-1)M}} - 1\right) = n\left(\left(\frac{n-1}{n}\right)^{\alpha} - 1\right)$$

$$= n\left(\left(1 - \frac{\alpha}{n} + o\left(\frac{1}{n}\right) \right) - 1\right)$$

$$= -\alpha + o(1)$$

Thus, taking the limit as $n\to\infty$ we have

$$1 - \nu = -\alpha.$$

That is for the sequence $h_n = B\, n^{-\alpha}$, $0 < \alpha < 1$, we have that

$$\nu = 1 + \alpha \text{ and } \lim_{n\to\infty} n^2 M\, h_{nM}\, A_n(x) = \alpha\, f(x) \int_{-\infty}^{\infty} K^2(u)\, du.$$

Lemma 2.2.6.

If $K(u)$ and $\{h_n\}$ satisfy (2.1.3) and (2.1.4) then

$$B_n(x) \le \frac{M}{nMh_{nM}} \sup_u K(u) \to 0 \text{ as } n\to\infty$$

Proof: By the definition of $B_n(x)$

$$B_n(x) = \frac{1}{nM} \sum_{j=(n-1)M+1}^{nM} \frac{1}{h_{nM}} K\left(\frac{x-X_1}{h_{nM}}\right)$$

$$\le \frac{1}{nMh_{nM}} \sum_{j=(n-1)M+1}^{nM} \sup_u K(u)$$

$$= \frac{M}{nMh_{nM}} \sup_u K(u)$$

which is finite, since by assumption $\sup_u K(u)<\infty$. Clearly, since $nMh_{nM}\to\infty$ as $n\to\infty$ the bound tends to zero as $n\to\infty$.

We will now show that $A_n(x)$ is asymptotically normally distributed. To do this we first prove two lemmas, in which we will require the definition,

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

Lemma 2.2.7.

If $K(u)$ and $\{h_n\}$ satisfy (2.1.3) and (2.1.4) then if

$$\lim_{n\to\infty} \left(\frac{(n-1)^2}{n} + n \frac{h_{nM}}{h_{(n-1)M}} - 2(n-1) \frac{h_{nM}}{h_{(n-1)M}}\right) = \nu_0$$

$$\lim_{n\to\infty} n\, M\, h_{nM} \operatorname{Var} A_{n1}(x) = \nu_0\, f(x) \int_{-\infty}^{\infty} K^2(u)\, du$$

Proof: The proof exactly parallels that of lemma 2.2.4. and so will be omitted.

Example 2.2.2.

If $h_n = B\, n^{-\alpha}$, $0 < \alpha < 1$

then ν_0 as defined in Lemma 2.2.7 equals α, that is $\nu_0 = \alpha$.

Lemma 2.2.8.

If K(u) and $\{h_n\}$ satisfy (2.1.3), (2.1.4) and (2.1.5) and if

$$\int_{-\infty}^{\infty} |K^3(u)\, du < \infty, \text{ and}$$

$$\lim_{n\to\infty} \left((n-1) - \frac{n\, h_{nM}}{h_{(n-1)M}}\right) = \nu_1$$, where ν_1 is a constant and,

if $\{C_n\}$ is a sequence such that $C_n \to 1$ as $n\to\infty$ and $K(C_n u) \to K(u)$ uniformly

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

Proof: By definition

$$E|A_{n1}(x)|^3 = \int_{-\infty}^{\infty} \left|\frac{(n-1)}{n\, h_{nM}}\, K\left(\frac{x-y}{h_{nM}}\right) - \frac{1}{h_{(n-1)M}}\, K\left(\frac{x-y}{h_{(n-1)M}}\right)\right|^3 f(y)\, dy$$

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

Thus multiplying both sides by $n^2 h_{nM}^2$ and making the transformation on the right hand side of $u = \frac{x-y}{h_{nM}}$ we obtain

$n^2 h_{nM}^2\, E\, A_{n1}(x)$

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

But,

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

$$= \nu_1 K(u) \quad \text{by assumption.}$$

Thus there exists a $\delta = \delta(n)$ such that for $n > n_0$,

$$\left| \left| (n-1) K(u) - \frac{nh_{nM}}{h_{(n-1)M}} K\left(\frac{h_{nM}}{h_{(n-1)M}} u\right) \right|^3 - \nu_1^3 |K(u)|^3 \right| \le \delta$$

Thus for $n > n_0$,

$$n^2 h_{nM}^2 E|A_{n1}(x)|^3 \le \frac{\nu_1^3}{n} \int_{-\infty}^{\infty} |K(u)|^3 f\left(x-h_{nM} u\right) du$$

$$+ \frac{\delta}{n} \int_{-\infty}^{\infty} f\left(x-h_{nM} u\right) du$$

But,

$$\frac{1}{n} \int_{-\infty}^{\infty} f\left(x-h_{nM} u\right) du \to 0 \text{ as } n\to\infty,$$

so that,

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

To prove the asymptotic normality of $A_{n1}(x)$ we will now use a theorem from Loève. The theorem is stated here without proof.

Theorem 2.2.9. (Normal Convergence Criterion - Loève (1960) p 316)

If X_{nk} are independent summands, then for every $\varepsilon>0$,

$$P\left(\frac{\sum_k X_{nk} - \mu}{\sigma} < c\right) \to \Phi(c)$$

and,

$$\max_k P\left(|X_{nk}| \ge \varepsilon\right) \to 0$$

if and only if, for every $\varepsilon>0$ and a $\tau>0$,

(i) $\sum_k P\left(|X_{nk}| \ge \varepsilon\right) \to 0$

(ii) $\sum_k \sigma_{nk}^2 (\tau) \to \sigma^2$

where $$\mu_{nk}(\tau) = \int_{|x|<\tau} x\,dF_{nk}(x)$$

$$\sigma^2_{nk}(\tau) = \int_{|x|<\tau} x^2\,dF_{nk}(x) - \left(\int_{|x|<\tau} x\,dF_{nk}(x)\right)^2$$

Using lemmas 2.2.7 and 2.2.8 together with Theorem 2.2.9 we now prove the asymptotic normality of $A_n(x)$.

Theorem 2.2.10.

If $\int_{-\infty}^{\infty} |K(u)|^3\,du < \infty$ and conditions (2.1.3), (2.1.4) and (2.1.5) are satisfied, then provided K(u) is piecewise continuous

$$(2.2.6) \quad \lim_{n\to\infty} P\left(\frac{A_n(x) - E\,A_n(x)}{\left(\mathrm{Var}\,A_n(x)\right)^{1/2}} \quad c\right) = \int_{-\infty}^{c} \frac{1}{\sqrt{2\pi}} e^{-\,^1/_2 u^2}\,du = \Phi(c)$$

Proof: By theorem 2.2.9 a necessary and sufficient condition for (2.2.6) to hold is that

$$(2.2.7) \qquad (n-1)M.P\left(\left|\frac{A_{n1}(x) - E\,A_{n1}(x)}{\left(\mathrm{Var}\,A_{n1}(x)\right)^{1/2}}\right| \geq \varepsilon \left((n-1)M\right)^{1/2}\right) \to 0$$

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

A sufficient condition (Liapounov's condition) for (2.2.7) to hold is that for some $\delta>0$,

$$(2.2.8) \qquad \frac{E\left|A_{n1}(x) - E\left(A_{n1}(x)\right)\right|^{2+\delta}}{(nM)^{\delta/2}\,\sigma^{2+\delta}\left(A_{n1}(x)\right)} \to 0 \quad \text{as } n\to\infty$$

where $\sigma^2\left(A_{n1}(x)\right)$ denotes the variance of $A_{n1}(x)$.

We will consider $\delta = 1$.

Since $(a + b)^3 \quad 4(a^3 + b^3)$ we obtain

$$E\left|A_{n1}(x) + EA_{n1}(x)\right|^3 \le E\left(\left|A_{n1}(x)\right| + \left|EA_{n1}(x)\right|\right)^3$$

$$\le 4\left(E\left|A_{n1}(x)\right|^3 + \left|EA_{n1}(x)\right|^3\right)$$

But,

$$\left|EA_{n1}(x)\right| \le E\left|A_{n1}(x)\right|$$

so that

$$(2.2.9) \qquad E\left|A_{n1}(x) + EA_{n1}(x)\right|^3 \le 8\, E\left|A_{n1}(x)\right|^3.$$

Substituting (2.2.9) in (2.2.8) we then get

$$(2.2.10) \qquad \frac{E\left|A_{n1}(x) - E\left(A_{n1}(x)\right)\right|^3}{(nM)^{1/2}\sigma^3\left(A_{n1}(x)\right)} \le \frac{8\, E\left|A_{n1}(x)\right|^3}{(nM)^{1/2}\sigma^3\left(A_{n1}(x)\right)}$$

But, from lemmas 2.2.7 and 2.2.8 we have that

$$E\left|A_{n1}(x)\right|^3 = o\left(\frac{1}{n^2 h_{nM}^2}\right)$$

and

$$\sigma^2\left(A_{n1}(x)\right) = O\left(\frac{1}{nh_{nM}}\right)$$

so that (2.2.10) is of the order, $o\left(\frac{1}{(nh_{nM})^{1/2}}\right)$. That is

$$\frac{E\left|A_{n1}(x) - E\, A_{n1}(x)\right|^3}{(nM)^{1/2}\, \sigma^3\left(A_{n1}(x)\right)} \to 0 \quad \text{as } n\to\infty.$$

This completes the proof of the Theorem.

We have now shown that $V_n(x) = A_n(x) + B_n(x)$ consists of one component, $A_n(x)$, that has an asymptotic normal distribution and another component, $B_n(x)$ which is bounded above by a bound which tends to zero as $n\to\infty$. Notice, however, that the variances of the two components are of the same order, $\left\{O\left(\frac{1}{n^2 h_{nM}}\right)\right\}$, for estimators of the Parzen type.

2.3 Finiteness of EN:

A property that is desirable for any stopping variable is that the expected sample size, EN, should be finite. If we assume no knowledge of the density function, f(x), we will in general not be able to determine the exact value of EN. Indeed, if f(x) is known it is likely that we can obtain an exact expression for EN in only a relatively small number of cases. For this reason we will confine ourselves to proving in this section that with our "naive" stopping rule the expected sample size is finite given certain restrictions on the kernel. To do this we prove a sequence of lemmas.

Lemma 2.3.1

For arbitrary $t>0$, and given $\varepsilon>0$,

$$P\left\{V_n(x) > \varepsilon\right\} \le e^{-nMh_{nM}\varepsilon t}\, Ee^{S_n(x)t} \tag{2.3.1}$$

where

$$S_n(x) = \sum_{j=1}^{nM} K\left(\frac{x-X_j}{h_{nM}}\right) - \sum_{j=1}^{(n-1)M} \frac{n}{n-1}\,\frac{h_{nM}}{h_{(n-1)M}}\, K\left(\frac{x-X_j}{h_{(n-1)M}}\right) \tag{2.3.2}$$

Proof:

From the definition of $V_n(x)$, we have

$$P\left\{V_n(x) > \varepsilon\right\} = P\left\{S_n(x) > nMh_{nM}\,\varepsilon\right\}$$

Define, (using a method described in Hoeffding 1963)

$$T(x) = \begin{cases} 1 & \text{if } S_n(x) - nMh_{nM}\,\varepsilon > 0 \\ 0 & \text{otherwise} \end{cases}$$

But for arbitrary $t>0$,

$$T(x) \le e^{t(S_n(x) - nMh_{nM}\,\varepsilon)}$$

so that

$$P\left(S_n(x) > nMh_{nM}\,\varepsilon\right) \le Ee^{t(S_n(x) - nMh_{nM}\,\varepsilon)}$$

$$= e^{-nMh_{nM}\,\varepsilon t}\, Ee^{tS_n(x)} .$$

Lemma 2.3.2.

For arbitrary $t>0$ and given $\varepsilon>0$

(2.3.3) $$P\left(V_n(x) < -\varepsilon\right) \quad e^{-nMh_{nM}\varepsilon t}\, Ee^{-tS_n(x)}$$

where $S_n(x)$ is given by (3.4.2)

Proof:

The proof follows as for lemma 3.4.1 if we note

$$P\left(V_n(x) < -\varepsilon\right) = P\left(S_n(x) < -nMh_{nM}\,\varepsilon\right)$$

and define

$$T'(x) = \begin{cases} 1 & \text{if } S_n(x) + nMh_{nM}\,\varepsilon < 0 \\ 0 & \text{otherwise} \end{cases}$$

To see the relevance of Lemmas 2.3.1 and 2.3.2 we will now look at $P(N>nM)$ and obtain a simple inequality concerning this quantity.

Define

(2.3.4) $$\Lambda_k = \left\{(X_1, X_2, \cdots, X_{nM}) : |V_k(x) > \varepsilon\right\}, \; 2 \le k \le n.$$

Then, from definition (2.1.2) we have that for $n \ge 2$,

$$P\left(N>nM\right) = P\left(\bigcap_{k=2}^{n} \Lambda_k\right)$$

$$\leq P\left(\Lambda_n\right)$$

$$= P\left(|V_n(x)| > \varepsilon\right)$$

Thus, since we can write

$$P\left(|V_n(x)| > \varepsilon\right) = P\left(V_n(x) < -\varepsilon\right) + P\left(V_n(x) > \varepsilon\right)$$

we have that

(2.3.5) $$P\left(N>nM\right) \leq P\left(V_n(x) > \varepsilon\right) + P\left(V_n(x) < -\varepsilon\right)$$

Further, we note that

(2.3.6) $$EN = \sum_{k=0}^{\infty} P\left(N>kM\right)$$

where we will assume $P\left(N>0\right) = P\left(N>M\right) = 1$. (This is consistent with the way we define our stopping rule since we need to take at least two samples of size M to define $V_n(x)$). Thus, provided we can show the convergence of the right hand side of (2.3.6), we have reached our goal. To do this we use equation (2.3.6), together with Lemmas (2.3.1) and (2.3.2). Some further results concerning $Ee^{tS_n(x)}$ must now be obtained however, before the latter two lemmas will be useful.

Lemma 2.3.3

If $\{h_n\}$ is a sequence of positive constants satisfying (2.1.4) and such that there exists an $N_0 = N_0(\varepsilon)$ for which if $n > N_0(\varepsilon)$,

(A1) $$e^{-nMh_{nM}\varepsilon} \leq \frac{1}{(nM)^{\gamma + (1+\delta)}} \quad ,\delta > 0$$

and if

$$Ee^{\pm S_n(x)} \leq C.\ n^{\gamma}$$

where both C and γ are finite positive constants, then

$$EN < \infty.$$

Note 1: Since lemmas 3.3.1 and 3.3.2 are true for arbitrary t we will only consider the case t = 1 in the remainder of this section.

Note 2: γ can be negative in the statement of this lemma, however, we will see that to obtain meaningful results in later lemmas and theorems (in particular theorem 2.3.6) we require $\gamma \geq 0$.

Proof of Lemma 2.3.3.

It will be sufficient to prove that

$$\sum_{n=N_0}^{\infty} P\left(N>nM\right) < \infty$$

if the above conditions are satisfied since this represents the tail of an infinite series for EN. Now from (2.3.5)

$$\sum_{n=N_0}^{\infty} P\left(N>nM\right) \leq \sum_{n=N_0}^{\infty} e^{-nMh_{nM}\varepsilon}\left(Ee^{S_n(x)} + Ee^{-S_n(x)}\right)$$

$$\leq 2C \sum_{n=N_0}^{\infty} n^{\gamma}\, e^{-nMh_{nM}\varepsilon}.$$

But, since for $n \geq N_0$ we assume that (A1) holds we have that

$$\sum_{n=N}^{\infty} P\left(N>nM\right) \leq \frac{2C}{M^{\gamma}} \sum_{n=N_0}^{\infty} \frac{1}{(nM)^{1+\delta}} < \infty.$$

Thus, since the tail of the series (2.3.6) is convergent, EN < ∞ as required.

Example 2.3.1: Consider

$$h_n = B\,n^{-\alpha} \qquad 0 < \alpha < 1, \qquad B \text{ a constant.}$$

If condition (A1) is to hold then for $n \geq N_0$

$$-nMh_{nM}\varepsilon \leq -\big(\gamma + (1+\delta)\big) \log nM$$

so that

$$\frac{nMh_{nM}}{\log nM} \geq \frac{\gamma + (1+\delta)}{\varepsilon}$$

This means that nMh_{nM} must increase faster than log nM. But

$$\frac{(nM)^{1-\alpha}}{\log nM} \to \infty \quad \text{as } n\to\infty$$

so that the condition will hold for the sequence given.

We now turn our attention to $Ee^{S_n(x)}$. Notice first that we can write

$$S_n(x) = \sum_{k=1}^{(n-1)M} A_n(x,X_k) + \sum_{k=(n-1)M+1}^{nM} B_n(x,X_k)$$

where

$$A_n(x,X_k) = K\left(\frac{x-X_k}{h_{nM}}\right) - \frac{n}{(n-1)} \frac{h_{nM}}{h_{(n-1)M}} K\left(\frac{x-X_k}{h_{(n-1)M}}\right), \quad k = 1, 2, \cdots, (n-1)M$$

$$B_n(x,X_k) = K\left(\frac{x-X_k}{h_{nM}}\right) \qquad k = (n-1)M+1, \cdots, nM$$

But $A_n(x,X_k)$ depends only on X_k and $B_n(x,X_\ell)$ depends only on X_ℓ so that we have written $S_n(x)$ as the sum of nM mutually independent random variables. Further $A_n(x,X_1), \cdots, A_n(x,X_{(n-1)M})$ are identically distributed random variables, as are $B_n(x,X_{(n-1)M+1}), \cdots, B(x,X_{nM})$. This means we may write

$$Ee^{S_n(x)} = Ee^{\sum_{k=1}^{(n-1)M} A_n(x,X_k) + \sum_{k=(n-1)M+1}^{nM} B_n(x,X_k)}$$

$$= Ee^{\sum_{k=1}^{(n-1)M} A_n(x,X_k)} \; Ee^{\sum_{k=(n-1)M+1}^{nM} B_n(x,X_k)}$$ (as the sums are

independent), so that

$$(2.3.7) \qquad Ee^{S_n(x)} = \left(Ee^{A_n(x,X_1)}\right)^{(n-1)M} \left(Ee^{B_n(x,X_{nM})}\right)^{M}$$

We now proceed to show that the second term in the product on the right hand side is bounded above.

Lemma 2.3.4

If K(u) satisfies (2.1.3) then

$$\left(Ee^{B_n(x,X_{nM})}\right)^M \le e^{ML}$$

where
$$L = \sup_{-\infty < u < \infty} K(u) \quad (< \infty \quad \text{by } (2.1.3))$$

Proof: By definition

$$Ee^{B_n(x,X_{nM})} = \int_{-\infty}^{\infty} e^{K\left(\frac{x-u}{h_{nM}}\right)} f(u)\, du$$

$$\le e^L \int_{-\infty}^{\infty} f(u)\, du$$

$$= e^L$$

and the result is immediate.

Before considering the first term in the product (2.3.7) we now examine "how large" $S_n(x)$ can be and still allow us to retain $EN < \infty$.

Lemma 2.3.5

If there exists an N_0 such that for $n > N_0$

$$|S_n(x)| \le c \log n \qquad \text{(a.s)}$$

then

(2.3.8) $$P(N>nM) \le 2\, n^c\, e^{-nMh_{nM}\varepsilon}.$$

Further, if

$$|S_n(x)| \le c \qquad \text{(a.s)}$$

then this inequality can be improved to

(2.3.9) $$P(N>nM) \le c\, e^{-nMh_{nM}\varepsilon}.$$

If either (2.3.8) or (2.3.9) holds and if in addition $\{h_n\}$ satisfies condition (A1) then $EN < \infty$.

