DB
530
1989 (H)

C:417.1

寄贈
磯貝英一氏
平成 年 月 日

NONPARAMETRIC PROBABILITY DENSITY ESTIMATION

USING RECURSIVE KERNEL ESTIMATORS

Eiichi Isogai

Department of Mathematics

Niigata University

THESIS

Submitted in partial fulfillment of the requirements for the degree of Doctor of Science in Doctoral Program in Mathematics in the University of Tsukuba

February 1989

92303088

ABSTRACT

Consider the problem of nonparametric probability density estimation. Recursive kernel estimators are proposed and their asymptotic properties are investigated. Also, a class of stopping rules based on the idea of fixed-width interval estimation is proposed. Finally, applying the idea of sequential density estimation, the problem of estimating a multiple regression function is considered.

CONTENTS

INTRODUCTION

Recently, nonparametric density estimation has been attracted by statisticians and scientists. In density estimation the useful statistical methods are parametric. The assumption in these methods is that the sample of observations comes from a population with a known parametric family. But when no assumptions about the functional form of the density function are made we should use another methods. These methods are nonparametric. In this thesis we consider the problem of nonparametric probability density estimation.

For the nonparametric estimation of density functions several methods have been developed. Furthermore, applications of nonparametric density estimators have been studied. Prakasa Rao [25] has made a survey of the methods developed in this area. Since Rosenblatt [27] has proposed the kernel method in 1956, Parzen [23] and many authors have investigated two types of the kernel estimators. One is nonrecursive and the other is recursive. The Parzen-Rosenblatt kernel estimator is nonrecursive. A disadvantage of this type is that the estimators are not easy to update when new data become available. In order to improve this disadvantage, Wolverton and Wagner [37], and Yamato [38] have developed the recursive kernel estimators. In this thesis we treat the recursive kernel estimators. Devroye and Györfi [12] published a book about nonparametric probability density estimation, in which they discussed various types of kernel estimators containing our recursive kernel estimators.

On the other hand, in estimating the probability density function there are many situations in practice where the number of observations on which the density estimators are to be calculated is not fixed but random. Carroll [6] has introduced two classes of integer-valued random variables (stopping rules) and developed sequential density estimators. By using these sequential density estimators he has given confidence intervals for the density function of fixed-width and prescribed coverage probability.

In Chapter 1 we propose recursive kernel estimators $f_n(x)$ of the nonparametric probability density function $f(x)$ in the multivariate case. Also, we study their asymptotic properties including consistency and asymptotic normality.

Chapter 2 deals with sequential density estimators in the univariate case. We define a certain class of stopping rules $N(d)$ and take $I_{N(d),d}(x) = [f_{N(d)}(x) - d\,,\, f_{N(d)}(x) + d]$ as a 2d-width confidence interval for $f(x)$ with prescribed coverage probability α. The rate of convergence of $P\{f(x) \varepsilon I_{N(d),d}(x)\}$ to α as $d \to 0$ is investigated.

In Chapter 3 we consider a modified class of the stopping rules in Chapter 2 and study the limiting behavior of the moments of the stopping rules.

In the last chapter we apply the idea of sequential density estimation to the problem of estimating a multiple regression function $m(x)$. We define a class of stopping rules $N(d)$, and by using the recursive kernel estimators $m_n(x)$ of $m(x)$ we give $I_{N(d),d}(x) = [m_{N(d)}(x) - d\,,\, m_{N(d)}(x) + d]$ as a 2d-width confidence interval for $m(x)$ with prescribed coverage probability α. We also show the convergence of $P\{m(x) \varepsilon I_{N(d),d}(x)\}$ to α as $d \to 0$.

1. RECURSIVE DENSITY ESTIMATION

1.1. FORMULATION OF THE PROBLEM

Let f(x) be a (unknown) probability density function (p.d.f.) on R^p with respect to the Lebesgue measure where R^p denotes the p-dimensional Euclidean space. There is a vast literature on the problem of recursively estimating the p.d.f. f(x). Yamato [38] introduced the recursive estimator defined by

$$\hat{f}_n(x) = (1-n^{-1})\hat{f}_{n-1}(x) + (nh_n^p)^{-1}K((x-x_n)/h_n). \qquad (\hat{F})$$

Carroll [6] and Davies [9] considered this estimator $\hat{f}_n$. The following recursive kernel estimator f_n^* was introduced by Wegman and Davies [36]:

$$f_n^*(x) = (1-n^{-1})(h_{n-1}/h_n)^{1/2}f_{n-1}^*(x) + (nh_n)^{-1}K((x-X_n)/h_n)$$

for $x \varepsilon R^1$.

In this section we shall propose a recursive kernel estimator $f_n(x)$ in a modified form of $\hat{f}_n(x)$, that is, n^{-1} is replaced by an^{-1} with $1/2 < a \leqq 1$. The estimator $f_n(x)$ defined in Section 1.2 and the estimator $\hat{f}_n(x)$ can be rewritten, respectively, as

$$f_n(x) = \sum_{m=1}^{n} a_m \beta_{mn} h_m^{-p} K((x-x_m)/h_m) + \beta_{0n} f_0(x)$$

and

$$\hat{f}_n(x) = \sum_{m=1}^{n} (nh_m^p)^{-1} K((x-x_m)/h_m).$$

If we put a=1 then it turns out that $a_m\beta_{mn} = n^{-1}$ for all $1 \leqq m \leqq n$ and $\beta_{0n}=0$. Hence $\hat{f}_n(x)$ can be obtained by substituting a=1 into $f_n(x)$.

In this chapter we shall consider the problem of estimating f(x) at a given point x. Yamato [38] showed the weak consistency of $\hat{f}_n(x)$ and its asymptotic normality. Davies [9] showed the strong consistency of $\hat{f}_n(x)$. We shall show the strong consistency of $f_n(x)$ and its asymptotic normality. The estimator $f_n(x)$ will also be shown to be better than $\hat{f}_n(x)$, from the viewpoint of the criterion of the asymptotic rate of variances, for some choice of the smoothing parameter h_n.

In Section 1.2 we shall define the recursive kernel estimator $f_n(x)$ in a certain form. Also, auxiliary results will be given. In Section 1.3 the estimator $f_n(x)$ will be shown to be strongly consistent and its asymptotic normality will also be shown. Further, the rate of convergence of the mean square error will be given. Section 1.4 is devoted to comparison between $f_n(x)$ and $\hat{f}_n(x)$ under the criterion of the limit of $Var(f_n(x))/Var(\hat{f}_n(x))$. We shall give a special type of the sequence $\{h_n\}$, by which it will be demonstrated that the above limit is strictly less than 1 for certain a depending on this sequence $\{h_n\}$. Thus, in this case the estimator $f_n(x)$ is better than the estimator $\hat{f}_n(x)$ under the above criterion. Further, an optimal choice of the coefficient a in the estimator $f_n(x)$ will be given for the special type of $\{h_n\}$.

1.2. ASSUMPTIONS AND AUXILIARY RESULTS

In this section we shall give a recursive kernel estimator of the p.d.f. f(x) and give several results which are needed for the sections that follow.

Let K be a real-valued Borel measurable function on R^p satisfying

$$\int |K(y)|dy < \infty \quad \text{and} \quad \int K(y)dy = 1, \tag{K1}$$

$$||K||_\infty = \sup_{y \in R^p} |K(y)| < \infty \tag{K2}$$

and

$$\lim_{||y|| \to \infty} ||y||^p |K(y)| = 0, \tag{K3}$$

where $||\cdot||$ denotes the Euclidean norm on R^p and the domain of integral is R^p unless otherwise specified. Let $\{a_n\}$ be a sequence of positive numbers defined by

$$a_n = a/n \text{ with } 1/2 < a \leqq 1 \quad \text{for all } n \geqq 1. \tag{A}$$