Proof: By definition

$$Ee^{|S_n(x)|} = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{|S_n|} f(x_1,\cdots, x_{nM})\, dx_1\cdots dx_{nM}$$

$$\le \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{c \log n} f(x_1,\cdots,x_{nM})\, dx_1\cdots dx_{nM}$$

$$= n^c \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1,\cdots,x_{nM})\, dx_1\cdots dx_{nM}$$

$$= n^c$$

But both,

$$Ee^{S_n} \le Ee^{|S_n|}$$

and

$$Ee^{-S_n} \le Ee^{|S_n|}$$

Thus from lemmas (2.3.1) and (2.3.2) and using inequality (2.3.5) we get,

$$P\left(N>nM\right) \le 2\, n^{c}\, e^{-nMh_{nM}\varepsilon}, \qquad n > N_0.$$

Hence, by lemma 2.3.3 it follows that $EN < \infty$.

Using a similar argument we see that if $|S_n(x)| \le c$ then

$$Ee^{|S_n(x)|} \le e^{c} = \text{constant}$$

and lemma 2.3.3 again gives us that $EN < \infty$.

Example 2.3.2

The conditions of lemma 2.3.5 are satisfied by the normal kernel and the "usual" $\{h_n\}$ sequence, that is

$$K(u) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}u^2}$$

and

$$h_n = B\, n^{-\alpha} \qquad , \ 0 < \alpha < 1.$$

To show this we write (with B=1 for convenience)

$$n|A_n(x,u)| = \frac{n}{\sqrt{2\pi}} \left| e^{-\frac{(nM)^{2\alpha}}{2}(x-u)^2} - \left(\frac{n}{n-1}\right)^{1-\alpha} e^{-\frac{[(n-1)M]^{2\alpha}}{2}(x-u)^2} \right|$$

$$= \frac{n}{\sqrt{2\pi}} \left| \left(\frac{n}{n-1}\right)^{1-\alpha} e^{-\frac{[(n-1)M]^{2\alpha}}{2}(x-u)^2} - e^{-\frac{(nM)^{2\alpha}}{2}(x-u)^2} \right|$$

Differentiating this with respect to u we obtain

$$\frac{n}{\sqrt{2\pi}}\Big[(x-u)[(n-1)M]^{2\alpha}\left(\frac{n}{n-1}\right)^{1-\alpha} e^{-\frac{[(n-1)M]^{2\alpha}}{2}(x-u)^2} - (x-u)(Mn)^{2\alpha}\, e^{-\frac{(Mn)^{2\alpha}}{2}(x-u)^2}\Big]$$

which is zero when either $(x-u) = 0$ or

$$(x-u)^2 = \frac{2\ \log\left(\frac{n}{n-1}\right)^{3\alpha-1}}{(Mn)^{2\alpha} - [(n-1)M]^{2\alpha}} \quad .$$

The turning points represented by the latter equation are real only if $\alpha \geq \frac{1}{3}$ (and for $\alpha = \frac{1}{3}$, $u=x$). By noting that $n|A_n(x,u)|$ is continuous and that $\lim_{u\to\pm\infty} n|A_n(x,u)| = 0$ we see that for $0 < \alpha < \frac{1}{3}$, $n|A_n(x,u)|$ is always less than $\frac{n}{\sqrt{2\pi}} |1 - (\frac{n}{1-n})^{1-\alpha}|$, its maximum value at $u = x$. But $\lim_{n\to\infty} \frac{n}{\sqrt{2\pi}} |1 - (\frac{n}{1-n})^{1-\alpha}| = \frac{1-\alpha}{\sqrt{2\pi}}$ so that there exists a constant c_1 and an $n_1 = n_1(c_1)$ such that for $n > n_1$, $n|A_n(x,u)| \leq c_1$ for all u. However, $|S_n(x)| \leq n \sup_u |A_n(x,u)| \leq c_1$ for all u when $m > n_1$ and $0 < \alpha < \frac{1}{3}$.

For $\frac{1}{3} \leq \alpha < 1$ we have three real turning points. By differentiating $n|A_n(x,u)|$ a second time we can show that $n|A_n(x,u)|$ has a local minimum at $u = x$ and from the nature of the function we can conclude that $n|A_n(x,u)|$ is always less than

$$\frac{n}{\sqrt{2\pi}}[(\frac{n}{n-1})^{1-3\alpha}]^{1-((n-1)^{2\alpha}/(n^{2\alpha}-(n-1)^{2\alpha}))} [(\frac{n}{n-1})^{2\alpha}-1]$$ which tends to

$\frac{2}{\sqrt{2\pi}} e^{3\alpha-1}$ as $n \to \infty$. Thus, as above there exists a constant c_2 and an $n_2 = n_2(c)$ such that $|S_n(x)| \leq c_2$ for all u when $n > n_2$ and $\frac{1}{3} \leq \alpha < 1$. If we choose $c = \max(c_1,c_2)$ and $n_0=\max(n_1,n_2)$ then for $n>n_0$, $|S_n(x)| \leq c$ for all u, $0 < \alpha < 1$, and $EN < \infty$ in this case.

To this point we have obtained lemmas that give sufficient conditions for either $S_n(x)$ or $Ee^{S_n(x)}$ to imply $EN < \infty$. In practice however it is the kernel function K(u) and the sequence $\{h_n\}$ that we control. We now give a theorem that places a condition directly on the kernel to ensure $EN < \infty$.

Theorem 2.3.6:

For a piecewise continuous kernel, K(u), and a sequence of positive constants $\{h_n\}$ satisfying (2.1.3), (2.1.4) and (2.1.5) such that condition (A1) holds and if

$$(2.3.10)\qquad \lim_{n\to\infty} \frac{n}{\log n} \int_{-\infty}^{\infty} |A_n(x,u)|\ e^{|A_n(x,u)|}\ f(u)\ du = \gamma \geq 0.$$

then

$$EN < \infty.$$

Proof:

From equation (2.3.7) and using lemma (2.3.4) we may write

$$(2.3.11)\qquad Ee^{S_n(x)} \leq \left\{Ee^{A_n(x,X_1)}\right\}^{(n-1)M} e^{ML}$$

But,

$$e^{A_n(x,X_1)} \leq e^{|A_n(x,X_1)|} = 1 + |A_n(x,X_1)| + \frac{|A_n(x,X_1)|^2}{2!} + \cdots$$

$$\leq 1 + |A_n(x,X_1)|\ e^{|A_n(x,X_1)|}$$

so that

$$Ee^{A_n(x,X_1)} \leq 1 + a_n \qquad \text{(say)}$$

where

$$a_n = \int_{-\infty}^{\infty} |A_n(x,u)|\ e^{|A_n(x,u)|}\ f(u)\ du.$$

Now, it follows from assumption (2.3.10) that

$$(2.3.12)\qquad a_n = 0\left(\gamma\frac{\log n}{n}\right)$$

Thus,

$$(2.3.13)\qquad (1+a_n)^{n-1} \sim \left(1 + \gamma\frac{\log n}{n}\right)^{n-1}.$$

Now let

$$y = \left(1 + \gamma\frac{\log n}{n}\right)^{n} \text{ so that}$$

$$\log y = n \log \left(1 + \gamma\frac{\log n}{n}\right),$$

and since $\frac{\log n}{n} \to 0$ as $n \to \infty$, there exists an n_0 such that for $n \geq n_0$, $\gamma \frac{\log n}{n} < 1$ and

(2.3.14) $$\log y \leq n. \gamma \frac{\log n}{n} = \gamma \log n$$

Hence, asymptotically

$$(1+a_n)^{n-1} \leq O(n^\gamma)$$

and

$$\left(Ee^{A_n(x,X_1)}\right)^{n-1} \leq O(n^\gamma)$$

so that

$$Ee^{S_n(x)} \leq O(n^\gamma). \ e^{ML}.$$

Similarly we can show that

$$Ee^{-S_n(x)} \leq (1+a_n)^{(n-1)M} e^{ML}$$

$$\leq O(n^\gamma) \ e^{ML}.$$

It now follows from lemma 2.3.3 that $EN < \infty$.

Clearly condition (2.3.10) still involves the unknown density function, f(u). Theorem 2.3.6 by itself therefore does not completely solve the problem of putting a condition onK(u) directly. In lemma 2.3.7 we obtain a sufficient condition that (2.3.10) is satisfied for any density f(u), though example 2.3.2 will show the condition is not necessary if a weak condition is placed on f(u).

Lemma 2.3.7

If K(u) satisfies (2.1.3) and if $\{h_n\}$ satisfies the conditions of theorem (2.3.6) and if in addition

(2.3.15) $$\lim_{n\to\infty} \frac{n}{\log n} |A_n(x,u)| = \gamma < \infty \text{ uniformly in } u,$$

where γ is a constant independent of u, then condition (2.3.10) holds so that $EN < \infty$.

Proof:

By assumption (2.3.15), $|A_n(x,u)| \to 0$ uniformly in u. Thus given a $\delta > 0$, there exists an n_0 such that for $n \geq n_0$ both $|A_n(x,u)| \leq 1$ and $\frac{n}{\log n} |A_n(x,u)| \leq \gamma + \delta$.

But for $n > n_0$,

$$\frac{n}{\log n} |A_n(x,u)| \; e^{|A_n(x,u)|} f(u) \leq (\gamma+\delta) \; e^1 f(u) \; \varepsilon L_1$$

so that using the dominated convergence theorem

$$\lim_{n\to\infty} \frac{n}{\log n} \int_{-\infty}^{\infty} |A_n(x,u)| \; e^{|A_n(x,u)|} f(u) \, du = \int_{-\infty}^{\infty} \gamma \, e^0 f(u) \, du$$

$$= \gamma$$

and condition (2.3.10) holds.

Comment:

(1) It is not difficult to show that both the normal kernel and the double exponential kernel satisfy the condition in lemma 2.3.7 (with $\gamma = 0$).

(2) In example 2.3.3 we show that with the addition of a weak condition on f(u), the uniform kernel satisfies (2.3.10). However, the uniform kernel does not satisfy condition (2.3.15) since

$$\frac{n}{\log n} \frac{n}{(n-1)} \frac{h_{nM}}{h_{(n-1)M}} \to \infty \quad \text{as} \quad n\to\infty.$$

(3) We note that for the case of the uniform kernel condition (2.3.10) does answer our question since it will hold for any density f(u) that satisfies the condition given.

(4) Comments 2 and 3 considered together show that condition

(2.3.15) is sufficient to ensure $EN < \infty$ (with the other conditions of lemma 2.3.7) but it is not a necessary condition.

Example 2.3.3

Consider the uniform kernel defined by

$$K(u) = \begin{cases} \frac{1}{2} & \text{if } |u| \le 1 \\ 0 & \text{otherwise} \end{cases}$$

and again let $h_n = B\, n^{-\alpha}$, $0 < \alpha < 1$

Then, from the definition of a_n,

$a_n = W_1 + W_2 + W_3$,

where

$$W_1 = \int_{x-h_{nM}}^{x+h_{nM}} \frac{1}{2} \left|1 - \frac{n}{(n-1)} \frac{h_{nM}}{h_{(n-1)M}}\right| e^{\frac{1}{2}\left|1 - \frac{h_{nM}}{h_{(n-1)M}} \frac{n}{(n-1)}\right|} f(u)\, du$$

$$W_2 = \int_{x-h_{(n-1)M}}^{x-h_{nM}} \frac{1}{2} \frac{n}{(n-1)} \frac{h_{nM}}{h_{(n-1)M}} e^{\frac{1}{2} \frac{n}{(n-1)} \frac{h_{nM}}{h_{(n-1)M}}} f(u)\, du$$

$$W_3 = \int_{x+h_{nM}}^{x+h_{(n-1)M}} \frac{1}{2} \frac{n}{(n-1)} \frac{h_{nM}}{h_{(n-1)M}} e^{\frac{1}{2} \frac{n}{(n-1)} \frac{h_{nM}}{h_{(n-1)M}}} f(u)\, du.$$

For the purpose of this example we will make the additional assumption that there exists an N_0 such that for $n \ge N_0$,

$$\sup_u f(u) = C < \infty \quad \text{for} \quad x - h_{(n-1)M} \le u \le x + h_{(n-1)M}$$

Then,

$$\frac{n}{\log n} W_1 \le \frac{n}{\log n} \left(2h_{nM} \cdot \frac{1}{2} \left| 1 - \frac{n}{(n-1)} \left(\frac{n-1}{n}\right)^{\alpha} \right| e^{\frac{1}{2}\left|1 - \frac{n}{(n-1)} \left(\frac{n-1}{n}\right)^{\alpha}\right|}\right.$$

But, expanding $\frac{1}{1-\frac{1}{n}}$ and $(1-\frac{1}{n})^{\alpha}$ in powers of $\frac{1}{n}$ we get,

$$\frac{n^{-\alpha}}{\log n}\left|1-\frac{n}{n-1}\cdot\left(\frac{n-1}{n}\right)^{\alpha}\right| = \frac{n^{1-\alpha}}{\log n}\left|1-\left(1+\frac{1}{n}+o\left(\frac{1}{n}\right)\right)\left(1-\frac{\alpha}{n}+o\left(\frac{1}{n}\right)\right)\right|$$

$$= \frac{n^{1-\alpha}}{\log n}\left(\frac{1-\alpha}{n}+o\left(\frac{1}{n}\right)\right)$$

so that $\frac{n}{\log n} W_1$ tends to zero as $n\to\infty$.

We may consider W_2 and W_3 together since they are both bounded above by

$$C\left(h_{(n-1)M}-h_{nM}\right)\cdot\frac{1}{2}\ \frac{n}{(n-1)}\left(\frac{n-1}{n}\right)^{\alpha} e^{\frac{n}{2(n-1)}\left(\frac{n-1}{n}\right)^{\alpha}}$$

Now,

$$\frac{n}{\log n}\left(h_{(n-1)M}-h_{nM}\right) = \frac{B.n}{M^{\alpha}\log n}\left(\frac{1}{(n-1)^{\alpha}}-\frac{1}{n^{\alpha}}\right)$$

$$= \frac{Bn}{M^{\alpha}n^{\alpha}\log n}\left(\frac{1}{(1-\frac{1}{n})^{\alpha}}-1\right)$$

which equals, after some minor algebraic manipulation

$$\frac{Bn}{M^{\alpha}n^{\alpha}\log n}\left(\frac{\alpha}{n}+o\left(\frac{1}{n}\right)\right)$$

Thus $\frac{n}{\log n}(W_2+W_3)$ tends to zero as $n\to\infty$.

This means that

$$\lim_{n\to\infty}\frac{n}{\log n}a_n = o$$

and so $EN < \infty$ by theorem 2.3.6 when we use the uniform kernel and the sequence $h_n = B\,n^{-\alpha}$.

2.4 Variance of N:

When considering the distributions of a random variable we usually like to obtain its variance. As we noted for EN, unless f(x) is known, it is unlikely that we will be able to determine Var (N). However, through the next lemma we can at least see that the variance of N is finite.

Lemma 2.4.1

Under the same conditions as theorem 2.3.6 and provided h_n satisfies condition (A1) of lemma 2.3.3 then

$$\text{Var } N < \infty.$$

Proof:

Since in theorem 2.3.6 we have already shown $EN < \infty$, it will be sufficient to show $EN^2 < \infty$. Defining Λ_j as in equation (2.3.4) and Λ_j^c by

$$\Lambda_j^c = \left\{(x_1, \circ\circ\circ, x_{nM}): \ |V_j(x)| \le \varepsilon\right\} \tag{2.4.1}$$

we have from the definition of N that

$$P\left(N=nM\right) = P\left(\Lambda_n^c \cap . \left(\bigcap_{j=1}^{n-1} \Lambda_j\right)\right)$$

$$\le P\left(\Lambda_{n-1}\right) = P\left(|V_{n-1}(x)| > \varepsilon\right).$$

The conditions of Theorem 2.3.6 ensure that

$Ee^{|S_n|} \le n^c$, so that using lemmas 2.3.1 and 2.3.2

$$P\left(N=nM\right) \le 2(n-1)^c \, e^{-(n-1)M \, h_{(n-1)M}\varepsilon} \tag{2.4.2}$$

Now, since by condition (A1) we have that for $n > N_0$

$$e^{-(n-1)Mh_{(n-1)M}\varepsilon} \le \frac{1}{(nM)^{\gamma + (1+\delta)}} \quad , \quad \gamma \ge 0, \ \delta \ge 0 \tag{2.4.3}$$

Let us consider the tail of the series for EN^2 beyound the first $N_0 - 1$ terms. Then using the inequality (2.4.2)

$$\sum_{n=N_0}^{\infty} n^2 P(N=nM) \le \sum_{n=N_0}^{\infty} 2 n^2 (n-1)^c e^{-(n-1)Mh_{(n-1)M}\varepsilon}$$

and further, using (2.4.3) we get, putting $\gamma = 2 + c$, that

$$\sum_{n=N_0}^{\infty} n^2 P(N=nM) \le 2 \sum_{n=N_0}^{\infty} n^{c+2} \cdot \frac{1}{(nM)^{c+(1+\delta)+2}} = 2 \sum_{n=N}^{\infty} \frac{M^{-(c+2)-(1+\delta)}}{n^{1+\delta}} < \infty.$$

Thus since the tail of EN^2 is finite, EN^2 is finite and the proof is complete.

Observations:

(1) We have obtained an upper bound on P(N=nM) and we note that P(N=nM) → 0 as $n\to\infty$ at an exponential rate provided h_n satisfies appropriate conditions.

(2) Because of the form of the upper bound on P(N=nM) we can say that for finite r, $EN^r < \infty$.

To prove this we need only look at the tail of the series for EN^r and proceed as in lemma 2.4.1.

(3) Evidently, (using the information in (2)) we have that the distribution for the random variable N is a discrete probability distribution with all finite moments existing.

(4) Professor W. Hoeffding has pointed out the existence of a method for showing $Ee^{tN} < \infty$ for some $t > 0$ given a (suitable) exponential bound on P(N>nM). This implies $EN^r < \infty$ for all $r > 0$. See C. Stein, "A note on cumulative sums," Ann. Math. Statist. 17,(1946). 498 - 499. Unfortuneately, this method can not be applied directly to our problem as the bound on P(N>nM) is not of the right form.

2.5 Closure of the Stopping Rule:

By definition, the closure of a stopping rule is the property by which P(N<∞) = 1. (If a stopping rule is not closed it is called an extended stopping rule). The property of closure will now be shown to hold for our "naive" stopping rule.

Lemma 2.5.1

If the conditions of theorem 2.3.6 hold then the stopping rule defined by (2.1.2) is a closed stopping rule. That is, $P(N<\infty) = 1$.

Proof:

The stated conditions, are sufficient to ensure

$$P(N>nM) \leq C\, n^{\gamma}\, e^{-nMh_{nM}\,\varepsilon}\quad, \gamma > 0.$$

Thus,

$$P(N=\infty) = \lim_{n\to\infty} P(N>nM) = 0$$

2.6 Divergence of N as $\varepsilon \to 0$:

In the previous sections we have considered a fixed value of ε. We will now consider the problem of the behaviour of the stopping variable $N(\varepsilon,M)$ as $\varepsilon \to 0$. Intuition tells us that if $\varepsilon \to 0$, since $N(\varepsilon,M)$ depends on the criterion $|V_n(x)| \leq \varepsilon$, then $N(\varepsilon,M)$ should increase and that it should be unbounded above. That is it would be desirable if we could obtain that $\lim_{\varepsilon\to 0} N(\varepsilon,M) = \infty$ in probability (or with probability one). Unfortunately, we show that for a certain class of kernel this result may not always be true for our naive stopping rule.

Lemma 2.6.1

For finite n, if we define

$$\xi^0_{nM} = \{(x_1, x_2, \dots, x_{nM}) \;;\; |V_j(x)| = 0 \text{ for some } j\leq n\}$$

then as $\varepsilon \to 0$,

$$P(N\leq nM) \to P(\xi^0_{nM})$$

Proof:

By the definition of our stopping rule, for any fixed, finite n,

$$P(N \le nM) = P\left(\bigcup_{k=1}^{n} \Lambda_k^c\right)$$

where Λ_k^c is defined by (2.4.1). Let,

$$\xi_{nM}(\varepsilon) = \bigcup_{k=1}^{n} \Lambda_k^c.$$

Since n is fixed and finite, from the definition of Λ_k^c we see that $\xi_{nM}(\varepsilon)$ decreases monotonely as ε decreases so that we may write

$$\lim_{\varepsilon \to 0} P\left(\xi_{nM}(\varepsilon)\right) = P\left(\lim_{\varepsilon \to 0} \xi_{nM}(\varepsilon)\right)$$

But,

$$\lim_{\varepsilon \to 0} \xi_{nM}(\varepsilon) = \overset{0}{\xi}_{nM}$$

so that the result follows.

If $P\left(\overset{0}{\xi}_{nM}\right) = 0$, then it would follow that $N(\varepsilon,M) \to \infty$ in probability as $\varepsilon \to 0$, since we would have then proved that

$$\lim_{\varepsilon \to 0} P(N > nM) = 1$$

for any finite n. However, $P\left(\overset{0}{\xi}_{nM}\right)$ is not zero for all kernels as example 2.6.1 will show. That is we will show that $P\left(\overset{0}{\xi}_{nM}\right)$ can be strictly positive.

Example 2.6.1: Let

$$K(u) = \begin{cases} \frac{1}{2} & \text{if } |u| \le 1 \\ 0 & \text{otherwise} \end{cases}$$

For convenience in calculation we will assume that $\{h_n\}$ is a monotone decreasing sequence of positive constants(the monotone property is not necessary). That is, for all n, $h_n \ge h_{n+1}$.

Certainly, $V_n(x) = 0$ if $K\left(\frac{x-X_k}{h_{nM}}\right) = 0$ for all $k \le nM$, and from the definition of $K(u)$ we have that

$$(2.6.1)\quad P\left(K\left(\frac{x-X_k}{h_{nM}}\right) = 0\right) = P\left(\left|\frac{x-X_k}{h_{nM}}\right| > 1\right)$$

$$= 1 - P\left(x - h_{nM} \le X_k \le x + h_{nM}\right)$$

$$= 1 - \int_{x-h_{nM}}^{x+h_{nM}} f(u)\,du$$

Certainly for a wide class of density functions,

$$0 < \int_{x-h_{nM}}^{x+h_{nM}} f(u)\,du < 1$$

so that the right hand side of (2.6.1) is strictly positive.

Since h_n is monotone we have that

$$K\left(\frac{x-X_k}{h_{(n-1)M}}\right) = 0 \quad \text{implies} \quad K\left(\frac{x-X_k}{h_{nM}}\right) = 0$$

so that

$$(2.6.2)\quad P\left(\xi^0_{nM}\right) \ge P(V_n(x) = 0) \ge \left(P\left(\left|\frac{x-u}{h_{(n-1)M}}\right| > 1\right)\right)^{(n-1)M}\left(P\left(\left|\frac{x-u}{h_{nM}}\right| > 1\right)\right)^M$$

and the last quantity on the right hand side of (2.6.2) can be strictly positive.

Thus we have an example for which $P\left(\xi^0_{nM}\right)$ can be strictly positive and hence for which $N(\varepsilon,M) \not\to \infty$ in probability.

Remark:

We note that the essential feature in the example is that the uniform kernel has only bounded support which allows the possibility of obtaining a sample for which $K\left(\frac{x-X_k}{h_{nM}}\right) = 0$ for $k = 1,2, \cdots, nM$. This reasoning could be applied to any kernel with only bounded support. Thus we have found a class of kernels for which $N(\varepsilon,M)$ does not necessarily tend to infinity in probability as $\varepsilon \to 0$.

Two questions will now be considered. The first is "Are there any (or is there a class) of kernels for which $\lim_{\varepsilon\to 0} N(\varepsilon,M) = \infty$ in probability?"

Secondly, "Can we redefine our "naive" stopping rule to include the kernels with bounded support in the class above?"

The answer is yes to both questions and we now proceed to a proof of this statement.

Lemma 2.6.2

Let K_0 be the class of kernels satisfying condition (2.1.3) and for which

$$P(V_n(x) = 0) = 0$$

for any finite n, then for $K \in K_0$,

$$\lim_{\varepsilon\to 0} N(\varepsilon,M) = \infty \quad \text{in probability.}$$

Proof:

By definition,

$$P(N \le nM) = P\left(\bigcup_{k=1}^{n} \Lambda_k^c\right)$$

and from lemma 2.6.1 we have that

$$\lim_{\varepsilon\to 0} P\left(\bigcup_{k=1}^{n} \Lambda_k^c\right) = P\left(\xi_{nM}^0\right).$$

But,

$$P\left(\xi_{nM}^0\right) = P(V_j(x) = 0 \quad \text{for some } j \le n)$$

$$\leq P(V_1(x) = 0) + \cdots + P(V_n(x) = 0)$$

$$= 0 \text{ by assumption.}$$

Thus, $\lim_{\varepsilon \to 0} P(N \leq nM) = 0$ and the result follows.

It remains to show that K_0 is not an empty class (and if possible to show it includes at least some of the more commonly used kernels). To accomplish this we will show in lemma 2.6.3 that under appropriate conditions, $V_n(x)$ has an absolutely continuous distribution.

Lemma 2.6.3

If K(u) is a kernel satisfying conditions (2.1.3) and if $X_1, \cdots, X_{nM}$ are independently and identically distributed random variables with density function f(x), and, if in addition K(u) satisfies

(i) K(u) is differentiable for all but a finite number of values of u

(2.6.3) (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.

Proof:

Let $$y_k = \frac{1}{nMh_{nM}} K\left(\frac{x-X_k}{h_{nM}}\right), \quad k = 1,2,\cdots, nM.$$

then under the conditions stated, y_k, $k = 1,2,\cdots$, nM, have absolutely continuous distribution. (Parzen 1960 p313)

But then $z_2 = y_1 + y_2$, has a density function which is the convolution of the densities of y_1 and y_2. (Parzen (1960) p317) By induction we can thus show that

$$\hat{f}_{nM}(x) = z_{nM} = y_1 + \cdots + y_{nM}$$

has a density function that is the nM - fold convolution of the density of y_1. It then follows that

$$V_n(x) = z_{nM} - z_{(n-1)M}$$

has a density function. (Parzen (1960), p318)

Remarks:

1. If $V_n(x)$ has an absolutely continuous distribution then clearly $P(V_n(x) = 0) = 0$.

2. The normal, double exponential and Cauchy kernels satisfy the conditions stated so that

3. $\mathcal{K}_0$ is not empty.

4. We note that for the uniform kernel, $K'(u) = 0$ a.e. so that the uniform kernel is not in the class considered.

Before redefining our stopping variable we will first prove that for kernels, $K \varepsilon \mathcal{K}_0$, that N tends to infinity with probability one.

Theorem 2.6.4:

Define $V_n(x)$ and $\hat{f}_n(x)$ as previously. The kernel, $K(u)$, is assumed to satisfy conditions (2.1.3) and (2.6.3) and $\{h_n\}$ satisfies conditions (2.1.4) and (2.1.5). Then the stopping variable N, defined by (2.1.2) is such that

$$N(\varepsilon,M) \to \infty \text{ as } \varepsilon \to 0 \text{ with probability one.}$$

Proof:

In lemma 2.2.1 we proved that under conditions (2.1.3), $V_n(x) \to 0$ a.s. There are three possibilities to consider

(i) $V_n(x) \not\to 0$ as $n \to \infty$

(ii) $V_n(x) \to 0$ as $n \to \infty$ but $V_n(x) \neq 0$ for any finite n, and,

(iii) $V_n(x) \to 0$ as $n \to \infty$ and for at least one finite n, $V_n(x)=0$.

We consider these in turn.

(i) If $V_n(x) \not\to 0$ then by the definition of $N(\varepsilon,M)$, $N = \infty$ so that the theorem holds in this case. (Note also that $V_n(x) \not\to 0$

only on a set with probability zero).

(ii) By assumption, $V_n(x) \to 0$ as $n \to \infty$ but $|V_n(x)| > 0$ for all finite n.

Choose a finite, fixed number N_0. We will now show that for any point ω in the subspace for which (ii) holds, we can choose an $\varepsilon = \varepsilon(\omega, N_0)$ such that $N(\omega) > N_0$, for any finite N_0.

By assumption, $|V_j(x)| > 0$ for $j \leq N_0$. Assume $N(\omega) \leq N_0$ for all ε. This implies $|V_j(x)| < \varepsilon$ for at least one $j \leq N_0$. Since $|V_j(x)|>0$ for all $j \leq N_0$ we can choose $\varepsilon(\omega) < \min_{1 \leq j \leq N_0} |V_j(x)|$ so that $N(\omega) > N_0$ and we have a contradiction.

(iii) Since $K \varepsilon \mathcal{K}_0$ we have by assumption that $P(V_n(x) = 0) = 0$ for all finite n so that this third case can occur at most on a set of probability zero. This completes the proof of the theorem.

We have seen however that all kernels are not in the class $\mathcal{K}_0$. It is now our purpose to redefine N say N' in such a way so that even if $K \notin \mathcal{K}_0$ we may still obtain $\lim_{\varepsilon \to 0} N'(\varepsilon, M) \to \infty$ at least in probability, where N' will be the new stopping variable. In redefining N we endeavour to keep the properties of the class $\mathcal{K}_0$ from changing.

Definition (2.6.4):

Given $\varepsilon > 0$, and $M > 0$,

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

Remarks:

1. The only difference between definition (2.6.4) and definition (2.1.2) is the requirement that we continue sampling if $V_n(x) = 0$. We thus remove the problems caused by the fact that

$P(V_n(x) = 0)$ may be strictly positive.

2. If $K \varepsilon \mathcal{K}_0$ then by the definition of the class $\mathcal{K}_0$, $P(V_n(x) = 0) = 0$, so that $N = N'$ a.s. Thus, for the class $\mathcal{K}_0$, the stopping rules defined by N and N' are essentially the same.

Theorem 2.6.6:

If $K(u)$ and h_n satisfy the conditions of theorem 2.3.6 then

$$EN' < \infty .$$

Proof:

By definition (2.6.4)

$$P(N'>nM) = P\left(\bigcap_{k=1}^{n} \Lambda_k'\right)$$

where $\Lambda_k' = \Lambda_k \cup \Lambda_k^0$, Λ_k' being defined by (2.3.4) and Λ_k^0 by,

$$\Lambda_k^0 = \{(x_1, \cdots, x_{nM}) \; ; \; |V_k(x)| = 0 \quad , \; k \leq nM\}$$

Thus,

$$P(N'>nM) \leq P(\Lambda_n') \tag{2.6.5}$$

$$\leq P(\Lambda_n) + P(\Lambda_n^0).$$

Now, using (2.6.5) we get

$$EN' = \sum_{n=0}^{\infty} P(N'>nM)$$

$$\leq \sum_{n=0}^{\infty} P(\Lambda_n) + \sum_{n=0}^{\infty} P(\Lambda_n^0) \tag{2.6.6}$$

But by theorem (2.3.6) the first term on the right hand side of (2.6.6) is finite so it only remains to prove

$$\sum_{n=o}^{\infty} P(\Lambda_n^0) < \infty.$$

Now, by definition (2.1.2)

$$P(N=nM) = P\left(\left(\bigcup_{k=1}^{n} \Lambda_k\right) \cup \Lambda_n^c\right)$$

$$\geq P(\Lambda_n^c)$$

$$= P\left(|V_n(x)| \leq \varepsilon\right)$$

This means that we have the inequality

$$P(N=nM) \geq P\left(|V_n(x)| = 0\right)$$

But,

$$EN = \sum_{n=0}^{\infty} nM\, P(N=nM)$$

$$\geq \sum_{n=0}^{\infty} nM\, P\left(|V_n(x)| = 0\right)$$

$$\geq \sum_{n=0}^{\infty} P\left(|V_n(x)| = 0\right)$$

$$= \sum_{n=0}^{\infty} P(\Lambda_n^0)$$

However, by Theorem 2.3.6, we have that EN is finite so that we have proved that

$$\sum_{n=0}^{\infty} P(\Lambda_n^0) < \infty$$

and the proof of the theorem is complete.

Remarks:

1. With N' defined by (2.6.4) and using the uniform kernel we will now have

$$\lim_{\varepsilon\to 0} N'(\varepsilon,M) = \infty \quad \text{in probability}$$

since we now have

$$P(N \leq nM) \to 0 \quad \text{as } \varepsilon \to 0.$$

2. If $K \varepsilon \mathcal{K}_0$, since $P(V_n(x) = 0) = 0$ for all finite n we have that $N = N'$ a.s. and this will imply that $EN = EN'$.

3. If $K \notin \mathcal{K}_0$ then from the definitions of the two stopping variables we have that

$$EN \leq EN'$$

2.7 Mean Square Error:

We would like to show that

$$E\left(\hat{f}_N(x) - f(x)\right)^2 \to 0 \quad \text{as } \varepsilon \to 0. \tag{2.7.1}$$

Under appropriate conditions we can show that result (2.7.1) is true. However, we will leave the proof of this until the theory for a general stopping rule has been developed in Chapter 3. The result will then follow as a special case of the general theory.

We state the result here only for the information of the reader and since the question of mean square error might logically have been treated at this point.

CHAPTER 3

SOME GENERAL RESULTS

3.1. Introduction and Assumptions:

In chapter 2 we discussed a "naive" sequential procedure, $N(\varepsilon,M)$. Clearly, there are many possible sequential procedures that could be considered and it is our purpose in this chapter to discuss some of the properties of an estimator $f_N(x)$ that depend on the specific properties of a stopping rule, N. We note that these properties will also depend on the properties of the fixed sample estimate, $f_n(x)$. For the purposes of this chapter we will make as few assumptions as possible about the estimator $f_n(x)$ and the stopping rule N, noting in particular that N should not be considered as the "naive" stopping variable of chapter 2 and $f_n(x)$ should not be considered as a Parzen type estimator of a density function (although the "naive" stopping rule and the Parzen estimator will provide examples of the general theory). Throughout this chapter however, we will assume that $f_n(x) \geq 0$.

We begin by making several assumptions about the sequential procedure N.

Assume we are given a set of observations $X_1, X_2, \cdots$ which are random variables on the same measurable space $(\mathcal{X},\mathcal{F})$ and assume there exists an (unknown) probability measure P over the measurable space. P determines the distribution functions of finite subsets of the X's.

The sequential procedure will consist of a stopping rule N and a terminal decision rule D. Note that N is often used to denote the random size of the sample which results from applying the rule N. The sense in which we use the symbol N will be clear from the context in which it appears.

Similarly, D is often used to denote the action taken.

Let,

$$\mathcal{B}_n = \mathcal{B}(X_1, X_2, \cdots, X_n)$$

$$= \{A \varepsilon \mathcal{F} : A = (X_1, X_2, \cdots, X_n)^{-1} (B) \varepsilon \mathcal{F},\ B \varepsilon R^n \text{ a Borel set}\} \subset \mathcal{F}.$$

$$\mathcal{B}_0 = \{\phi, \mathcal{F}\}.$$

We then assume

(i) N is an extended random variable on $(\mathcal{X}, \mathcal{F})$, that is

$$N \varepsilon \{0, 1, 2, \cdots, \infty\}$$

(ii) $[N=n] = \{\text{set on which } N = n\} \ \varepsilon\ \mathcal{B}_n$

(iii) $[N=n, D=a] \ \varepsilon\ \mathcal{B}_n$.

Throughout this chapter we will use a slightly modified notation for the estimator $f_n(x)$ at the point x. We will write

(3.1.1) $$f_n(x) \equiv f_n(x; x_1, x_2, \cdots, x_n)$$

where we wish to demonstrate more clearly the dependence of $f_n(x)$ on the random variables, $X_1, X_2, \cdots, X_n$.

In the sections that follow we will make use of a theorem giving the expectation of a function $g_N(x_1, \cdots x_N)$ where N is a stopping variable. We now state this theorem without proof.

Theorem 3.1.1

Let $g_1(x_1)$, $g_2(x_1, x_2), \cdots$ be a sequence of non-negative, Borel measurable functions on R^1, R^2, . Let N be a stopping variable and suppose $P(1 < N < \infty) = 1$. Then $Y = g_N(X_1, \cdots, X_N)$ is a random variable, EY is defined (possibly = + ∞) and

$$Eg_N(x_1,\cdots, x_N) = \sum_{\ell=1}^{\infty} \int \cdots \int_{\Omega_\ell} g_\ell(x_1,\cdots, x_\ell)\, dF(x_1,\cdots, x_\ell)$$

where

$$\Omega_\ell = \{(x_1,\cdots, x_\ell) : [N = \ell]\},$$

$$= \text{subset of } R^\ell \text{ on which } [N = \ell]$$

3.2 Finite Expected Value of $f_N(x)$:

In chapter 2 considerable effort was placed in proving $EN < \infty$. This result is desirable in its own right, but also implies that $Ef_N(x) < \infty$. We prove this in the next lemma.

Lemma 3.2.1

If $f_n(x) \geq 0$ and if $\sup_n \sup_{X_1,\cdots, X_n} \frac{1}{n} f_n(x ; X_1,\cdots, X_n) < \infty$,

then a sufficient condition that $Ef_N(x)$ is finite is that $EN < \infty$.

Proof:

By theorem 3.1.1,

$$Ef_N(x) = \sum_{\ell=1}^{\infty} \int_{\Omega_\ell} f_\ell(x ; x_1, x_2,\cdots, x_\ell)\, dF(x_1, x_2,\cdots, x_\ell)$$

where $\Omega_\ell = \{(x_1, x_2,\cdots) ; N = \ell\} = \{\text{set on which we stop when } N=\ell\}$.

Thus,

$$Ef_N(x) = \sum_{\ell=1}^{\infty} \ell \int_{\Omega_\ell} \frac{1}{\ell} f_\ell(x ; x_1, x_2,\cdots, x_\ell)\, dF(x_1, x_2,\cdots, x_\ell)$$

$$\leq \sup_n \sup_{(x_1,\cdots,x_n)} \left\{\frac{1}{\ell} f_\ell(x ; x_1,\cdots,x_\ell)\right\} \sum_{\ell=1}^{\infty} \ell \int_{\Omega_\ell} dF(x_1,\cdots,x_\ell)$$

and since $\int_{\Omega_\ell} dF(x_1,\cdots,x_\ell) = P[N = \ell]$ we get

$$Ef_N(x) \le \sup_\ell \sup_{X_1, X_2, \cdots, X_n} \frac{1}{\ell} f_\ell(x ; X_1,\cdots, X_n) \, EN < \infty$$

by assumption.

Corollary 3.2.2

If $f_n(x) = \hat{f}_n(x) = \frac{1}{nh_n} \sum_{j=1}^{n} K\left(\frac{x-X_j}{h_n}\right)$, and N is defined by

(2.1.2) then under the assumptions of Theorem (2.3.6), $Ef_N(x) < \infty$.

Proof:

By theorem (2.3.6), $EN < \infty$. Thus it remains only to show that $\sup_n \frac{1}{n} \hat{f}_n(x) < \infty$. Now,

$$\frac{1}{n} \hat{f}_n(x) = \frac{1}{n^2 h_n} \sum_{j=1}^{n} K\left(\frac{x-X_j}{h_n}\right)$$

$$\le \frac{1}{n^2 h_n} \, n \sup_u K(u) = \frac{1}{nh_n} \sup_u K(u)$$

and as $nh_n \to \infty$ as $n\to\infty$ and $\sup_u K(u) < \infty$ we have that $\sup_n \frac{1}{n} \hat{f}_n(x) < \infty$.

Thus the result follows.

3.3 Convergence Theorems:

Previously in chapter 2 we made a point of obtaining results such as $N(\varepsilon,M)$ tending to infinity as ε tends to zero in both probability and with probability one. These results are desirable in themselves, but, in this section we will show that they are also required to obtain results concerning the density estimator $f_N(x)$ where N is a random variable. We will discuss these results as a series of theorems .