Let $\{h_n\}$ be a sequence of positive numbers satisfying the following conditions:

$$\lim_{n \to \infty} h_n = 0, \tag{H1}$$

$$h_{n_0} \geqq h_{n_0+1} \geqq \cdots \quad \text{for some } n_0 \geqq 1, \tag{H2}$$

$$\lim_{n \to \infty} nh_n^p = \infty, \tag{H3}$$

$$\lim_{n \to \infty} (nh_n^p)^{-1/2} \log n = 0, \tag{H4}$$

$$\sum_{n=1}^{\infty} (n^2 h_n^p)^{-1} < \infty. \tag{H5}$$

Define K_n by

$$K_n(x,y) = h_n^{-p} K((x-y)/h_n) \quad \text{for all } x,y \in R^p,\ n=1,2,\cdots.$$

Consider the modified recursive kernel estimator $f_n(x)$ for $f(x)$ given by

$$f_0(x) = K(x) \quad \text{for all } x \in R^p \tag{F}$$

$$f_n(x) = (1-a_n) f_{n-1}(x) + a_n K_n(x, X_n) \quad \text{for all } x \in R^p,\ n=1,2,\cdots,$$

where $X_1, X_2, X_3, \cdots$ is a sequence of independent and identically distributed p-dimensional random vectors with the common p.d.f. f, and the conditions (K1)$\sim$(K3), (A) and (H1) are assumed to be satisfied. In what follows, for the estimator f_n we shall assume the conditions (K1)$\sim$(K3), (A) and (H1) without restating them repeatedly. Throughout this chapter $C_1, C_2, \cdots$ denote appropriate positive constants, and for any function g on R^p $C(g)$ stands for the set of all points of continuity of g.

REMARK. If $K(x) \geqq 0$ for all $x \in R^p$, then it is easy to see that $f_n(x)$ $(n=1,2,\cdots)$ are actually probability density functions. If we put a=1 in (A) then the estimator $f_n(x)$ coincides with the estimator $\hat{f}_n(x)$.

Now, we introduce notation. Let

$$\beta_{mn} = \begin{cases} \prod_{k=m+1}^{n} (1-a_k) & \text{for } 0 \leqq m < n \\ 1 \quad \text{for } m=n \geqq 0 \end{cases}$$

$$\gamma_0 = \gamma_1 = 1$$

and

$$\gamma_n = \prod_{j=2}^{n} (1-a_j) \quad \text{for all } n \geqq 2.$$

It is clear that $\gamma_n > 0$ for all $n \geqq 1$, $\beta_{mn} \leqq 1$ for all $n \geqq m \geqq 0$ and

$\beta_{mn} = \gamma_n \gamma_m^{-1}$ for all $n \geqq m \geqq 1$.

Sacks [29] showed that

$$(1-\varepsilon_m')m^a n^{-a} \leqq \beta_{mn} \leqq (1+\varepsilon_m')m^a n^{-a} \quad \text{for all } n \geqq m \geqq 1, \tag{1.2.1}$$

where $\varepsilon_m' \to 0$ as $m \to \infty$. By (1.2.1) and the fact that $1-a_j > 0$ for all $j \geqq 2$, there exist two positive constants C_1 and C_2 such that

$$C_1 m^a n^{-a} \leqq \beta_{mn} \leqq C_2 m^a n^{-a} \quad \text{for all } n \geqq m \geqq 1. \tag{1.2.2}$$

In particular, it holds that

$$C_1 n^{-a} \leqq \gamma_n \leqq C_2 n^{-a} \quad \text{for all } n \geqq 1. \tag{1.2.3}$$

LEMMA 1.2.1. Let $\{h_n\}$ satisfy (H1) and (H2). If for some $a > 1/2$ there exists a positive constant β such that

$$n^{1-2a} h_n^p \sum_{m=1}^{n} m^{2(a-1)} h_m^{-p} \to \beta \quad \text{as } n \to \infty, \tag{1.2.4}$$

then for any positive integer m_0

$$\sum_{m=m_0}^{n} a^2 m^{-2} \gamma_m^{-2} h_m^{-p} \sim a^2 \beta (n h_n^p \gamma_n^2)^{-1} \quad \text{as } n \to \infty,$$

where "$\phi_n \sim \psi_n$ as $n\to\infty$" means that $\phi_n/\psi_n \to 1$ as $n\to\infty$.

PROOF. It suffices to show that

$$\beta^{-1} n h_n^p \sum_{m=m_0}^{n} m^{-2} \beta_{mn}^2 h_m^{-p} \to 1 \quad \text{as } n \to \infty. \tag{1.2.5}$$

Let any ε with $0 < \varepsilon < 1$ be fixed. Choose ξ with $0 < \xi < 1$ such that

$$(1+2\varepsilon/3)(1+\xi) < 1+\varepsilon \quad \text{and} \quad (1-2\varepsilon/3)(1-\xi) > 1-\varepsilon. \tag{1.2.6}$$

By (1.2.1) there exists a positive integer m_1 greater than both of m_0 and n_0, where n_0 is given in (H2), such that

$$(1-\varepsilon/3)m^{2a} n^{-2a} \leqq \beta_{mn}^2 \leqq (1+\varepsilon/3)m^{2a} n^{-2a} \tag{1.2.7}$$

for all $n \geqq m \geqq m_1$. By the monotonicity of h_n and $a > 1/2$ we get

$$\sum_{m=1}^{n} m^{2(a-1)} h_m^{-p} \to \infty \quad \text{as } n \to \infty. \tag{1.2.8}$$

From (H1), $a > 1/2$, (1.2.4) and (1.2.8) there exists a positive integer $m_2 \geqq m_1$ such that

$$1-\xi < \beta^{-1} n^{1-2a} h_n^p \sum_{m=m_1}^{n} m^{2(a-1)} h_m^{-p} < 1+\xi \tag{1.2.9}$$

and

$$-\varepsilon/3 < C_2^2 \sum_{m=m_0}^{m_1-1} m^{-2} \gamma_m^{-2} h_m^{-p} \Big/ \sum_{m=m_1}^{n} m^{2(a-1)} h_m^{-p} < \varepsilon/3 \tag{1.2.10}$$

for all $n \geqq m_2$. Combining (1.2.3),(1.2.6),(1.2.7),(1.2.9) and (1.2.10) we obtain

$$\beta^{-1} n h_n^p \sum_{m=m_0}^{n} m^{-2} \beta_{mn}^2 h_m^{-p} \tag{1.2.11}$$

$$< \beta^{-1} n^{1-2a} h_n^p \sum_{m=m_1}^{n} m^{2(a-1)} h_m^{-p} \{ (C_2^2 \sum_{m=m_0}^{m_1-1} m^{-2} \gamma_m^{-2} h_m^{-p} \Big/ \sum_{m=m_1}^{n} m^{2(a-1)} h_m^{-p}) + 1 + \varepsilon/3 \}$$

$$< (1+\xi)(1+2\varepsilon/3) < 1+\varepsilon \quad \text{for all } n \geqq m_2.$$

In the same manner as above we have

$$\beta^{-1} n h_n^p \sum_{m=m_0}^{n} m^{-2} \beta_{mn}^2 h_m^{-2} > 1-\varepsilon \quad \text{for all } n \geqq m_2. \tag{1.2.12}$$

Thus the combined use of (1.2.11) and (1.2.12) yields (1.2.5), which completes the proof.

The next lemma is one of the results given by Sacks [29].

LEMMA 1.2.2. Let $q > -1$. Then for any positive integer m_0

$$\sum_{m=m_0}^{n} m^q \sim (1+q)^{-1} n^{q+1} \quad \text{as } n \to \infty.$$

The following two propositions can be found in Watanabe [34].