Theorem 3.3.1:

If $N_\varepsilon \to \infty$ in probability and if $f_n(x) \to f(x)$ a.s. as $n \to \infty$ and if in addition

$$\sup_n \sup_{(x_1,\ldots,x_n)} f_n(x;\ x_1,\ldots,x_n) < \infty$$

then $Ef_{N_\varepsilon}(x) \to f(x)$ as $\varepsilon \to 0$.

Proof:

Since $f_n(x) \to f(x)$ a.s. as $n \to \infty$, then by Egoroff's theorem we have that $f_n(x) \to f(x)$ a.u. That is, given $\delta > 0$ there exists an $n_0 = n_0(\delta)$ such that if $n > n_0$,

$$|f_n(x) - f(x)| < \delta ,$$

except possibly on a set A of probability less than δ.

Also, since $N_\varepsilon \to \infty$ in probability as $\varepsilon \to 0$, there exists an $\varepsilon_0 = \varepsilon_0(\delta, n_0)$ such that for $\varepsilon \le \varepsilon_0$ we have

$$P(N_\varepsilon \ge n_0) > 1 - \delta .$$

Now, by Theorem 3.1.1

$$Ef_{N_\varepsilon}(x) = \sum_{\ell=1}^{\infty} \int_{\Omega_\ell} f_\ell(x;\ x_1,\ldots,x_\ell)dF(x_1,\ldots,x_\ell)$$

$$= \sum_{\ell=1}^{\infty} \{ \int_{A^c\cap\Omega_\ell} f_\ell(x;\ x_1,\ldots,x_\ell)dF(x_1,\ldots,x_\ell)$$

$$+ \int_{A\cap\Omega_\ell} f_\ell(x;\ x_1,\ldots,x_\ell)dF(x_1,\ldots,x_\ell)\}$$

But for $n \geq n_0$ on the set A^c we have that

$$f(x) - \delta \leq f_n(x) \leq f(x) + \delta$$

so that

$$\sum_{\ell=n_0}^{\infty} \{ \int_{A^c \cap \Omega_\ell} f_\ell(x;\, x_1,\ldots,x_\ell) dF(x_1,\ldots,x_\ell)$$

$$+ \int_{A \cap \Omega_\ell} f_\ell(x;\, x_1,\ldots,x_\ell) dF(x_1,\ldots,x_\ell) \}$$

$$\leq (f(x)+\delta) \sum_{\ell=1}^{\infty} \int_{A^c \cap \Omega_\ell} dF(x_1,\ldots,x_\ell) + c \sum_{\ell=1}^{\infty} P(A \cap \Omega_\ell)$$

where $c = \sup_\ell \sup_{(x_1,\ldots,x_\ell)} f_\ell(x;\, x_1,\ldots,x_\ell)$. The right hand side of the inequality is then less than or equal to

$$f(x) + \delta + c\, P(A) \leq f(x) + \delta + c\delta$$

since $P(A) \leq \delta$ by assumption. Also,

$$\sum_{\ell=1}^{n_0-1} \int_{\Omega_\ell} f_\ell(x;\, x_1,\ldots,x_\ell) dF(x_1,\ldots,x_\ell)$$

$$\leq c \sum_{\ell=1}^{n_0-1} \int_{\Omega_\ell} dF(x_1,\ldots,x_\ell)$$

$$= c \sum_{\ell=1}^{n_0-1} P(N = \ell)$$

$$= c\, P(N < n_0)$$

$\leq c\delta$ by assumption for $\varepsilon \leq \varepsilon_0$.

Thus we have shown that for $\varepsilon \le \varepsilon_0$,

(3.3.1) $$Ef_{N_\varepsilon}(x) \le f(x) + \delta(1+2c) .$$

But,

$$Ef_N(x) \ge \sum_{\ell=n_0}^{\infty} \int_{A^c \cap \Omega_\ell} f_\ell(x;\, x_1,\ldots,x_\ell) dF(x_1,\ldots,x_\ell)$$

$$\ge (f(x)-\delta) \sum_{\ell=n_0}^{\infty} P(A^c \cap \Omega_\ell) .$$

Now, $P(\Omega_\ell) = P(\Omega_\ell \cap A^c) + P(\Omega_\ell \cap A)$ and

$$\sum_{\ell=n_0}^{\infty} P(\Omega_\ell \cap A) \le \sum_{\ell=1}^{\infty} P(\Omega_\ell \cap A) = P(A) \le \delta .$$

Further, if $\varepsilon \le \varepsilon_0$

$$\sum_{\ell=n_0}^{\infty} P(\Omega_\ell) = \sum_{\ell=n_0}^{\infty} P(\Omega_\ell \cap A^c) + \sum_{\ell=n_0}^{\infty} P(\Omega_\ell \cap A) \ge 1 - \delta .$$

Thus,

$$\sum_{\ell=n_0}^{\infty} P(\Omega_\ell \cap A^c) \ge 1 - \delta - \sum_{\ell=n_0}^{\infty} P(\Omega_\ell \cap A)$$

$$\ge 1 - 2\delta$$

and so finally we obtain

(3.3.2) $$Ef_{N_\varepsilon}(x) \ge (f(x)-\delta)(1-2\delta) .$$

Since δ is arbitrary, the required result now follows from the inequalities (3.3.1) and (3.3.2).

EXAMPLE 3.3.1:

Let $f_n(x) = \hat{f}_n(x)$ and let N_ε be defined by (2.1.2). Since $\sup_n \sup_{(x_1,\ldots,x_n)} \hat{f}_n(x)$ may be ∞, let us define

$$\hat{g}_n(x) = \begin{cases} \hat{f}_n(x) & \text{if } \hat{f}_n(x) \le R \\ R & \text{if } \hat{f}_n(x) > R \end{cases}$$

where R is a finite constant. We assume R is chosen very large and that $R > f(x)$. Note we still use the stopping rule defined by (2.1.2) since (i) to calculate $\hat{g}_n(x)$ we first must calculate $\hat{f}_n(x)$ and (ii) if the stopping rule were based on $\hat{g}_n(x)$ there would be some difficulty as the property $N_\varepsilon \to \infty$ a.s. as $\varepsilon \to 0$ would be lost. Notice also that since $R > f(x)$ then provided $\hat{f}_n(x) \to f(x)$ a.s. as $n \to \infty$ (for conditions see Lemma 2.2.2) then $\hat{g}_n(x) \to f(x)$ a.s. as $n \to \infty$, so that the conditions of theorem 3.3.1 are satisfied and $E\hat{g}_{N_\varepsilon}(x) \to f(x)$ as $\varepsilon \to 0$.

Remarks:

1. The theorem remains true if we have only $\sup_n \sup_{(X_1,\ldots,X_n)} f_n(x) < \infty$ (a.s.).

2. From a practical standpoint, provided R is made large enough, the truncation of $f_n(x)$ should have little effect on the actual estimates we obtain for $f(x)$.

3. We now show that asymptotically as $n \to \infty$, $\hat{f}_n(x)$ and $\hat{g}_n(x)$ have the same properties. To do this we will consider the general case where $g_n(x)$ is the truncated estimator corresponding to $f_n(x)$, where it is assumed $f_n(x) \to f(x)$ a.s.

Asymptotic Properties of a Truncated Estimator:

Define,

$$g_n(x) = \begin{cases} f_n(x) & \text{if } f_n(x) \le R \\ R & \text{if } f_n(x) > R \,. \end{cases} \tag{3.3.3}$$

Lemma 3.3.2 (Asymptotic unbiasedness):

If $f_n(x) \to f(x)$ a.s. and if $g_n(x)$ is defined by (3.3.3) then $Eg_n(x) \to f(x)$ as $n \to \infty$, provided that $R > f(x)$.

Proof:

Since $R > f(x)$, $g_n(x) \to f(x)$ a.s. so that by Egoroff's theorem, $g_n(x) \to f(x)$ a.u. That is, given $\delta > 0$ there exists an $n_0 = n_0(\delta)$ such that if $n > n_0$, $|g_n(x) - f(x)| < \delta$ except possibly on a set A of probability less than δ. But

$$Eg_n(x) \ge \int_{A^c} g_n(x) dF(x_1,\ldots,x_n)$$

$$\ge (f(x)-\delta)P(A^c) \qquad \text{for } n \ge n_0$$

so that

$$Eg_n(x) \ge (f(x)-\delta)(1-\delta) \,. \tag{3.3.4}$$

Now

$$Eg_n(x) = \int_{A^c} g_n(x)dF(x_1,\ldots,x_n) + \int_A g_n(x)dF(x_1,\ldots,x_n)$$

$$\le (f(x)+\delta)P(A^c) + RP(A)$$

$$\le (f(x)+\delta) + R\delta$$

so that

(3.3.5) $$Eg_n(x) \le f(x) + \delta(1+R) .$$

Since δ is arbitrary, the result follows immediately from inequalities (3.3.4) and (3.3.5).

Lemma 3.3.3:

(i) $Ef_n^j(x) \ge Eg_n^j(x)$ for every $j > 0$.

(ii) $Var(f_n(x)) \ge Var(g_n(x))$

Proof: These results are well known. We give only a sketch of the proof.

$$Ef_n^j(x) = \int f_n^j(x)dF(x_1,\ldots,x_n)$$

$$= \int_{B_n} f_n^j(x)dF(x_1,\ldots,x_n) + \int_{B_n^c} f_n^j(x)dF(x_1,\ldots,x_n) ,$$

where $B_n = \{(x_1,\ldots,x_n) : f_n(x) > R\}$.

Hence $$Ef_n^j(x) \ge \int_{B_n} R^j dF(x_1,\ldots,x_n) + \int_{B_n^c} f_n^j(x)dF(x_1,\ldots,x_n)$$

$$= \int g_n^j(x)dF(x_1,\ldots,x_n) = Eg_n^j(x) .$$

The result ii for variances follows by subtracting R from both $f_n(x)$ and $g_n(x)$ and then examining $\text{cov}(g_n(x) - R, h_n(x) - R)$ where $h_n(x) = \max(f_n(x), R)$.

Observations:

1. If $\text{var } f_n(x) \to 0$ as $n \to \infty$, then so also does $\text{var}(g_n(x)) \to 0$ as $n \to \infty$. The rate of convergence of $\text{var}(g_n(x))$ to zero is at least as fast as the rate of $\text{var}(f_n(x))$.

2. Since
$$E(g_n(x) - f(x))^2 = E(g_n(x) - Eg_n(x) + Eg_n(x) - f(x))^2$$
$$= E(g_n - Eg_n(x))^2 + (Eg_n(x) - f(x))^2,$$
using Lemmas 3.3.2 and 3.3.3, we see that $E(g_n(x) - f(x))^2 \to 0$ as $n \to \infty$, so that $g_n(x)$ is mean square consistent.

Comment on Notation:

We use the stopping variable N_ε to denote the dependence of N on some criterion. For consistency with chapter 2 we take this criterion as dependent on a parameter ε for which we wish to consider the behavior of N and $f_N(\cdot)$ as $\varepsilon \to 0$. Note, however, that only minor modifications are necessary in the proofs in this chapter if the criterion were assumed to depend on a parameter t, and t were assumed to tend to infinity.

Theorem 3.3.4:

If $N_\varepsilon \to \infty$ in probability as $\varepsilon \to 0$ and if $f_n(x) \to f(x)$ a.s.

as $n \to \infty$ and $\sup_n \sup_{(X_1,\ldots,X_n)} f_n(x; X_1,\ldots,X_n) < \infty$ then $E(f_{N_\varepsilon}(x) - Ef_{N_\varepsilon}(x))^2 \to 0$ as $\varepsilon \to 0$.

Proof:

Since we have already shown that $Ef_{N_\varepsilon}(x) \to f(x)$ under these conditions, it will be sufficient to show $Ef^2_{N_\varepsilon}(x) \to f^2(x)$. We will note that $Ef^2_{N_\varepsilon}(x) \geq 0$. An upper bound is obtained using a similar method to that employed in obtaining (3.3.1). The proof is therefore omitted.

Theorem 3.3.5:

If the conditions of Theorem 3.3.4 hold, then $E(f_{N_\varepsilon}(x) - f(x))^2 \to 0$ as $\varepsilon \to 0$.

Proof:

Since,

$$E(f_{N_\varepsilon}(x) - f(x))^2 = E(f_{N_\varepsilon}(x) - Ef_{N_\varepsilon}(x))^2 + (Ef_{N_\varepsilon}(x) - f(x))^2,$$

the result follows immediately from theorem 3.3.1 and 3.3.4.

Theorem 3.3.6:

If $f_n(x) \to f(x)$ a.s. as $n \to \infty$ and if $N_\varepsilon \to \infty$ a.s. as $\varepsilon \to 0$ then $f_{N_\varepsilon}(x) \to f(x)$ with probability one as $\varepsilon \to 0$.

Proof:

By combining the two sets of probability zero implicit in the statement of the lemma, the result is immediate.

Corollary 3.3.7:

Let $f_n(x) = \hat{f}_n(x)$ and let N be defined by (2.1.12). Then if K(u) and h_n satisfy the conditions of Lemma 2.2.2 and $K \in \mathcal{K}_0$ then

$$\lim_{\varepsilon\to 0} \hat{f}_{N_\varepsilon}(x) = f(x) \quad \text{with probability one.}$$

Proof:

The conditions are sufficient for both $\hat{f}_n(x) \to f(x)$ a.s. as $n \to \infty$ and $N \to \infty$ as $\varepsilon \to 0$ so the result follows immediately from theorem 3.3.6.

Remarks on Other Solutions:

In the course of this investigation two other methods were studied for obtaining results on the unbiasedness, asymptotic variance and mean square error for $f_{N_\varepsilon}(x)$. The following result constitutes alternate sufficient conditions for theorem 3.3.1. The difficulty with both methods is that I cannot prove the conditions can be satisfied for my proposed sequential procedures. We now state the theorem for unbiasedness with these conditions and sketch the proofs.

Theorem 3.3.1.A:

If $N_\varepsilon \to \infty$ in probability and if

(a) $$E[f_n(x)|N = n] \to f(x) \quad \text{as } n \to \infty$$

or (b) $$E(f_n(x) - f(x))^2 \to 0 \quad \text{as} \quad n \to \infty \quad \text{and}$$

$$\sum_{\ell=1}^{\infty}\{P(N = \ell)\}^{\frac{1}{2}} = c < \infty \text{ , where } c \text{ is a constant}$$

independent of ε, then

$$E(f_{N_\varepsilon}(x) - f(x)) \to 0 \quad \text{as} \quad \varepsilon \to 0.$$

Proof (a): This proof is similar to that used by Srivastava (1973) for the case when N_ε is independent of the observations since we can write

$$Ef_{N_\varepsilon}(x) = \sum_{n=1}^{\infty} E(f_{N_\varepsilon}(x)|N_\varepsilon = n)P(N_\varepsilon = n)$$

$$= \sum_{n=1}^{\infty} E(f_n(x)|N_\varepsilon = n)P(N_\varepsilon = n) \ .$$

Paralleling Srivastava's proof using $E(f_n(x)|N_\varepsilon = n)$ in place of $Ef_n(x)$ now gives the result. The details will be omitted.

The difficulty we encounter is that we have not been able to verify that condition (a) is able to be satisfied. Notice that

$$E[f_n(x)|N_\varepsilon = n] = \frac{\int_{[N_\varepsilon=n]} f_n(x)dF(x_1,\dots,x_n)}{P(N_\varepsilon = n)}$$

and for fixed ε , both numerator and denominator tend to zero. We need to know the rates of convergence involved. The inequality we have for $P(N_\varepsilon = n)$ is much too crude as $\varepsilon \to 0$ and we know little about the other rate so the condition has not been verified for $\hat{f}_n(x)$ and N as defined in chapter 2 (or for any other example).

Proof (b):

By theorem 3.1.1 we have

$$E(f_{N_\varepsilon}(x) - f(x)) = \sum_{\ell=1}^{\infty} \int_{\Omega_\ell} (f_\ell(x) - f(x))dF(x_1,\dots,x_\ell) .$$

Choosing n_0 in a similar manner as in theorem 3.3.1 we obtain

$$\sum_{\ell=n_0}^{\infty} \int_{\Omega_\ell} (f_\ell(x) - f(x))dF(x_1,\dots,x_\ell)$$

$$= \sum_{\ell=n_0}^{\infty} \int (f_\ell(x) - f(x))I_{\Omega_\ell}(x_1,\dots,x_\ell)dF(x_1,\dots,x_\ell)$$

where $I_{\Omega_\ell}(x_1,\dots,x_\ell)$ is the indicator function of Ω_ℓ . Using Schwarz's inequality we see that this is less than

$$\sum_{\ell=n_0}^{\infty} [\int (f_\ell(x) - f(x))^2 dF(x_1,\ldots,x_\ell)]^{\frac{1}{2}} [\int I_{\Omega_\ell}(x_1,\ldots,x_\ell) dF(x_1,\ldots,x_\ell)]^{\frac{1}{2}}$$

$$= \sum_{\ell=n_0}^{\infty} [E(f_\ell(x) - f(x))^2]^{\frac{1}{2}} \{P(N = \ell)\}$$

and by the way we choose n_0 we ensure this is less than

$$\delta \sum_{\ell=n_0}^{\infty} \{P(N = \ell)\}^{\frac{1}{2}} \le \delta c \quad \text{(by assumption)}.$$

Given this step it is now not difficult to complete the proof. However, the difficulty lies in showing $\sum_{\ell=1}^{\infty} \{P(N = \ell)\}^{\frac{1}{2}} \le c < \infty$, where c is independent of ε, can be satisfied in practice. We can show that for any fixed $\varepsilon > 0$, that $\sum_{\ell=1}^{\infty} \{P(N = \ell)\}^{\frac{1}{2}} = c(\varepsilon) < \infty$. However, the supremum over all such sums may be infinite. Once again the problem reduces to finding a suitable bound for $P(N = \ell)$.

Observations:

i. The proof of theorem 3.3.1.A under condition (b) is based on the use of Schwarz's inequality. A similar result follows by using the more general Hölder's Inequality. The sufficient condition (b) may be replaced by

(b') $E(f_n(x) - f(x))^p \to 0$ as $n \to \infty$ and

$$\sum\{P(N = \ell)\}^{1/q} \le c < \infty$$

where c is a constant independent of ε and $1/p + 1/q = 1$.

The same difficulty in determining the existence of a uniform bound, c, applies here also.

ii. The proof of theorem 3.3.1 involves the assumption that $\sup_n \sup_{(X_1,\ldots,X_n)} f_n(x; X_1,\ldots,X_n) < \infty$. This is made to ensure that

$$\sum_{\ell=1}^{\infty} \int_{A\cap\Omega_\ell} f_\ell(x; x_1,\ldots,x_\ell)dF(x_1,\ldots,x_\ell) \tag{3.3.6}$$

is sufficiently small whenever $P(A)$ is sufficiently small. Letting P be the probability measure on the usual infinite product space, (3.3.6) may be rewritten

$$\sum_{\ell=1}^{\infty} \int_{A\cap\Omega_\ell} f_\ell(x)dP \quad .$$

Let us assume

$$\int_B f_\ell(x)dP \le kP(B) \tag{3.3.7}$$

for some $k \ge 0$ and for every measurable set B. Clearly

$$\sum_{\ell=1}^{\infty} \int_{A\cap\Omega_\ell} f_\ell(x)dP \le \sum_{\ell=1}^{\infty} kP(A\cap\Omega_\ell)$$

$$= kP(A) \quad ,$$

hence (3.3.7) is a more general condition than $\sup_n \sup_{(X_1,\ldots,X_n)} f_n(x;\, X_1,\ldots,X_n) < \infty$. Unfortunately, it appears that this is difficult to verify (P being unknown) except in the circumstance that f_n is uniformly bounded in n and in $(X_1,\ldots,X_n)$.

iii. Analogues of theorems 3.3.4 and 3.3.5 can be proved with sufficient conditions similar to those of theorem 3.3.1.A.