PROPOSITION 1.2.3. Let $\{A_n\}$ be a sequence of nonnegative numbers. Suppose that there exist two sequences of nonnegative numbers $\{b_n\}$ and $\{d_n\}$ such that

$$A_{n+1} \leqq (1-b_{n+1})A_n + b_{n+1}d_{n+1} \quad \text{for all } n \geqq 1, \tag{1.2.13}$$

$$\sum_{n=1}^{\infty} b_n = \infty \quad \text{and} \quad \lim_{n\to\infty} b_n = 0, \tag{1.2.14}$$

and

$$\lim_{n\to\infty} d_n = 0. \tag{1.2.15}$$

Then we have

$$\lim_{n\to\infty} A_n = 0.$$

PROPOSITION 1.2.4. Let $\{U_n\}$ and $\{V_n\}$ be two sequences of random variables on some probability space $(\Omega, \mathcal{F}, P)$. Let $\{\mathcal{F}_n\}$ be a sequence of σ-fields, $\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{F}$ for all $n \geqq 1$, where U_n and V_n are measurable with respect to $\mathcal{F}_n$ for each $n \geqq 1$. And let $\{b_n\}$ be a sequence of real numbers. Suppose that the following conditions are satisfied:

$$0 \leqq U_n \quad \text{a.s.} \quad \text{for all } n \geqq 1, \tag{1.2.16}$$

$$E[U_1] < \infty, \tag{1.2.17}$$

$$E[U_{n+1}|\mathcal{F}_n] \leqq (1-b_{n+1})U_n + V_n \quad \text{a.s.} \quad \text{for all } n \geqq 1, \tag{1.2.18}$$

$$\sum_{n=1}^{\infty} E|V_n| < \infty \tag{1.2.19}$$

and

$$0 \leqq b_n \leqq 1 \ (n=1,2,\cdots), \ \lim_{n\to\infty} b_n = 0 \text{ and } \sum_{n=1}^{\infty} b_n = \infty, \tag{1.2.20}$$

where $E[\cdot]$ and $E[\cdot|\cdot]$ denote the expectation and the conditional expectation operators, respectively. Then we have

$$\lim_{n\to\infty} U_n = 0 \quad \text{a.s.} \quad \text{and} \quad \lim_{n\to\infty} E[U_n] = 0.$$

Cacoullos [5] gave the following proposition.

PROPOSITION 1.2.5. Let $K(y)$ be a real-valued Borel measurable function on R^p satisfying (K1)$\sim$(K3) without $\int K(y)dy = 1$. Let $g(y)$ be a real-valued Borel measurable function on R^p such that $\int|g(y)|dy < \infty$, and let

$$g_n(x) = h_n^{-p} \int K(y/h_n) g(x-y) dy,$$

where $\{h_n\}$ is a sequence of positive constants satisfying (H1). Then at each point $x \varepsilon C(g)$

$$\lim_{n\to\infty} g_n(x) = g(x) \int K(y) dy.$$

DEFINITION 1.2.6. A bounded real-valued function g defined on R^p is said to be locally Lipschitz of order λ, $1/2 < \lambda \leqq 1$, at x_0 (abbreviated as loc. Lip. λ at x_0) if there exist two positive constants L and η, which may depend on x_0, such that $||y|| < \eta$ implies $|g(x_0+y) - g(x_0)| \leqq L\,||y||^{\lambda}$.

1.3. STRONG CONSISTENCY AND ASYMPTOTIC NORMALITY

In this section we shall show the strong consistency of the estimator f_n defined by (F) of Section 1.2 and we shall discuss the asymptotic normality of the estimator.

The following theorem shows the strong consistency of the estimator f_n.

THEOREM 1.3.1. Let $\{h_n\}$ satisfy (H1) and (H5). Then for each $x \in C(f)$

$$\lim_{n\to\infty} f_n(x) = f(x) \quad \text{a.s.} \tag{1.3.1}$$

and

$$\lim_{n\to\infty} E(f_n(x) - f(x))^2 = 0. \tag{1.3.2}$$

PROOF. From the algorithm (F) it follows that

$$|Ef_n(x) - f(x)| \tag{1.3.3}$$

$$\leqq (1-a_n)|Ef_{n-1}(x) - f(x)| + a_n|EK_n(x,X_n) - f(x)|$$

for all $n \geqq 1$. By Proposition 1.2.5 we get

$$\lim_{n\to\infty} |EK_n(x,X_n) - f(x)| = 0. \tag{1.3.4}$$

Thus, making use of (1.3.3),(1.3.4) and Proposition 1.2.3 we have

$$\lim_{n\to\infty} |Ef_n(x) - f(x)| = 0. \tag{1.3.5}$$

Now, by the algorithm (F) and the independence of X_n we obtain

$$E[(f_n(x) - Ef_n(x))^2 | X_1, \cdots, X_{n-1}] \tag{1.3.6}$$

$$\leqq (1-a_n)(f_{n-1}(x) - Ef_{n-1}(x))^2 + a_n^2 \mathrm{Var}[K_n(x,X_n)]$$

$$\leq (1-a_n)(f_{n-1}(x) - Ef_{n-1}(x))^2 + a_n^2 E[K_n^2(x,X_n)] \quad \text{a.s.}$$

for all $n \geq 2$, where Var(X) denotes the variance of X.

By Proposition 1.2.5 we have

$$\lim_{n\to\infty} h_n^p E[K_n^2(x,X_n)] = f(x)\int K^2(y)dy,$$

which yields by using of (K1) and (K2) that

$$h_n^p E[K_n^2(x,X_n)] \leq C_3 \quad \text{for all } n \geq 1. \tag{1.3.7}$$

Hence, combining (1.3.6) and (1.3.7) we obtain

$$E[(f_n(x) - Ef_n(x))^2 | X_1, \cdots, X_{n-1}] \tag{1.3.8}$$

$$\leq (1-a_n)(f_{n-1}(x) - Ef_{n-1}(x))^2 + C_3 a_n^2 h_n^{-p} \quad \text{a.s.} \quad \text{for all } n \geq 2.$$

From (H5) it follows that

$$\sum_{n=1}^{\infty} a_n^2 h_n^{-p} < \infty. \tag{1.3.9}$$

By (K2) we have

$$E(f_1(x) - Ef_1(x))^2 < \infty. \tag{1.3.10}$$

Thus, from (A),(1.3.8),(1.3.9),(1.3.10) and Proposition 1.2.4 we get

$$\lim_{n\to\infty} (f_n(x) - Ef_n(x))^2 = 0 \quad \text{a.s.} \tag{1.3.11}$$

and

$$\lim_{n\to\infty} E(f_n(x) - Ef_n(x))^2 = 0. \tag{1.3.12}$$

On the other hand

$$|f_n(x) - f(x)| \leq |f_n(x) - Ef_n(x)| + |Ef_n(x) - f(x)| \tag{1.3.13}$$

and

$$E(f_n(x) - f(x))^2 = E(f_n(x) - Ef_n(x))^2 + (Ef_n(x) - f(x))^2 \tag{1.3.14}$$

for $n \geq 1$. Hence we get (1.3.1) from (1.3.5),(1.3.11) and (1.3.13), and get (1.3.2) from (1.3.5),(1.3.12) and (1.3.14), which complete the proof.

We shall give the following theorem concerning the order of convergence of $Var(f_n(x))$.

THEOREM 1.3.2. Let $\{h_n\}$ satisfy (H1) and (H2). Assume the following condition:

For some a in (A) with $1/2 < a \leq 1$ there exists a positive constant β such that (1.3.15)

$$n^{1-2a} h_n^p \sum_{m=1}^{n} m^{2(a-1)} h_m^{-p} \to \beta \quad \text{as } n \to \infty.$$

Then for each $x \in C(f)$

$$\lim_{n\to\infty} n h_n^p Var(f_n(x)) = Bf(x), \tag{1.3.16}$$

where

$$B = a^2 \beta \int K^2(y)\,dy > 0.$$

PROOF. Let

$$Z_m = K_m(x,X_m) - EK_m(x,X_m) \quad \text{for all } m \geq 1. \tag{1.3.17}$$

From the algorithm (F) we get

$$f_n(x) - Ef_n(x) = \sum_{m=1}^{n} a_m \beta_{mn} Z_m \quad \text{for all } n \geq 1. \tag{1.3.18}$$

It follows from (H1) and Proposition 1.2.5 that

$$\lim_{n\to\infty} h_n^p E Z_n^2 = f(x) \int K^2(y)\,dy. \tag{1.3.19}$$

Let

$$b_n = \sum_{m=1}^{n} a_m^2 \gamma_m^{-2} h_m^{-p} \quad \text{for all } n \geq 1.$$