Theorem 3.3.4.A:

If $N_\varepsilon \to \infty$ in probability and if

(a) $E[(f_n(x) - Ef_n(x))^2|N=n] \to 0$ as $n \to \infty$

or (b) $E(f_n(x) - Ef_n(x))^{2p} \to 0$ as $n \to \infty$ and

$$\sum_{\ell=1}^{\infty} \{P(N = \ell)\}^{1/q} \le c < \infty \ ,$$

where c is a constant independent of ε and $1/p + 1/q = 1$, then

$$E(f_{N_\varepsilon}(x) - Ef_{N_\varepsilon}(x))^2 \to 0 \quad .$$

Moreover, if the bias tends to zero (theorem 3.3.1.A)

$$E(f_{N_\varepsilon}(x) - f(x))^2 \to 0 \ .$$

Note, however, that to prove the mean square error, $E(f_{N_\varepsilon}(x) - f(x))^{2p} \to 0$

we would also need to show $E(f_n(x) - f(x))^{2p} \to 0$ as $\varepsilon \to 0$. Using the results of Parzen (1962) this can be shown to be true for the estimator $\hat{f}_n(x)$. The proof is not difficult, involving only some algebraic manipulation and will be omitted.

iv. Finally we observe that theorem 3.3.6 and corollary 3.3.7 do not have any of the requirements for a uniform bound or for uniform integrability as required by theorems 3.3.1, 3.3.4 and 3.3.5 or their analogues. Moreover, theorems 3.3.6 and corollary 3.3.7 also guarantee strong convergence.

CHAPTER 4

THE YAMATO SEQUENTIAL ESTIMATOR.

4.1 Introduction:

Let $X_1, X_2, \cdots, X_n$ be a sequence of independently distributed, m - dimensional random vectors having a probability density function f(x). Yamato (1972) then defines a "sequential" estimator of the density function f(x) as

$$\tilde{f}_n(x) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{h_j^m} K\left(\frac{x-X_j}{h_j}\right) \tag{4.1.1}$$

Notice that the estimate $\tilde{f}_n(x)$ may be written as

$$\tilde{f}_n(x) = \frac{(n-1)}{n} \tilde{f}_{n-1}(x) + \frac{1}{nh_n^m} K\left(\frac{x-X_n}{h_n}\right) \tag{4.1.2}$$

This means that $\tilde{f}_n(x)$ may be more suitable that $\hat{f}_n(x)$ for use as a "sequential estimator" since it has the property of correcting the estimate succesively as observations are added.

A further reason for considering the Yamato type estimator is that Yamato proves that in the case m = 1,

$$\lim_{n\to\infty} \frac{\text{Var}\left(\tilde{f}_n(x)\right)}{\text{Var}\left(\hat{f}_n(x)\right)} = \nu_0 \leq 1, \text{ where } \nu_0 \text{ depends on the} \tag{4.1.3}$$

sequence $\{h_n\}$, so that $\tilde{f}_n(x)$ is at least as efficient as $\hat{f}_n(x)$, the Parzen type estimate.

We will first show that the estimator $\tilde{f}_n(x)$ defined by (4.1.1) is strongly consistent. That is, if x is contained in the continuity set C(f), then $\tilde{f}_n(x) \to f(x)$ with probability one. Our method of proof is to use a technique due to Van Ryzin (1969), who studied the strong consistency of estimates of the form $\hat{f}_n(x)$. Notice, that due to the special form of $\tilde{f}_n(x)$, Van Ryzin's conditions have been considerably weakened. Since we will be considering m - dimensional space and not 1 - dimensional space as in Chapters 2 and 3 we will require the following conditions:

(4.1.4)
(i) $K(u)$ is a density on R^m

(ii) $\sup_{u \varepsilon R^m} |K(u)| < \infty$

(iii) $||u|| \, K(u) \to 0$ as $||u||^2 = \sum_{i=1}^{m} u_i^2 \to \infty$

where $K(u) = K(u_1, u_2, \cdots, u_m)$ is a real valued, Borel measurable function on R^m and R^m is an m - dimensional Euclidean space.

Further we will require that $\{h_n\}$ is a sequence of real numbers such that condition (2.1.4) is satisfied, that is (i) $h_n > 0$, $n = 1,2,\cdots$; (ii) $\lim_{n\to\infty} h_n = 0$ and $\lim_{n\to\infty} nh_n = \infty$.

The strong consistency of $\tilde{f}_n(x)$ is proved in section 4.2.

In sections 4.3, through 4.7 we will parallel the work of Chapter 2 and define a "naive" stopping rule based on

$$\tilde{V}_n(x) = \tilde{f}_{nM}(x) - \tilde{f}_{(n-1)M}(x) \tag{4.1.5}$$

where $\tilde{f}_{nM}(x)$ is defined by (4.1.1) with m = 1.

4.2 Strong Consistency:

In order to show strong consistency we will use the following lemma which is proved in Van Ryzin (1969). The proof is omitted here.

Lemma 4.2.1:

Let $\{Y_n\}$ and $\{Y'_n\}$ be two sequences of random variables on a probability space $(\mathcal{X}, \mathcal{F}, P)$. Let $\{\mathcal{F}_n\}$ be a sequence of Borel fields, $\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{F}$, where Y_n and Y_n' are measurable with respect to $\mathcal{F}_n$. If

(4.2.1)

(i) $0 \le Y_n$ a.e.

(ii) $EY_1 < \infty$

(iii) $E\left(Y_{n+1} \mid \mathcal{F}_n\right) \le Y_n + Y_n'$ a.e.

(iv) $\sum_{n=1}^{\infty} E|Y_n'| < \infty$

then Y_n converges a.e. to a finite limit.

Theorem 4.2.2 now gives the pointwise strong consistency for the estimator, $\tilde{f}_n(x)$.

Theorem 4.2.2: (Pointwise Strong Consistency).

If K(u) satisfies (4.1.4) and $\{h_n\}$ is a monotone decreasing sequence of numbers satisfying (2.1.4) and if in addition

$$\sum_{n=1}^{\infty} \frac{1}{n^2 h_n^m} < \infty \tag{4.2.2}$$

and if

$$0 < \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \frac{h_n^m}{h_j^m} = \nu_0 < 1 \tag{4.2.3}$$

then $\tilde{f}_n(x) \to f(x)$ with probability one if $x \in C(f)$.

Proof:

First note that $E\tilde{f}_n(x) \to f(x)$ as $n\to\infty$ (Yamato (1972)). Thus it is sufficient to show $\tilde{f}_n(x) - E\tilde{f}_n(x) \to 0$ with probability one as $n\to\infty$.

Let $Y_n = \left(\tilde{f}_n(x) - E\tilde{f}_n(x)\right)^2$ and $\mathcal{F}_n$ = Borel field generated by $X_1, X_2, \cdots, X_n$ and note that $\lim_{n\to\infty} E(Y_n) = 0$. (Yamato(1972)). From (4.1.2) we now have that

$$\tilde{f}_{n+1}(x) - E\tilde{f}_{n+1}(x) = \tilde{f}_n(x) - E\tilde{f}_n(x) - \frac{1}{n+1}\left(\tilde{f}_n(x) - E\tilde{f}_n(x)\right)$$

$$+ \frac{1}{(n+1)}\left(\frac{1}{h_{n+1}^m} K\left(\frac{x-X_{n+1}}{h_{n+1}}\right) - \frac{1}{h_{n+1}^m} EK\left(\frac{x-X_{n+1}}{h_{n+1}}\right)\right)$$

Thus,

$$E\left(Y_{n+1} \mid F_n\right) = Y_n + \frac{1}{(n+1)^2} Y_n - \frac{2}{(n+1)} Y_n + \frac{1}{(n+1)^2 h_{n+1}^{2m}} \text{Var } K\left(\frac{x-X}{h_{n+1}}\right).$$

Now define

$$Y_n' = \frac{1}{(n+1)^2} Y_n - \frac{2}{(n+1)} Y_n + \frac{1}{(n+1)^2 h_{n+1}^{2m}} \text{Var } K\left(\frac{x-X}{h_{n+1}}\right)$$

In order to use Lemma 1 we must verify conditions (4.2.1), of which we already have that (i), (ii), and (iii) and satisfied. Thus to complete the proof we must verify

$$\sum_{n=1}^{\infty} E|Y_n'| < \infty.$$

But,

$$\sum_{n=1}^{\infty} \frac{1}{(n+1)} E|Y_n| = \sum_{n=1}^{\infty} \frac{1}{(n+1)} \text{Var } \tilde{f}_n(x).$$

Now, Yamato (1972), using a monotone sequence satisfying (2.1.4) and with ν_0 defined by (4.2.3) has shown (Theorem 3) that

$$\text{(4.2.4)} \qquad \lim_{n\to\infty} nh_n^m \text{ Var } \tilde{f}_n(x) = \nu_0 f(x) \int_{R^m} K^2(y)\, dy.$$

We note that since $\{h_n\}$ is a decreasing sequence, $\nu_0 \le 1$.

Now, for large n we thus have,

$$\frac{1}{n} \text{Var } \tilde{f}_n(x) = O\left(\frac{1}{n^2 h_n^m}\right).$$

Hence, using condition (4.2.2) we have that

$$\sum_{n=1}^{\infty} \frac{1}{(n+1)} \text{ Var } \tilde{f}_n(x) < \infty.$$

Further, Van Ryzin (1969) shows that

$$\sum_{n=1}^{\infty} \frac{1}{(n+1)^2 h_{n+1}^{2m}} \text{ Var } K\left(\frac{x-X}{h_{n+1}}\right) < \infty$$

using condition (4.2.2).

Thus $\sum_{n=1}^{\infty} E|Y_n'| < \infty$ and the result follows from Lemma 4.2.1 and the fact that the mean square limit and the almost sure limit coincide with probability one.

In order to prove the uniform result with respect to x, we introduce the Fourier transform of the kernel. Define

(4.2.5) $$k(t) = k(t_1, \cdots, t_m) = \int e^{it'u} K(u)\, du$$

where $$t'u = \sum_{j=1}^{m} t_j u_j.$$

Also let $$\phi_n(t) = \frac{1}{n} \sum_{j=1}^{m} e^{it'X_j}$$

and let $$\phi(t) = Ee^{it'X}.$$

Theorem 4.2.3: (Uniform Strong Consistency)

If K(u) satisfies conditions (4.1.4) and $\{h_n\}$ is a monotone decreasing sequence of real numbers satisfying (2.1.4) and if both

$\lim_{n\to\infty} \frac{h_n}{h_{n+1}} = 1$ and $\lim_{n\to\infty} nh_n^{2m} \to \infty$ and if

(4.2.6) $$\sum_{n=1}^{\infty} \frac{1}{(nh_n^m)^2} < \infty$$

and

(4.2.7) $$\sum_{n=1}^{\infty} \frac{1}{nh_n^{2m-1}} \left| \frac{1}{h_{n+1}} - \frac{1}{h_n} \right| < \infty$$

and if also

(4.2.8) $$\int | k(t) | \, dt < \infty$$

where k(u) is defined by (4.2.5) and $|k(u)|$ is non-decreasing on $u < 0$ and non-increasing on $u \geq 0$, then if f(x) is uniformly continuous on R^m, $\sup_x |\tilde{f}_n(x) - f(x)| \to 0$ with probability one as $n\to\infty$.

Proof: (All integrals are over R^m unless otherwise specified).

We first prove $\sup_x |\tilde{f}_n(x) - E\tilde{f}_n(x)| \to 0$ a.s. as $n\to\infty$. Since both k(u) and K(u) are in L_1, using the inversion theorem for a Fourier Transform

$$\sup_x |\tilde{f}_n(x) - E\tilde{f}_n(x)| = \sup_x \left| \frac{1}{(2\pi)^m} \int \frac{1}{n} \sum_{j=1}^{n} \left(e^{iu'X_j} - \phi(u)\right) k(h_j u)\, e^{-iu'x} du \right|$$

Taking the absolute value signs inside the integral and noting $|e^{-iu'x}| \leq 1$, we now have that the right hand side is less than

$$\frac{1}{(2\pi)^m} \int \left| \frac{1}{n} \sum_{j=1}^{n} \left(e^{iu'X_j} - \phi(u)\right) k(h_j u) \right| du$$

and we can write this as

$$\frac{1}{(2\pi)^m} \int \left| \frac{1}{n} \sum_{j=1}^{n} \left(e^{iu'X_j} - \phi(u)\right) \frac{k(h_j u)}{k(h_n u)} \right| \left| k(h_n u) \right| du.$$

Using Schwarz's inequality this is less than

$$\frac{1}{(2\pi)^m} \left(\int |k(h_n u)| du \right)^{1/2} \left(\int \left| \frac{1}{n} \sum_{j=1}^{n} \left(e^{iu'X_j} - \phi(u)\right) \frac{k(h_j u)}{k(h_n u)} \right|^2 |k(h_n u)| \, du \right)^{1/2}$$

$$= \frac{1}{(2\pi)^m} \left(\int \frac{1}{h_n^m} |k(u)| \, du \right)^{1/2} \left(\int \left| \frac{1}{n} \sum_{j=1}^{n} \left(e^{iu'X_j} - \phi(u)\right) \frac{k(h_j u)}{k(h_n u)} \right|^2 |k(h_n u)| du \right)^{1/2}$$

Since $\int |k(u)| \, du < \infty$ by assumption (4.2.8), we need only consider the term

$$\int \frac{1}{h_n^m} \left| \frac{1}{n} \sum_{j=1}^{n} \left(e^{iu'X_j} - \phi(u)\right) \frac{k(h_j u)}{k(h_n u)} \right|^2 \left| k(h_n u) \right| du$$

which we will denote by Y_n. Taking expectations we get

$$EY_n = \int \frac{1}{h_n^m} E \left| \frac{1}{n} \sum_{j=1}^{n} \left(e^{iu'X_j} - \phi(u)\right) \frac{k(h_j u)}{k(h_n u)} \right|^2 \left| k(h_n u) \right|^2 du,$$

the interchange of the integral and expectation sign being justified by the use of Fubini's Theorem for positive functions. But,

$$E \left| e^{iu'X_j} - \phi(u) \right|^2 = E\left(e^{iu'X_j} - \phi(u)\right) \left(e^{-iu'X_j} - \overline{\phi(u)}\right)$$

$$= 1 - |\phi(u)|^2$$

so that we obtain

$$\text{(4.2.9)} \quad EY_n = \frac{1}{n^2 h_n^m} \int \left(1 - |\phi(u)|^2\right) \sum_{j=1}^{n} \left| \frac{k(h_j u)}{k(h_n u)} \right|^2 \left| k(h_n u) \right| du.$$

However, since $|k(u)|$ is non-decreasing on $u < 0$ and non-increasing on $u \geq 0$ we have

$$\text{(4.2.10)} \qquad \left| \frac{k(h_j u)}{k(h_n u)} \right| \leq 1 \quad , \; j \leq n.$$

Thus, using inequality (4.2.10) in equation (4.2.9) we get

$$EY_n \le \frac{1}{nh_n^m} \int |k(h_n u)\ du| = \frac{1}{nh_n^{2m}} \int |k(u)|\ du.$$

Condition (4.2.8) and the assumption $nh_n^{2m} \to \infty$ ensure that EY_n tends to zero as $n\to\infty$. We now prove that the convergence of Y_n occurs with probability one.

Write,

$$Y_{n+1} = \frac{1}{h_{n+1}^m} \int \frac{1}{(n+1)^2} \left| \sum_{j=1}^{n+1} \left(e^{iu'X_j} - \phi(u)\right) \frac{k(h_j u)}{k(h_{n+1}u)} \right|^2 |k(h_{n+1}u)|\ du$$

$$= \frac{1}{(n+1)^2} \frac{1}{h_{n+1}^m} \int \left| \sum_{j=1}^{n} \left(e^{iu'X_j} - \phi(u)\right) \frac{k(h_j u)}{k(h_n u)} + e^{iu'X_{n+1}} - \phi(u) \right|^2 |k(h_{n+1}u)|du.$$

Let F_n be defined as in Theorem 4.2.2. We now expand the first term in the integral using the fact that $|Z|^2 = Z\bar{Z}$ and that

$E\left(e^{iu'X_{n+1}} - \phi(u)\right)|F_n) = 0$ and $E\left(e^{iu'X_{n+1}} - \phi(u) \mid F_n\right) = 0.$

Thus,

$$E\left(Y_{n+1}|F_n\right) = \frac{1}{(n+1)^2} \frac{1}{h_{n+1}^m} \int \left| \sum_{j=1}^{n} \left(e^{iu'X_j} - \phi(u)\right) \frac{k(h_j u)}{k(h_n u)} \right|^2 |k(h_{n+1}u)|\ du$$

$$+ \frac{1}{h_{n+1}^m (n+1)^2} \int \left(1 - |\phi(u)|^2\right) |k(h_{n+1}u)|\ du.$$

But, $1 - |\phi(u)|^2 \le 1$ and $|k(u)| \le 1$ since we assume $K(u)$ is a density function. Thus

$$E\left(Y_{n+1}|F_n\right) \le \frac{1}{n^2 h_{n+1}^m} \int \left| \sum_{j=1}^{n} \left(e^{iu'X_j} - \phi(u)\right) \frac{k(h_j u)}{k(h_n u)} \right|^2 |k(h_n u)|\ du$$

$$+ \frac{1}{n^2 h_{n+1}^{2m}} \int |k(u)| \, du$$

Adding and subtracting Y_n we obtain

$$E\left(Y_{n+1} | F_n\right) \leq Y_n - \frac{1}{n^2}\left(\frac{1}{h_n^m} - \frac{1}{h_{n+1}^m}\right) \int \left| \sum_{j=1}^{n} \left(e^{iu'X_j} - \phi(u)\right) \frac{k(h_j u)}{k(h_n u)} \right|^2 |k(h_n u)| \, du$$

$$+ \frac{1}{n^2 h_{n+1}^{2m}} \int |k(u)| \, du$$

Let,

$$W_n = \frac{1}{n^2 h_{n+1}^{2m}} \int |k(u)| du \tag{4.2.11}$$

and

$$U_n = \frac{1}{n^2}\left(\frac{1}{h_n^m} - \frac{1}{h_{n+1}^m}\right) \int \left| \sum_{j=1}^{n} \left(e^{iu'X_j} - \phi(u)\right) \frac{k(h_j u)}{k(h_n u)} \right|^2 |k(h_n u)| \, du. \tag{4.2.12}$$

Note that W_n is not a random variable. Hence $EW_n = W_n$, and using condition(4.2.6), we have

$$\sum_{n=1}^{\infty} E|W_n| = \sum_{n=1}^{\infty} |W_n| < \infty. \tag{4.2.13}$$

Thus, in order to use lemma 4.2.1 we need to show that $\sum_{n=1}^{\infty} |U_n| < \infty$.

From (4.2.12) we have on taking expectations

$$E|U_n| = \frac{1}{n^2}\left|\frac{1}{h_n^m} - \frac{1}{h_{n+1}^m}\right| \int \left(1 - |\phi(u)|^2\right) \sum_{j=1}^{n} \left|\frac{k(h_j u)}{k(h_n u)}\right|^2 |k(h_n u)| \, du \tag{4.2.14}$$

$$\leq \frac{1}{n^2} \left| \frac{1}{h_n^m} - \frac{1}{h_{n+1}^m} \right| \frac{1}{h_n^m} \int n \, |k(u)| \, du.$$

By using the fact that $(1 - t^m) = \left(\sum_{j=1}^{m} t^{j-1} \right) (1-t)$ and using the condition that $\frac{h_n}{h_{n+1}} \to 1$ as $n \to \infty$, Van Ryzin (1969) shows

$$\left| \frac{1}{h_{n+1}^m} - \frac{1}{h_n^m} \right| \sim \frac{m}{h_n^{m-1}} \left| \frac{1}{h_{n+1}} - \frac{1}{h_n} \right|.$$

Thus the upper bound in (4.2.14) is asymptotically equivalent to $\frac{m}{nh_n^{2m-1}} \left| \frac{1}{h_{n+1}} - \frac{1}{h_n} \right| \int |k(u)| \, du$. Hence, using condition (4.2.7), $\sum_{n=1}^{\infty} E|U_n| < \infty$. Thus conditions (iii) and (iv) of lemma (4.2.1) are satisfied, where $Y_n' = U_n + W_n$.