By Lemma 1.2.1 we obtain

$$b_n n h_n^p \gamma_n^2 \sim a^2 \beta \ (>0) \quad \text{as } n \to \infty. \tag{1.3.20}$$

On the other hand, by (1.2.3) we have

$$0 < n h_n^p \gamma_n^2 \leq C_4 n^{1-2a} h_n^p ,$$

which yields, together with $a > 1/2$ and (H1), that

$$\lim_{n\to\infty} n h_n^p \gamma_n^2 = 0. \tag{1.3.21}$$

Thus it follows from (1.3.20) and (1.3.21) that

$$b_n \uparrow \infty \quad \text{as } n \to \infty. \tag{1.3.22}$$

By (1.3.19),(1.3.22) and the Toeplitz lemma (see Loève [20]) we have

$$b_n^{-1} \sum_{m=1}^{n} a_m^2 \gamma_m^{-2} E Z_m^2 \to f(x) \int K^2(y)\,dy \quad \text{as } n \to \infty. \tag{1.3.23}$$

From (1.3.18) we get

$$\begin{aligned} \mathrm{Var}(f_n(x)) &= \sum_{m=1}^{n} a_m^2 \beta_{mn}^2 E Z_m^2 \\ &= (n h_n^p)^{-1} (b_n n h_n^p \gamma_n^2) b_n^{-1} \sum_{m=1}^{n} a_m^2 \gamma_m^{-2} E Z_m^2 , \end{aligned}$$

which yields

$$n h_n^p \mathrm{Var}(f_n(x)) = (b_n n h_n^p \gamma_n^2) b_n^{-1} \sum_{m=1}^{n} a_m^2 \gamma_m^{-2} E Z_m^2 . \tag{1.3.24}$$

Combining (1.3.20),(1.3.23) and (1.3.24) we obtain

$$\lim_{n\to\infty} n h_n^p \mathrm{Var}(f_n(x)) = B f(x) ,$$

which concludes the theorem.

LEMMA 1.3.3. Suppose that in addition to (K1) $\sim$ (K3), $K(y)$ satisfies the condition

$$\int \|y\| \, |K(y)| dy < \infty . \tag{K4}$$

Then for each x at which f is loc. Lip. λ, there exists a positive constant C, which may depend on x, such that

$$|EK_n(x, X_n) - f(x)| \leqq C h_n^{\lambda} \quad \text{for all } n \geqq 1. \tag{1.3.25}$$

PROOF. Let

$$\delta_n = EK_n(x, X_n) - f(x) \quad \text{for all } n \geqq 1 \tag{1.3.26}$$

and

$$\|f\|_{\infty} = \sup_{y \in R^p} f(y). \tag{1.3.27}$$

The boundedness of the p.d.f. f implies $\|f\|_{\infty} < \infty$. By (K1) and Definition 1.2.6 we have

$$|\delta_n| \tag{1.3.28}$$

$$\leqq \int |K(y)| \, |f(x - h_n y) - f(x)| dy$$

$$\leqq C_3 \int_{\{h_n \|y\| < \eta\}} \|y\|^{\lambda} |K(y)| dy \, h_n^{\lambda} + 2\|f\|_{\infty} \int_{\{h_n \|y\| \geqq \eta\}} |K(y)| dy$$

$$\leqq C_3 \int \|y\|^{\lambda} |K(y)| dy \, h_n^{\lambda} + 2 \, \|f\|_{\infty} \eta^{-\lambda} h_n^{\lambda} \int \|y\|^{\lambda} |K(y)| dy$$

$$= (C_3 + 2 \, \|f\|_{\infty} \eta^{-\lambda}) \int \|y\|^{\lambda} |K(y)| dy \, h_n^{\lambda}$$

for all $n \geqq 1$. If $\lambda = 1$, then it follows from (K4) that

$$\int \|y\|^{\lambda} |K(y)| dy < \infty$$

If $1/2 < \lambda < 1$, then it follows from (K1),(K4) and the Hölder inequality that

$$\int \|y\|^{\lambda} |K(y)| dy$$

$$\leq [\int \|y\| |K(y)| dy]^{\lambda} [\int |K(y)| dy]^{1-\lambda} < \infty.$$

Hence, putting $C = (C_3 + 2 \|f\|_{\infty} \eta^{-\lambda}) \int \|y\|^{\lambda} |K(y)| dy$ we get

$$C < \infty. \tag{1.3.29}$$

Finally we get (1.3.25) from (1.3.28) and (1.3.29). The proof is complete.