The conditions of lemma 4.2.1 are satisfied for Y_n and so Y_n tends to zero almost everywhere. Thus,

$$\sup_x |\tilde{f}_{n+1}(x) - E\tilde{f}_{n+1}(x)| \to 0 \quad \text{a.s. as } n \to \infty.$$

Now, using the definition of $\tilde{f}_n(x)$ and noting $\int K(u)\, du = 1$, we have that

$$\sup_x |E\tilde{f}_n(x) - f(x)| = \sup_x \left| \int \left(\frac{1}{n^2} \sum_{j=1}^{n} K_j(u) \, f(x-u) - \frac{1}{n} \sum_{j=1}^{n} K_j(u) \, f(x) \, du \right| \right.$$

where $K_j(u) = \frac{1}{h_n} K\left(\frac{x-X_j}{h_n}\right)$. Then the right hand side is less than

$$(4.2.15) \qquad \sup_x \int \frac{1}{n} \left| \sum_{j=1}^{n} K_j(u) \right| \left| f(x-u) - f(x) \right| du$$

and splitting the range of integration into two sets $||u|| \le \delta$ and $||u|| > \delta$ we have that (4.2.15) is less than

$$\sup_x \int_{||u||\le\delta} \frac{1}{n} \left|\sum_{j=1}^{n} K_j(u)\right| \left|f(x-u) - f(x)\right| du$$
$$+ \sup_x \int_{||u||>\delta} |f(x-u) - f(x)| \frac{1}{n} \sum_{j=1}^{n} K_j(u)\, du$$

$$\le \sup_x \sup_{||u||\le\delta} |f(x-u) - f(x)| + 2 \sup_x \frac{1}{n} f(x) \sum_{j=1}^{n} \int_{||u||>\frac{\delta}{h_j}} K(u)\, du.$$

Since f(x) is uniformly continuous, the first term can be made arbitrarily small by choosing δ sufficiently small. In the second term, notice that for this fixed δ,

$$f(x) \int_{||u||>\frac{\delta}{h_n}} K(u)\, du \to 0$$

as $n\to\infty$ at all points x. Hence the Cesàro sum approaches zero and the proof is complete.

Example 4.2.1

Consider the $\{h_n\}$ sequence given by

$$h_n = \frac{1}{(\log n)^{\frac{1}{m}}}$$

This sequence satisfies conditions (2.1.4) and in addition $\lim_{n\to\infty} \frac{h_n}{h_{n+1}} = 1$. Also

$$\sum_{j=1}^{\infty} \frac{1}{n^2 h_n^m} = \sum_{n=1}^{\infty} \frac{\log n}{n^2} \le \sum_{n=1}^{\infty} \frac{1}{n^{2-r}} \quad , 0 < r < 1$$

which is finite so that condition (4.2.2) is also satisfied.

Hence the sequence $h_n = \frac{1}{(\log n)^{\frac{1}{m}}}$ satisfies the conditions of Theorem 4.2.2. Let us now evaluate the constant we have called ν_0, defined by (4.2.3). From the definition of ν_0,

$$\nu_0 = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \frac{h_n^m}{h_j^m} = \lim_{n\to\infty} \left(\frac{1}{n \log n} \sum_{j=1}^{n} \log j \right).$$

Let $y = \log x$, so that

$$\int_1^{n+1} \log x \, dx \geq \sum_{j=1}^{n} \log j \geq \int_2^{n+1} \log(x-1) \, dx.$$

That is, on integration by parts

$$(n+1) \log(n+1) - n \geq \sum_{j=1}^{n} \log j \geq (n+1) \log n - (n-1) - \log n.$$

Hence, using these inequalities we see that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \frac{h_n^m}{h_j^m} = 1.$$

That is for the sequence $h_n = \frac{1}{(\log n)^{\frac{1}{m}}}$ the conditions of Theorem 4.2.2 are satisfied and $\nu_0 = 1$.

Example 4.2.2

Consider the $\{h_n\}$ sequence defined by

$$h_n = n^{\frac{-r}{m}} \qquad (0 < r < \tfrac{1}{2}).$$

This sequence satisfies (2.1.4) and $\lim_{n\to\infty} \frac{h_n}{h_{n+1}} = 1$. Also we have that $\sum_{j=1}^{\infty} \frac{1}{n^2 h_n^m} < \infty$. Thus this sequence satisfies the conditions of Theorem 4.2.2. In addition, using a method similar to that in example 4.2.1 we can show that

$$\nu_0 = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \frac{h_n^m}{h_j^m} = \frac{1}{r+1} \qquad (0 < r < \tfrac{1}{2}).$$

Now clearly $nh_n^{2m} \to \infty$ as $n\to\infty$ and

$$\sum_{n=1}^{\infty} \frac{1}{(nh_n^m)^2} = \sum_{n=1}^{\infty} \frac{1}{(n^{1-r})^2} < \infty \text{ since } 0 < r < \frac{1}{2}$$

But,

$$\sum_{n=1}^{\infty} \frac{1}{nh_n^{2m-1}} \left|\frac{1}{h_{n+1}} - \frac{1}{h_n}\right| = \sum_{n=1}^{\infty} \frac{1}{n.n^{-2r+\frac{r}{m}}} \left((n+1)^{\frac{r}{m}} - n^{\frac{r}{m}}\right)$$

$$= \sum_{n=1}^{\infty} \frac{1}{n^{1-2r}} \left((1+\frac{1}{n})^{\frac{r}{m}} - 1\right)$$

and,

$$(1+\frac{1}{n})^{\frac{r}{m}} - 1 \le \frac{r}{mn} \qquad \text{since } \frac{r}{m} < 1.$$

Thus,

$$\sum_{n=1}^{\infty} \frac{1}{n^{1-2r}} \left((1+\frac{1}{n})^{\frac{r}{m}} - 1\right) \le \frac{r}{m} \sum_{n=1}^{\infty} \frac{1}{n^{2-2r}} < \infty \qquad \text{since } 0 < r < \frac{1}{2}.$$

Hence, the sequence $h_n = n^{-\frac{r}{m}}$ also satisfies the conditions of Theorem 4.2.3.

We will now return to the case when $m = 1$ and consider a "naive" sequential procedure for the Yamato estimator.

4.3 The Sequential Procedure:

We will now consider $\tilde{f}_n(x)$ defined as in (4.1.1) but with $m = 1$, that is

$$\tilde{f}_n(x) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{h_j} K\left(\frac{x-X_j}{h_j}\right). \tag{4.3.1}$$

In section 4.2, Theorems 4.2.2 and 4.2.3 show that $\tilde{f}_n(x)$ is both pointwise and uniformly strongly consistent under appropriate conditions so that as in the case of Parzen type estimators a "naive" sequential procedure can be defined. Again the procedure consists of taking successive samples of size M consisting of M independently and identically distributed random variables from a population with density function $f(x)$. $\tilde{V}_n(x)$ is defined by (4.1.5)

Since we can write, using a modification of (4.1.2)

$$\tilde{f}_{nM}(x) = \frac{n-1}{n} \tilde{f}_{(n-1)M}(x) + \frac{1}{nM} \sum_{j=(n-1)M+1}^{nM} \frac{1}{h_j} K\left(\frac{x-X_j}{h_j}\right)$$

we see that $\tilde{V}_n(x)$ can be written as

$$\tilde{V}_n(x) = \tilde{f}_{nM}(x) - \tilde{f}_{(n-1)M}(x)$$

$$= \frac{(n-1)}{n} \tilde{f}_{(n-1)M}(x) - \tilde{f}_{(n-1)M}(x) + \frac{1}{nM} \sum_{j=(n-1)M+1}^{nM} \frac{1}{h_j} K\left(\frac{x-X_j}{h_j}\right)$$

so that

$$\tilde{V}_n(x) = -\frac{1}{n} \tilde{f}_{(n-1)M}(x) + \frac{1}{nM} \sum_{j=(n-1)M+1}^{nM} \frac{1}{h_j} K\left(\frac{x-X_j}{h_j}\right) \tag{4.3.2}$$

Formally, the stopping rule we will consider is defined by,

(4.3.3)
$$N_1(\varepsilon,M) = \begin{cases} \text{Ist } n \text{ such that } |\tilde{V}_n(x)| < \varepsilon \text{ for given } \varepsilon > 0, \\ \infty \text{ if no such } n \text{ exists} \end{cases}$$

We will assume that conditions (2.1.3) and (2.1.4) hold for $K(u)$ and the sequence $\{h_n\}$ respectively. Paralleling our previous discussion for the Parzen-type estimators, $\hat{f}_n(x)$, we now discuss some of the properties of $\tilde{f}_{N_1}(x)$ and $N_1(\varepsilon,M)$.

4.4 Properties:

Lemmas 4.4.1 and 4.4.2 correspond to lemmas 2.2.1 and 2.2.2 and give elementary properties of $\tilde{V}_n(x)$ that allow it's use in the stopping rule (4.3.3).

Lemma 4.4.1:

As $n\to\infty$, if $K(u)$ and $\{h_n\}$ satisfy conditions (2.1.3) and (2.1.4) and if h_n is monotone decreasing then

(a) $|\tilde{V}_n(x)| \to 0$ in probability if x is a continuity point of $f(x)$ and

(b) $\sup_x |\tilde{V}_n(x)| \to 0$ in probability if $f(x)$ is uniformly continuous.

Proof:

Yamato (1972) shows that given the above conditions

(a) $\tilde{f}_n(x) \to f(x)$ in probability and (b) $\sup_x |\tilde{f}_n(x) - f(x)| \to 0$

in probability. Thus from the definition of $\tilde{V}_n(x)$ we have that $|\tilde{V}_n(x)| = |\tilde{f}_{nM}(x) - \tilde{f}_{(n-1)M}(x)|$ is a Cauchy sequence in probability and part. (a) follows. Similarly, (b) $\sup_x |V_n(x)| = \sup_x |\tilde{f}_{nM}(x) - \tilde{f}_{(n-1)M}(x)|$.

This is also a Cauchy sequence under the given conditions. Thus the lemma is proved.

Lemma 4.4.2

If K(u) and the sequence $\{h_n\}$ satisfy the conditions of theorem 4.2.2 then $\tilde{V}_n(x) \to 0$ with probability one if x is a continuity point of f(x).

Proof:

Under the conditions stated $\tilde{f}_n(x)$ is a Cauchy sequence with probability one so the result follows.

Lemma 4.4.3

As $n\to\infty$, if K(u) and $\{h_n\}$ satisfy the conditions of Theorem 4.2.3 then

$$\sup_x |\tilde{V}_n(x)| \to 0 \text{ with probability one if } f(x) \text{ is uniformly continuous.}$$

Proof:

The result follows since under the conditions stated $\tilde{f}_n(x)$ is uniformly Cauchy.

In order to examine the properties of the stopping rule we must first study the properties of $\tilde{V}_n(x)$. We begin an examination of the properties of $\tilde{V}_n(x)$ by examining its variance.

Lemma 4.4.4

If K(u) and $\{h_n\}$ satisfy conditions (2.1.3) and (2.1.4) and if in addition h_n is monotone decreasing then we can write

$$\tilde{V}_n(x) = \tilde{A}_n(x) + \tilde{B}_n(x)$$

and

$$\lim_{n\to\infty} n^2 Mh_{nM} \text{ Var } \tilde{V}_n(x) = \lim_{n\to\infty} n^2 Mh_{nM} \text{ Var } \tilde{B}_n(x)$$

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

Proof:

From equation (4.3.2) we have that

$$\tilde{A}_n(x) = -\frac{1}{n}\tilde{f}_{(n-1)M}(x)$$

and

$$\tilde{B}_n(x) = \frac{1}{nM}\sum_{j=(n-1)M+1}^{nM} \frac{1}{h_j} K\left(\frac{x-X_j}{h_j}\right).$$

Further, since $\tilde{A}_n(x)$ depends only on $X_1, X_2, \cdots, X_{(n-1)M}$ and $\tilde{B}_n(x)$ depends only on $X_{(n-1)M+1}, \cdots, X_{nM}$, $\tilde{A}_n(x)$ and $\tilde{B}_n(x)$ are independent. Hence we can write

$$(4.4.1)\quad \text{Var } \tilde{V}_n(x) = \frac{1}{n^2} \text{Var } \tilde{f}_{(n-1)M}(x) + \frac{1}{(nM)^2}\sum_{j=(n-1)M+1}^{nM} \frac{1}{h_j^2} \text{Var } K\left(\frac{x-X_j}{h_j}\right)$$

In the last expression on the right hand side of (4.4.1) we notice that $(n-1)M+1 \le j \le nM$ so that $j\to\infty$ as $n\to\infty$ and so using Parzen (1962)

$$\lim_{n\to\infty} \frac{1}{h_j} \text{Var } K\left(\frac{x-X_j}{h_j}\right) = f(x)\int_{-\infty}^{\infty} K^2(u)\, du, \qquad (n-1)M < j \le nM.$$

But, by assumption $\lim_{n\to\infty} \frac{h_{nM}}{h_j} = 1 \qquad (n-1)M < j \le nM$

so that

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

That is,

$$\lim_{n\to\infty} n^2 M h_{nM} \text{Var } \tilde{B}_n(x) = f(x)\int_{-\infty}^{\infty} K^2(u)\, du.$$

Further, Yamato (1972) shows that

$$\lim_{n\to\infty} \text{Var } \tilde{f}_n(x) = 0$$

so that

$$\lim_{n\to\infty} n^2 Mh_{nM} \tilde{A}_n(x) = \lim_{n\to\infty} Mh_{nM} \text{ Var } \tilde{f}_{nM}(x) = 0$$

Hence,

$$\lim_{n\to\infty} n^2 Mh_{nM} \text{ Var } \tilde{V}_n(x) = 0 + \lim_{n\to\infty} n^2 Mh_{nM} \text{ Var } \tilde{B}_n(x)$$

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

This proves the lemma.

Remarks:

(1) As in the Parzen case, $\tilde{B}_n(x)$ is a finite sum, dependent only on the final sample which consists of M independent random variables. $\tilde{B}_n(x)$ is always positive.

(2) If the density function, $f(x)$, were known then under appropriate conditions on the kernel, $K(u)$, the exact distribution of $\tilde{B}_n(x)$ could be found (at least in theory). It would be more complicated than in the Parzen case since we no longer have M identically distributed random variables.

(3) $n\tilde{A}_n(x) = \tilde{f}_{(n-1)M}(x)$ will have an asymptotic normal distribution since $\tilde{f}_{(n-1)M}(x)$ is distributed normally asymptotically. However,

$$\text{Var } \tilde{A}_n(x) \sim \frac{1}{n^3 Mh_{nM}} \int_{-\infty}^{\infty} K^2(u)\, du.$$

(4) Some comparison between the stopping rules N and N_1 can be

made. At this point we can compare the variances of $V_n(x)$ and $\tilde{V}_n(x)$ using lemmas 4.4.4 and (2.2.5) and we show that the variance of $\tilde{V}_n(x)$ is always less than or equal to the variance of $V_n(x)$.

Lemma 4.4.5

Given the conditions of lemmas (4.4.4) and (2.2.4) then there exists a constant ν defined by (2.2.2) such that

$$\lim_{n\to\infty} \frac{\operatorname{Var} \tilde{V}_n(x)}{\operatorname{Var} V_n(x)} = \frac{1}{\nu} \le 1.$$

Proof:

From lemmas 4.4.4 and 2.2.4 we get that

$$\lim_{n\to\infty} \frac{\operatorname{Var} \tilde{V}_n(x)}{\operatorname{Var} V_n(x)} = \lim_{n\to\infty} \frac{n^2 Mh_{nM} \operatorname{Var} \tilde{V}_n(x)}{n^2 Mh_{nM} \operatorname{Var} V_n(x)}$$

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

But in lemma 2.2.4 we saw $\nu \ge 1$ so the result follows.

Example 4.4.1:

Let $h_n = B\, n^{-\alpha}$, $0 < \alpha < 1$.

In example 2.2.1 we showed that

$$\lim_{n\to\infty} \left(n + \frac{n^2}{(n-1)} \frac{h_{nM}}{h_{(n-1)M}} - 2n \frac{h_{nM}}{h_{(n-1)M}} \right) = 1 + \alpha$$

while in example 4.2.2 (with $m = 1$) we showed that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \frac{h_n}{h_j} = \frac{1}{1 + \alpha}$$

and we note that this last quantity is exactly the result of lemma 4.4.5.

We thus have a link between Parzen type and Yamato type estimators. Notice that in the variance of $V_n(x)$, the $A_n(x)$ term gives the term $\alpha f(x) \int_{\infty}^{\infty} K^2(u)\,du$ while $B_n(x)$ gives the term $f(x) \int K^2(u)\,du$, while for $\tilde{V}_n(x)$ the $\tilde{A}_n(x)$ term leads to a zero.

4.5 An Alternative Procedure:

In this section we will examine briefly an intuitively appealing, but unfortunately impractical stopping rule. We have seen that

$$\tilde{V}_n(x) = \frac{(n-1)}{n} \tilde{f}_{(n-1)M}(x) - \tilde{f}_{(n-1)M}(x) + \frac{1}{nM} \sum_{j=(n-1)M+1}^{nM} \frac{1}{h_j} K\left(\frac{x-X_j}{h_j}\right)$$

Notice that if we should define

$$\tilde{V}'_n(x) = \tilde{f}_{nM}(x) - \frac{(n-1)}{n} \tilde{f}_{(n-1)M}(x)$$

then on simplification

$$\tilde{V}'_n(x) = \frac{1}{nM} \sum_{j=(n-1)M+1}^{nM} \frac{1}{h_j} K\left(\frac{x-X_j}{h_j}\right)$$

which depends only on the last sample of M observations. That is we could then define

$$N_2(\varepsilon,M) = \begin{cases} \text{1st } n \text{ such that } \tilde{V}'_n(x) < \varepsilon, \text{ given } \varepsilon > 0. \\ \infty \text{ if no such } n \text{ exists.} \end{cases}$$

Lemma 4.5.1 shows that this stopping rule is unsuitable for use as a sequential stopping rule.

Lemma 4.5.1

If $K(u)$ and $\{h_n\}$ satisfy (2.1.3) and (2.1.4) then $N_2(\varepsilon,M)$ is bounded above. Further this bound does not depend on the convergence properties of $\tilde{f}_n(x)$ (thus making $N_2(\varepsilon,M)$ an unsuitable stopping rule).

Proof: By definition

$$\tilde{V}_n'(x) = \frac{1}{nM} \sum_{j=(n-1)M+1}^{nM} \frac{1}{h_j} K\left(\frac{x-X_j}{h_j}\right)$$

$$\leq \frac{M}{nMh_{nM}} \sup_u K(u)$$

$$\to 0 \quad \text{as} \quad n\to\infty.$$

Thus $N_2(\varepsilon,M)$ is always less than or equal to the smallest n that satisfies

$$\frac{\sup_u K(u)}{nh_{nM}} \leq \varepsilon$$

and this value does not depend on the convergence properties of $\hat{f}_n(x)$.

4.6 Finiteness of EN:

We now return to the stopping rule defined by (4.3.3) and will show that $EN_1 < \infty$. First, we will state two lemmas without proof, the proofs being similar to those of lemmas 2.3.1 and 2.3.2.

Lemma 4.6.1

For arbitrary $t > 0$,

$$P(V_n(x) > \varepsilon) \leq e^{-nMh_{nM}\varepsilon t} Ee^{\tilde{S}_n(x)t} \tag{4.6.1}$$

where

$$\tilde{S}_n(x) = \sum_{j=1}^{nM} \frac{h_{nM}}{h_j} K\left(\frac{x-X_j}{h_j}\right) - \sum_{j=1}^{(n-1)M} \frac{n}{(n-1)} \frac{h_{nM}}{h_j} K\left(\frac{x-X_j}{h_j}\right) \tag{4.6.2}$$

Lemma 4.6.2

For arbitrary $t > 0$,

(4.6.3) $$P(\tilde{V}_n(x) < -\varepsilon) \leq e^{-nMh_{nM}\varepsilon} Ee^{-t\tilde{S}_n(x)}$$

where $\tilde{S}_n(x)$ is defined by (4.6.2).

Since t is arbitrary we will consider t = 1 only, in the following lemmas. We also note at this stage that we may write by using (4.3.2) that

$$\tilde{S}_n(x) = \sum_{j=(n-1)M+1}^{nM} \frac{h_{nM}}{h_j} K\left(\frac{x-X_j}{h_j}\right) - h_{nM}\, M\, \tilde{f}_{(n-1)M}(x)$$

so that

(4.6.4) $$\tilde{S}_n(x) = \sum_{j=(n-1)M+1}^{nM} \frac{h_{nM}}{h_j} K\left(\frac{x-X_j}{h_j}\right) - \frac{1}{(n-1)} \sum_{j=1}^{(n-1)M} \frac{h_{nM}}{h_j} K\left(\frac{x-X_j}{h_j}\right)$$

Lemma 4.6.3

If $\{h_n\}$ is a monotone decreasing sequence of real numbers satisfying conditions (2.1.4), (2.1.5) and condition (A1) and if K(u) satisfies (2.1.3), and if in addition ν_0 is defined by (4.2.3) then

$$Ee^{+\tilde{S}_n(x)} \leq e^{ML}$$

where $L = \sup_u K(u)$

Proof:

From (4.6.4) we have that

$$Ee^{\tilde{S}_n(x)} = E\exp\left(\sum_{j=(n-1)M+1}^{nM} \frac{h_{nM}}{h_j} K\left(\frac{x-X_j}{h_j}\right)\right) E\exp\left(-\frac{1}{(n-1)} \sum_{j=1}^{(n-1)M} \frac{h_{nM}}{h_j} K\left(\frac{x-X_j}{h_j}\right)\right)$$

But, since $\frac{h_{nM}}{h_j} \leq 1$ for $j \leq nM$ we get that

(4.6.5) $$\sum_{k=(n-1)M+1}^{nM} \frac{h_{nM}}{h_j} K\left(\frac{x-X_j}{h_j}\right) \leq M.L$$

Further, since $K(u) \geq 0$ we have that

(4.6.6) $$\exp\left(- \frac{1}{(n-1)} \sum_{j=1}^{(n-1)M} \frac{h_{nM}}{h_j} K\left(\frac{x-X_j}{h_j}\right)\right) \leq 1$$

Thus, using inequalities (4.6.5) and (4.6.6) we have that

$$Ee^{\tilde{S}_n(x)} \leq Ee^{ML} E(1) = e^{ML}.$$

Lemma 4.6.4

Under the conditions of Lemma 4.6.3

$$Ee^{-\tilde{S}_n(x)} \leq e^{L}.$$

Proof: From (4.6.4) we have

$$Ee^{-\tilde{S}_n(x)} = E\exp\left(- \sum_{j=(n-1)M+1}^{nM} \frac{h_{nM}}{h_j} K\left(\frac{x-X_j}{h_j}\right)\right) E\exp\left(\frac{1}{(n-1)} \sum_{j=1}^{(n-1)M} \frac{h_{nM}}{h_j} K\left(\frac{x-X_j}{h_j}\right)\right)$$

Since $K(u) \geq 0$ for all u the first expection on the right hand side is less than one. Now,

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

since $\frac{1}{(n-1)} \sum_{j=1}^{(n-1)M} \frac{h_{nM}}{h_j} \leq 1$

Thus we have that

$$Ee^{-\tilde{S}_n(x)} \leq 1. \quad Ee^{L} = e^{L} \qquad \text{as required.}$$

Theorem 4.6.5:

If $K(u)$ and the sequence $\{h_n\}$ satisfy the conditions of lemma 4.6.3 then $EN_1 < \infty$.

Proof:

Using a method similar to that used in the Parzen case we can show

$$P(N_1 > nM) \le P(\tilde{V}_n(x) > \varepsilon) + P(\tilde{V}_n(x) < -\varepsilon),$$

and using lemma 4.6.2 together with lemmas 4.6.3 and 4.6.4 we get

$$P(N_1 > nM) \le C\, e^{-nMh_{nM}\varepsilon}$$

where $C = e^{ML} + e^{L}$. But,

$$EN_1 = \sum_{k=0}^{\infty} P(N_1 > kM)$$

$$\le \sum_{k=0}^{\infty} C\, e^{-kMh_{kM}\varepsilon}$$

which is finite since $\{h_n\}$ satisfies condition (A1). Thus the proof of the theorem is complete.

Lemma 4.6.6

If K(u) and the sequence $\{h_n\}$ satisfy the conditions of lemma 4.6.3 then Var $N_1 < \infty$.

Proof:

It will be sufficient to prove $EN_1^2 < \infty$ since we have already shown $EN_1 < \infty$.

We have from the definition of $N_1(\varepsilon, M)$ that

$$P(N_1 = nM) = P\left((\Lambda_n^1)^c \cap \left(\bigcap_{j=1}^{n-1} \Lambda_j^1\right)\right)$$

where

$$\Lambda_j^1 = \left\{(x_1, \ldots, x_{nM}) : |\tilde{V}_j(x)| \le \varepsilon\right\} \tag{4.6.7}$$

so that

$$P(N_1 = nM) \le P(\Lambda_{n-1}^1) = P\left(|\tilde{V}_n(x)| \le \varepsilon\right)$$

and using lemmas 4.6.3 and 4.6.4 we get that

$$P(N_1 = nM) \le C\, e^{-nMh_{nM}\varepsilon}.$$

Thus it follows that since $\{h_n\}$ satisfies condition (A1),

$$EN_1^2 = \sum_{k=1}^{\infty} C\, n^2\, e^{-nMh_{nM}\varepsilon} < \infty$$

and the proof is complete.

Observations:

1. As $\varepsilon \to 0$ the upper bound on EN_1 becomes infinite. This is what we would expect given the form of the stopping rule.

2. Given M, ε and the form of h_n we can calculate the upper bound of EN_1.

3. Notice that the Yamato estimator leads to a much simpler set of conditions to ensure the finiteness of the expected sample size than does the Parzen estimator. The main reason for this is the simpler form of equation (4.3.2) compared to that of equation (2.2.4A). This is one more advantage of working with $\tilde{f}_n(x)$ compared to $\hat{f}_n(x)$ when we wish to use sequential methods.

We now prove the closure of the sequential procedure.

Lemma 4.6.7 (Closure).

If K(u) and $\{h_n\}$ satisfy the conditions of lemma 4.6.3 then $P(N_1 < \infty) = 1$.

Proof:

Under the stated conditions we have that

$$P(N_1 > nM) \le C.\, e^{-nMh_{nM}\varepsilon}.$$

so the result is immediate.

4.7. Divergence of N_1 as $\varepsilon \to 0$:

We will now examine the stopping rule's behaviour when we let $\varepsilon \to 0$. As for the stopping rule $N(\varepsilon,M)$ defined in chapter 2 we would like to show that

$$\lim_{\varepsilon\to 0} N_1(\varepsilon,M) \quad \left| \begin{array}{l} \text{in probability} \\ \text{with probability one,} \end{array} \right.$$

and as in that case it can be shown that this result will not be true for all kernels. We state a lemma (that corresponds to lemma 2.6.1) without proof.

Lemma 4.7.1

If $\xi^1_{nM} = \{(x_1, x_2, \circ\circ\circ, x_{nM}) \; ; \; |V_j(x)| = 0 \text{ for some } j\}$ then as $\varepsilon \to 0$, $P(N_1 \leq nM) \to P(\xi^1_{nM})$.

Again we note that $P(\xi^1_{nM})$ may be greater than zero (the uniform kernel again provides an example) but that a class of kernels, K^1_0, exists for which $P(\tilde{V}_n(x) = 0) = 0$.

In fact $K \varepsilon K^1_0$ if it satisfies the conditions of lemma 2.7.3. That is if $K(u)$ satisfies (2.1.3) and

(i) $K(u)$ is differentiable for all u, and

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

Thus lemma's 2.6.2 and 2.6.3 will hold if we replace $V_n(x)$ by $\tilde{V}_n(x)$ and N by N_1 in their statements. Similarly Theorem 2.6.4 will hold with $V_n(x)$, $\hat{f}_n(x)$ and $N(\varepsilon,M)$ replaced by $\tilde{V}_n(x)$, $\tilde{f}_n(x)$ and $N_1(\varepsilon,M)$ respectively. That is, we can show under the conditions stated in the lemmas and theorem that

$$\lim_{n\to\infty} N_1(\varepsilon,M) = \infty \begin{cases} \text{in probability and} \\ \text{with probability one.} \end{cases}$$

We can also redefine $N_1(\varepsilon,M)$ in the same manner $N(\varepsilon,M)$ was redefined in order to ensure that as $\varepsilon\to 0$ $N\to\infty$ at least in probability. Thus we define

$$N_1'(\varepsilon,M) = \begin{cases} \text{1st } n \text{ such that } |\tilde{V}_n(x)| < \varepsilon \text{ and } |\tilde{V}_n(x)| > 0 \\ \infty \text{ if no such } n \text{ exists} \end{cases}$$

In this case, by using a similar method to that used in Theorem 2.6.6 we can show

$$EN_1' < \infty.$$

4.8. Mean Square Error:

One of the properties we wish to look at for $\tilde{f}_{N_1}(x)$, is mean square error. We wish to show that as $\varepsilon\to 0$ that mean square error tends to zero. This is done in Theorem 4.8.1

Theorem 4.8.1:

If $K\varepsilon\, K_0^1$ then $\lim_{\varepsilon\to 0} E[\tilde{f}_{N_1}(x) - f(x)]^2 = 0$ provided that the conditions of Theorem 4.2.2 are satisfied.

Proof:

If $K\varepsilon\, K_0^1$ then by our discussion in section 4.7 we know that $\lim_{\varepsilon\to 0} N_1(\varepsilon,M) = \infty$ with probability one. But by Theorem 4.2.2, $\tilde{f}_n(x) \to f(x)$ a.s. and the result follows from Theorem 3.3.4.

CHAPTER 5

COMPLEMENTS

5.1. Introduction:

In this chapter we give some results that complement those in chapters two through four. We also try to give some avenues for future research.

5.2 Mean and Variance:

By definition, the mean of a random variable with density f(x) is $\int_{-\infty}^{\infty} x\, f(x)dx$. Thus, for a given set of observations $x_1, x_2, \cdots, x_n$ let us calculate the mean of $\hat{f}_n(x)$ as

$$\mu(x_1, \cdots, x_n) = \int_{-\infty}^{\infty} x\, f_n(x)dx.$$

Note that in this chapter K(u) is always assumed to satisfy (2.1.3) and the sequence $\{h_n\}$ to satisfy (2.1.4) and (2.1.5).

Lemma 5.2.1

For the estimator $\tilde{f}_n(x)$ defined by (4.1.1)

$$\tilde{\mu}(x_1, \cdots, x_n) = \bar{x} + \mu_K \cdot \frac{1}{n} \sum_{j=1}^{n} h_j$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ and $\mu_K = \int_{-\infty}^{\infty} u\, K(u)du$

Proof:

By the definition of $\tilde{f}_n(x)$, and since $x_1, x_2, \cdots, x_n$ is a given sample,

$$\int_{-\infty}^{\infty} x\, \tilde{f}_n(x)\,dx = \int_{-\infty}^{\infty} x\, \frac{1}{n} \sum_{j=1}^{n} \frac{1}{h_j} K\left(\frac{x-x_j}{h_j}\right) dx$$

$$= \frac{1}{n} \sum_{j=1}^{n} \int_{-\infty}^{\infty} \frac{x}{h_j} K\left(\frac{x-x_j}{h_j}\right) dx.$$

Transforming by $u = \dfrac{x-x_j}{h_j}$ we get

$$\tilde{\mu}(x_1, \cdots, x_n) = \frac{1}{n} \sum_{j=1}^{n} \int_{-\infty}^{\infty} (x_j + h_j u)\, K(u)\,du$$

and the result follows immediately.

Corollary 5.2.2:

For Parzen-Type estimators, $\hat{f}_n(x)$, $h_j = h_n$ for $j = 1, 2, \cdots, n$ so that

$$\hat{\mu}(x_1, \cdots, x_n) = \bar{x} + h_n\, \mu_K.$$

Corollary 5.2.3:

If $X_1, X_2, \cdots, X_n$ is a sample of independently and identically distributed random variables and if $K(u) = K(-u)$ then

$$E\, \tilde{\mu}(X_1, \cdots, X_n) = E\, \hat{\mu}(X_1, \cdots, X_n) = \mu$$

where $\mu = EX_1$.

Proof:

Since $K(u) = K(-u)$, $\mu_K = 0$ and the result is immediate.

Comment:

For the case $K(u) = K(-u)$ we have that $\mu_K = 0$ so that we can write $\hat{\mu}(x_1, \cdots, x_n) = \tilde{\mu}(x_1, \cdots, x_n) = \bar{x}$, and we have seen this is an unbiased estimate of μ. This gives us a reason for assuming the kernel is symmetric about zero.

Corollary 5.2.4:

As $n\to\infty$, $\tilde{\mu}(x_1,\cdots, x_n) \to \mu$ (if it exists)

Proof:

Clear.

Now let us calculate the variance of $\tilde{f}_n(x)$ for a given sample $x_1, x_2,\cdots, x_n$ as,

$$\tilde{\sigma}^2 (x_1,\cdots,x_n) = \int_{-\infty}^{\infty} x^2\tilde{f}_n(x)dx - \tilde{\mu}^2(x_1,\cdots, x_n)$$

Lemma 5.2.5

For Yamato's estimate, $\tilde{f}_n(x)$, given $x_1,\cdots, x_n$,

$$\tilde{\sigma}^2(x_1,\cdots, x_n) = s^2 + \frac{1}{n} \sum_{j=1}^{n} h_j^2 \mu_{2K} - \frac{\mu_K^2}{n^2} \left(\sum_{j=1}^{n} h_j\right)^2$$

where $s^2 = \frac{1}{n} \sum_{j=1}^{n} x_j^2 - \bar{x}^2$ and

$$\mu_{2K} = \int_{-\infty}^{\infty} u^2K(u)du.$$

Proof:

Let

$$\tilde{\mu}_2(x_1,\cdots, x_n) = \int_{-\infty}^{\infty} x^2 . \frac{1}{n} \sum_{j=1}^{n} \frac{1}{h_j} K\left(\frac{x-X_j}{h_j}\right) dx$$

$$= \frac{1}{n} \sum_{j=1}^{n} x_j^2 + \frac{2}{n} \sum_{j=1}^{n} x_jh_j \mu_K + \frac{1}{n} \sum_{j=1}^{n} h_j^2 \mu_{2,K}$$

after some simple calculations. Thus,

$$\tilde{\sigma}^2(x_1,\cdots, x_n) = \tilde{\mu}_2(x_1,\cdots,x_n) - \tilde{\mu}^2(x_1,\cdots, x_n)$$

$$= \frac{1}{n} \sum_{j=1}^{n} x_j^2 - \bar{x}^2 + \frac{1}{n} \sum_{j=1}^{n} h_j^2 \mu_{2K} - \frac{\mu_K^2}{n^2} \left(\sum_{j=1}^{n} h_j\right)^2$$

Corollary 5.2.6:

For $\hat{f}_n(x)$, since $h_j = h_n$ for $j = 1,2,\cdots, n$,

$$\hat{\sigma}^2(x_1,\cdots, x_n) = \frac{1}{n} \sum_{j=1}^{n} x_j^2 - \bar{x}^2 + h_n^2 \mu_{2K} - h_n^2 \mu_K^2$$

$$= \frac{1}{n} \sum_{j=1}^{n} x_j^2 - \bar{x}^2 + h_n^2 \sigma_K^2$$

Corollary 5.2.7:

If $X_1, X_2,\cdots, X_n$ is a sample of independently and identically distributed random variables

$$E\,\tilde{\sigma}^2(X_1,\cdots,X_n) = \frac{(n-1)}{n} \sigma^2 + \frac{1}{n} \sum_{j=1}^{n} h_j^2 \mu_{2K} - \frac{\mu_K^2}{n^2} \left(\sum_{j=1}^{n} h_j\right)^2$$

$$E\,\hat{\sigma}^2(X_1,\cdots,X_n) = \frac{(n-1)}{n} \sigma^2 + h_n^2 \sigma_K^2$$

and $\lim_{n\to\infty} E\,\tilde{\sigma}^2(X_1,\cdots,X_n) = \lim_{n\to\infty} E\hat{\sigma}^2(X_1,\cdots,X_n) = \sigma^2$

where $\sigma^2 = \text{Var } X_1 < \infty$.

5.3. Choice of ε:

The choice of ε is important in the use of our sequential procedure. We would like to choose ε in such a way as to link it to mean square error. An upper bound for ε is obtained through the following lemma.

Lemma 5.3.1

$$(5.3.1)\quad E|V_n(x)| \le \left(E|\hat{f}_{nM}(x) - f(x)|^2\right)^{1/2} + \left(E|\hat{f}_{(n-1)M}(x) - f(x)|^2\right)^{1/2}.$$

Proof:

From the definition of $V_n(x)$ we have

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

so that using the triangle inequality and taking expectations

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

The result follows by noting that for any distribution $EY^2 \ge [EY]^2$.

Now, $E|\hat{f}_n(x) - f(x)|^2 \to 0$ as $n\to\infty$. If we require that $E|\hat{f}_{nM}(x) - f(x)|^2 \le \gamma$ for some $\gamma > 0$ then it will certainly be sufficient to require that

$$\max\left(E|\hat{f}_{nM}(x) - f(x)|^2, E|\hat{f}_{(n-1)M}(x) - f(x)|^2\right) \le \gamma.$$

Then,

$$E|V_n(x)| \le 2\gamma^{1/2}.$$

Thus we should choose $\varepsilon \le 2\gamma^{1/2}$.

One question that arises is can we choose ε so that a specific value for mean square error is not exceeded. To do this we would have to find a function $g(\varepsilon)$ such that as $\varepsilon\to 0$, $g(\varepsilon) \to 0$ and $E|\hat{f}_n(x) - f(x)|^2 \le g(\varepsilon)$. Unfortuneately, I was unable to obtain this bound and this remains an open question.

5.4. Choice of M:

Our sequential procedure depends on the size, M, of the successive samples which we observe. We assume M is constant. Intuitively, if we choose M too large then the final total sample size attained will tend to be too large in the sense that many of the observations in the final sample may not be necessary. On the other hand, if M is too small the number of samples of size M required to stop will tend to be large. Clearly, in a practical situation the size of M will thus depend on the cost of individual observations and the extra cost involved (if any) in using small samples.

Formally, we stop the first time $|V_n(x)| < \varepsilon$. We actually stop when $|V_n(x)| = \varepsilon_1 \le \varepsilon$ and we would like to choose M so that ε_1 is approximately equal to ε. If ε_1 is very much different from ε it would indicate that either (a) M is too large

(b) $V_n(x)$ is very sensitive to extra observations.

From the nature of $V_n(x)$ we would conclude that except perhaps for small n, (a) is the case that applies. Thus one possible way of choosing M is to look at the effect of adding a sample of size M to the calculation of mean square error. Parzen (1962) show that for large n

$$E|\hat{f}_n(x) - f(x)|^2 = 0\left(\frac{1}{n^{4/5}}\right),$$

so that

$$E|\hat{f}_{nM}(x) - f(x)|^2 - E|\hat{f}_{(n-1)M}(x) - f(x)|^2 = 0\left(\frac{1}{n^{4/5}M^{4/5}}\left(1 - \frac{1}{(1-\frac{1}{n})^{4/5}}\right)\right)$$

(5.4.1) $$= 0\left(\frac{1}{n^{4/5}M^{4/5}}\left(1 - \left(1 + \frac{4}{5n} - \frac{18}{25n^2} + \cdots\right)\right)\right)$$

$$= \frac{4}{5}\cdot 0\left(\frac{1}{M^{4/5}n^{4/5}}\right).$$

In chapter 2 we assumed M was a constant independent of n. To be consistent with that assumption we assume that at any stage, (5.4.1) should be made less than $n^{-4/5}$ in which case $M^{4/5} \dot{\ge} \frac{4}{5} C$, where C is the constant of proportionality in (5.4.1) and C is given by.

$$C = 5\left(\frac{f(x)}{4} \cdot \int_{-\infty}^{\infty} K^2(y)dy\right)^{4/5} \cdot |k_2 f^{(2)}(x)|^{2/5}.$$

Comments:

1. We are "forcing" M to be a constant by our choice of rate of decrease for equation (5.4.1). Clearly, we can choose to make (5.4.1) decrease at a different rate than $n^{-9/5}$ in which case M will be a function of n. The choice of M as a function of n and its effect on the results in chapters two to four is a question for future study.

2. In section 5.3. we mentioned that it should be possible to find a $g(\varepsilon)$ such that $E|\hat{f}_{nM}(x) - f(x)|^2 \leq g(\varepsilon)$. Given $g(\varepsilon)$ it would probably be more desirable to use this inequality to choose M. For example, we might choose M so as to decrease the mean square error by a specified percentage at each stage of the procedure. This is also a topic for future research.

5.5. Global Stopping Rules:

This dissertation has been concerned only with pointwise stopping rules. It would be useful to also have stopping rules applicable to the global problem. Two possible quantities, upon which a global stopping rule could be based are

(i) $\sup_x |\hat{f}_{nM}(x) - \hat{f}_{(n-1)M}(x)| = \sup_x |V_n(x)|$

(ii) $\int_{-\infty}^{\infty} |\hat{f}_{nM}(x) - \hat{f}_{(n-1)M}(x)|^2 \, dx$

Since (i) is closely related to our $V_n(x)$ function of chapter 2 we will consider it. Lemma 2.2.1 (b) shows that $\sup_x |V_n(x)| \to 0$ in probability if f(x) is uniformly continuous. Lemma 5.5.1, which follows, gives an analogue of lemma 2.2.2.

Lemma 5.5.1

Let K(u) be such that (2.1.3) is satisfied and let k(u) be defined by (4.2.5). Assume $\int_{-\infty}^{\infty} |k(u)|du < \infty$ and that

$g(C) = \int |k(Ct) - k(t)|\, dt$ is locally Lipschitz of order 1 at C = 1.

Let $\{h_n\}$ be a sequence such that (2.1.4) and (2.1.5) are satisfied and also that

$$\sum_{n=1}^{\infty} \frac{1}{(nh_n)^2} < \infty$$

$$\sum_{n=1}^{\infty} \frac{1}{nh_n} \left| \frac{1}{h_{n+1}} - \frac{1}{h_n} \right| < \infty$$

$$\lim_{n\to\infty} nh_n^2 = \infty$$

then if f(x) is uniformly continuous on R, $\sup_x |\hat{f}_{nM}(x) - \hat{f}_{(n-1)M}(x)| \to 0$ a.s.

Proof:

The conditions stated are sufficient to ensure $\hat{f}_n(x)$ is a uniform Cauchy sequence a.s. (Van Ryzin, 1969) so that the result follows.

Corresponding to definition (2.1.2) we can define our stopping rule as given $\varepsilon > 0$,

(5.5.1)
$$N_S = \begin{cases} \text{First } n \text{ such that } \sup_x |V_n(x)| \le \varepsilon \\ \infty \text{ if no such } n \text{ exists.} \end{cases}$$

Lemma 5.5.2

If N is defined by 2.1.2 we have

$$EN_S \ge EN.$$

Proof:

From the definitions of N_S and N it is immediate that $N_S \geq N$ which implies $EN_S \geq EN$.

Remarks:

1. The result of this lemma is as expected. From the definition of N_S we intuitively expect to take more observations to satisfy the criterion. Also, from the nature of the problem we expect the result. That is, if we wish to satisfy some criterion for all x and not just a specific x we would expect to increase the sample size.

2. The results we would now like to prove are such things as $EN_S < \infty$. This should be possible. In fact we should be able to parallel the work in both chapters two and three for the global case.

5.6. Choice of h_n:

We saw in chapter 1, section 3, that one of the open problems in the use of kernel estimators, is the choice of h_n from the available observations. In this section we will suggest a method for choosing h_n. The idea has only intuition as a basis and will be a topic for future research.

Assume $X_1, X_2, \cdots, X_n$ is a sample of independently and identically distributed random variables with density function f(x). We will consider the estimate $\hat{f}_n(x)$, and since we assume $\int_{-\infty}^{\infty} K(u)du = 1$ we have that $\int_{-\infty}^{\infty} \hat{f}_n(x)dx = 1$.

Suppose now that we order the sample obtained to get the order statistics $-\infty < X_{(1)} < X_{(2)} < \cdots < X_{(n)} < \infty$. If x is the point at which we wish to estimate $\hat{f}_n(x)$ then for some r, $X_{(r)} \leq x \leq x_{(r+1)}$, ($X_{(r)}$ may be $-\infty$ and $X_{(r+1)}$ may be $+\infty$). Further, we can write

$$\int_{X_{(r)}}^{X_{(r+1)}} f(x)dx = F(x_{(r+1)}) - F(x_{(r)})$$

and

$$E \int_{X_{(r)}}^{X_{(r+1)}} f(x)dx = EF(x_{(r+1)}) - EF(x_{(r)}) = \frac{1}{n+1}.$$

Thus, since $\hat{f}_n(x)$ is an estimate of $f(x)$, let us estimate h_n by solving the equation

(5.6.1) $$\int_{X_{(r)}}^{X_{(r+1)}} \hat{f}_n(x)dx = \frac{1}{n+1}.$$

Since $K(u)$ is known and since we can write (5.6.1) as

$$\frac{1}{n} \sum_{j=1}^{n} \int_{X_{(r)}}^{X_{(r+1)}} \frac{1}{h_n} K\left(\frac{x-X_j}{h_n}\right) dx = \frac{1}{n+1}$$

we see that the only unknown is h_n for any given n. This corresponds to finding h_n for the pointwise problem.

For the global problem where we wish to find h_n so that we can use the same h_n for all points x we solve

$$\int_{X_{(1)}}^{X_{(n)}} \hat{f}_n(x)dx = \frac{n-1}{n+1} = 1 - \frac{2}{n+1}$$

where $\frac{n-1}{n+1}$ is chosen since

$$E \int_{X_{(1)}}^{X_{(n)}} f(x)dx = \frac{n-1}{n+1}.$$

Clearly there are many questions to be answered before we can claim this as a viable method for choosing a h_n sequence. At the very least we should prove (or disprove) that

(i) $\hat{h}_n \to 0$ ($^{\text{in prob}}_{\text{a.s.}}$?) as $n\to\infty$

(ii) $n\hat{h}_n \to \infty$ ($^{\text{in prob}}_{\text{a.s.}}$?) as $n\to\infty$.

In addition, since $\hat{h}_n$ is a random variable some attempt should be made to look at the distribution theory involved.

It may also be possible to make some judgement of "how good" a particular value of $\hat{h}_n$ is by use of a quantity we call the variation of $\hat{f}_n(x)$. This will be the subject of the next section.

5.7. Variation:

Define the variation of $\hat{f}_n(x)$ on a partition Y, $y_1 < y_2 < \cdots < y_n$, by

$$\zeta(h) = \sum_{j=1}^{m} |\hat{f}_n(y_j) - \hat{f}_n(y_{j-1})|.$$

As usual, total variation is defined as $\sup_Y \zeta(h)$ where $\sup_Y$ denotes taking the supremum over all possible partitions Y.

We believe that as h_n decreases, $\zeta(h_n)$ increases. If on the other hand h_n is too large the estimate may begin to look like the kernel function and the total variation would be approximately that of the kernel function. By studying the function $\zeta(h)$ we may be able to obtain some idea as to whether a given h is too large or too small.

Lemma 5.7.1

Given a sample $x_1, x_2, \cdots, x_n$,

$$h\,\zeta(h) \le \text{Total Variation } K(u)$$

Proof:

By definition, for a partition $y_1 < y_2 < \cdots < y_m$

$$\zeta(h) = \sum_{j=1}^{m} \frac{1}{nh}\left| \sum_{k=1}^{n}\left(K\left(\frac{y_j - X_k}{h}\right) - K\left(\frac{y_{j-1} - X_k}{h}\right) \right)\right|$$

$$\leq \sum_{j=1}^{m} \frac{1}{nh} \sum_{k=1}^{n}\left| K\left(\frac{y_j - X_k}{h}\right) - K\left(\frac{y_{j-1} - X_k}{h}\right)\right|$$

$$= \frac{1}{nh} \sum_{k=1}^{n} \sum_{j=1}^{m}\left| K\left(\frac{y_j - X_k}{h}\right) - K\left(\frac{y_{j-1} - X_k}{h}\right)\right|$$

$$\leq \frac{1}{nh} \sum_{k=1}^{n} \sup_{Y} \sum_{j=1}^{m}\left| K\left(\frac{y_j - X_k}{h}\right) - K\left(\frac{y_{j-1} - X_k}{h}\right)\right|$$

$$= \frac{1}{h} \text{ Total Var } K(u).$$

Observations:

1. As $h \to 0$ this bound becomes infinite. This agrees with our intuition that variation may be very large when h is small.

2. The upper bound decreases as h increases

3. Given $X_1, X_2, \cdots, X_n$ we can plot $\zeta(h)$. By looking at the value $\zeta(\hat{h}_n)$ and looking at its position on the curve we may be able to get some idea as "how good" $\hat{h}_n$ is as an estimate of h.

Theorem 5.7.2:

Total Variation $\hat{f}_n(x) \to \infty$ as $h \to 0$ for a fixed sample size.

Proof:

By definition

$$\zeta(h) = \sum_{j=1}^{m}\left| \frac{1}{nh} \sum_{k=1}^{n}\left(K\left(\frac{y_j - X_k}{h}\right) - K\left(\frac{y_{j-1} - X_k}{h}\right)\right)\right|$$

$$\geq \left| \frac{1}{nh} \sum_{k=1}^{n}\left(K\left(\frac{y_i - X_k}{h}\right) - K\left(\frac{y_{i-1} - X_k}{h}\right)\right)\right|$$

Choose a partition Y, such that for some $k = k_0$, $y_i = X_{k_0}$

$$\zeta(h) \geq \left| \frac{1}{nh} \sum_{k=1}^{n} \left(K\left(\frac{y_i - X_k}{h}\right) - K\left(\frac{y_{i-1} - X_k}{h}\right) \right) \right|$$

But, since $h \to 0$, with the partition chosen

$$\zeta(h) \to \infty.$$

But, $\sup_Y \zeta(h) \geq \zeta(h)$ so that

Total variation $\hat{f}_n(x) \to \infty$ as $h \to 0$ for a fixed sample size n.

Notes:

1. If sample size is allowed to vary and h_n is such that $nh_n \to \infty$ as $n \to \infty$ then the lower bound does not necessarily become infinite.

2. The condition $nh_n \to \infty$ as $n \to \infty$ is necessary for consistency so that the increase in variation as h_n becomes very small for fixed n is to be expected.

Lemma 5.7.3

$$\sup_Y \lim_{n\to\infty} \zeta(h) = \text{Total Var } f(x) \qquad \text{a.s.}$$

Proof:

$$\lim_{n\to\infty} \zeta(h) = \lim_{n\to\infty} \sum_{j=1}^{m} |\hat{f}_n(y_i) - \hat{f}_n(y_{i-1})|$$

$$= \sum_{j=1}^{m} \lim_{n\to\infty} |\hat{f}_n(y_i) - \hat{f}_n(y_{i-1})|$$

$$= \sum_{j=1}^{m} |f(y_i) - f(y_{i-1})| \qquad \text{a.s.}$$

Thus,

$$\sup_Y \lim_{n\to\infty} \zeta(h) = \sup_Y \sum_{j=1}^{m} |f(y_i) - f(y_{i-1})| = \text{Total Var. } f(x) \qquad \text{a.s.}$$

Lemma 5.7.4

$$\text{Total Var } f(x) \le \lim_{n\to\infty} \text{Total Var } \hat{f}_n(x)$$

where Total Var $\hat{f}_n(x) = \sup_{Y} \zeta(h)$.

Proof:

By definition

$$\zeta(h) = \sum_{j=1}^{m} |\hat{f}_n(y_j) - \hat{f}_n(y_{j-1})|$$

$$\le \sup_{Y} \sum_{j=1}^{m} |\hat{f}_n(y_j) - \hat{f}_n(y_{j-1})|$$

$$= \text{Total Var } \hat{f}_n(x)$$

Thus, $\lim_{n\to\infty} \zeta(h) \le \lim_{n\to\infty} \text{Total Var } \hat{f}_n(x)$.

But this inequality holds for all partitions Y so that

$$\sup_{Y} \lim_{n\to\infty} \zeta(h) \le \lim_{n\to\infty} \text{Total Var } \hat{f}_n(x)$$

and the left hand side is Total Var f(x) as required.

APPENDIX 1

TABLE 1 (Parzen)

$K(y)$	$k(u)$	$\int_{-\infty}^{\infty} K^2(y)dy$
$\frac{1}{2}, \ \lvert y\rvert \le 1$ $0, \ \lvert y\rvert \ge 1$	$\frac{\sin u}{u}$	$\frac{1}{2}$
$1 - \lvert y\rvert, \ \lvert y\rvert \le 1$ $0 \ , \ \lvert y\rvert \ge 1$	$\left(\frac{\sin(\frac{u}{2})}{\frac{u}{2}}\right)^2$	$\frac{2}{3}$
$(\frac{4}{3}) - 8y^2 + 8\lvert y\rvert^3; \ \lvert y\rvert < \frac{1}{2}$ $\frac{8}{3}(1 - \lvert y\rvert)^3, \ \frac{1}{2} \le y \le 1$ $0 \ , \ \lvert y\rvert > 1$	$\left(\frac{\sin(\frac{u}{4})}{\frac{u}{4}}\right)^4$	$0{\cdot}96$
$(2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}y^2}$	$e^{-\frac{1}{2}u^2}$	$\left(2(\pi^{\frac{1}{2}})\right)^{-1} = {\cdot}2821$
$\frac{1}{2} e^{-\lvert y\rvert}$	$(1 + u^2)^{-1}$	$\frac{1}{2}$
$\frac{1}{\pi}(1 + y^2)^{-1}$	$e^{-\lvert u\rvert}$	$\frac{1}{\pi} = {\cdot}3179$
$\frac{1}{2\pi}\left(\frac{\sin(\frac{y}{2})}{\frac{y}{2}}\right)^2$	$1 - \lvert u\rvert, \ \lvert u\rvert \le 1$ $0 \ , \ \lvert u\rvert \ge 1$	$\frac{1}{3\pi} = {\cdot}1061$

TABLE 2 (Epanechnikov)

$K(y)$	$L = \int_{-\infty}^{\infty} K^2(y)\,dy$	$r = L/\int_{-\infty}^{\infty} K_0^2(y)\,dy$
$K_0(y)$	$\frac{3}{5\sqrt{5}}$	1
$\frac{\sqrt{\pi^2-8}}{4} \cos \frac{\sqrt{\pi^2-8}}{2} y \; ; \; \lvert y\rvert \le \frac{\pi}{\sqrt{\pi^2-8}}$ $0 \; ; \; \lvert y\rvert \ge \frac{\pi}{\sqrt{\pi^2-8}}$	$\frac{\pi}{16}\sqrt{\pi^2-8}$	1·001
$\frac{1}{\sqrt{6}} - \frac{\lvert y\rvert}{6} \; ; \; \lvert y\rvert \le \sqrt{6}$ $0 \quad \lvert y\rvert > \sqrt{6}$	$\sqrt{6}/9$	1·015
$(2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}y^2}$	$\frac{1}{2\sqrt{5}}$	1·051
$\frac{1}{2\sqrt{3}} \; ; \; \lvert y\rvert \le \sqrt{3}$ $0 \; ; \; \lvert y\rvert > \sqrt{3}$	$\frac{1}{2\sqrt{3}}$	1·077
$\frac{1}{2} e^{-\sqrt{2}\lvert y\rvert}$	$\frac{1}{4\sqrt{2}}$	1·320

REFERENCES

[1] Anderson, G. D. (1969): "A comparison of methods for estimating a probability density function," Ph.D. Dissertation, University of Washington.

[2] Anscombe, F. J. (1952): "Large sample theory of sequential estimation," Proc. Camb. Philos. Soc. 48, 600 - 607.

[3] Davies, H. I. (1973): "Strong consistency of a sequential estimator of a probability density function," Bull. Math. Statist, 15, No. 3 ~ 4, 49 - 54.

[4] Doob, J. L. (1953): "Stochastic Processes," John Wiley, New York.

[5] Epanechnikov, V. A. (1969): "Non-parametric estimation of a multivariate probability density," Theor. Prob. Appl. 14, 153 - 158.

[6] Hill, D. W. (1973): "Estimation of probability functions using splines," Ph.D. Dissertation, University of New Mexico.

[7] Hoeffding, Wassily. (1963): "Probability inequalities for sums of bounded random variables," Journal of Amer. Stat. Assoc., 58, 13 - 20.

[8] Leadbetter, M. R. (1963): "On the nonparametric estimation of probability densities," Technical Report No. 11, Research Triangle Institute. (Doctoral dissertation at the University of North Carolina at Chapel Hill).

[9] Loève, M. (1960): "Probability Theory," Van Nostrand, Princeton, New Jersey.

[10] Moran, P. A. P. (1968): "Introduction to Probability Theory," Oxford University Press.

[11] Nadaraya, E. A. (1965): "On nonparametric estimates of density functions and regression surveys," Theor. Prob. Appl., 10. 186 - 190.

[12] Parzen, E. (1960): "Modern Probability Theory and its Applications," John Wiley, New York.

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

[14] Rosenblatt, M. (1956): "Remarks on some nonparametric estimates of a density function," Ann. Math. Statist., 42, 1815 - 1842.

[15] Srivastava, R. C. (1973): "Estimation of a probability function based on random number of observations with applications," Int. Stat. Rev., 41, Nol, 77 - 86.

[16] Van Ryzin, J. (1969): "On strong consistency of density estimates," Ann. Math. Statist., 40, 1765 - 1772.

[17] Wald, A. (1950): "Asymptotic minimax solutions of sequential point estimation problems," Proc. 2nd. Berkeley Symposium on Math. Stat and Probability, 1 - 12.

[18] Wahba, Grace (1971): "A polynomial algorithm for density estimation," Ann. Math. Statist, 42, 1870 - 1886.

[19] Watson, G. S. and Leadbetter, M. R. (1963): "On estimating a probability density, I," Ann. Math. Statist., 34, 480 - 491.

[20] Wegman, E. J. (1972): "Nonparametric probability density estimation: I. A summary of available methods," Technometrics, 14, 533 - 545.

[21] Wegman, E. J. (1972): "Nonparametric probability density estimation: II. A comparison of density estimation methods," J. Statist. Comput. Simul., 1, 225 - 245.

[22] Woodroofe, M. (1970): "On choosing a delta sequence," Ann. Math. Statist., 41, 1665 - 1671.

[23] Yamato, H. (1972): "Sequential estimation of a continuous probability density function and mode," Bull. Math. Statist., 14, No. 3 ~ 4, 1 - 12.