The following theorem presents the rate of convergence of mean square error.

THEOREM 1.3.4. Let x be a point such that $f(x) > 0$ and f is loc. Lip. λ at x. Let $K(y)$ be given in Lemma 1.3.3. Suppose that $\{h_n\}$ satisfies (H1),(H2),(H3) and (H5). In addition to (1.3.15), assume the following condition:

$$\sum_{m=1}^{n} m^{a-1} h_m^{\lambda} = O(n^a h_n^{\lambda}) \quad \text{as } n \to \infty, \tag{1.3.30}$$

where a is given in (1.3.15).
Then there exists a positive constant C, which may depend on x, such that

$$E(f_n(x) - f(x))^2 \leq C b_n \quad \text{for all } n \geq 1, \tag{1.3.31}$$

where $b_n = \max\{(nh_n^p)^{-1}, h_n^{2\lambda}, n^{-2a}\}$.

PROOF. By Theorem 1.3.2 there exists a positive constant

C_3 such that

$$\mathrm{Var}(f_n(x)) \leqq C_3(nh_n^p)^{-1} \quad \text{for all } n \geqq 1. \qquad (1.3.32)$$

Let δ_m be given in (1.3.26). From (1.2.2),(1.3.30) and Lemma 1.3.3 we get

$$\left|\sum_{m=1}^{n} a_m \beta_{mn} \delta_m\right| \leqq C_4 n^{-a} \sum_{m=1}^{n} m^{a-1} h_m^{\lambda} \leqq C_5 h_n^{\lambda} \quad \text{for all } n \geqq 1.$$

Hence we have

$$\left(\sum_{m=1}^{n} a_m \beta_{mn} \delta_m\right)^2 \leqq C_6 h_n^{2\lambda} \quad \text{for all } n \geqq 1. \qquad (1.3.33)$$

Since $|f_0(x) - f(x)| \leqq \|K\|_\infty + \|f\|_\infty < \infty$, where $\|f\|_\infty$ is defined by (1.3.27), it follows from (1.2.3) that

$$\beta_{0n}^2 (f_0(x) - f(x))^2 \leqq C_7 n^{-2a} \quad \text{for all } n \geqq 1. \qquad (1.3.34)$$

Since $Ef_n(x) - f(x) = \beta_{0n}(f_0(x) - f(x)) + \sum_{m=1}^{n} a_m \beta_{mn} \delta_m$, by (1.3.33) and (1.3.34) we have

$$(Ef_n(x) - f(x))^2 \leqq C_8 b_n \quad \text{for all } n \geqq 1. \qquad (1.3.35)$$

Let $C = C_3 + C_8 < \infty$. Hence by (1.3.14),(1.3.32) and (1.3.35) we obtain (1.3.31). This completes the proof.

We shall now show the asymptotic normality of the estimator f_n. In the remainder of this section $K(y)$ is assumed to satisfy (K1) ∿ (K4).

LEMMA 1.3.5. Let Z_m be given in (1.3.17). Suppose that $\{h_n\}$ satisfies (H1),(H2) and (H4). Assume the following conditions:

For some a in (A) with $2/3 \leq a \leq 1$ there exists (1.3.36)
a positive constant β such that

$$n^{1-2a} h_n^p \sum_{m=1}^{n} m^{2(a-1)} h_m^{-p} \to \beta \quad \text{as } n \to \infty$$

and

$$\|f\|_\infty = \sup_{y \in R^p} f(y) < \infty. \tag{1.3.37}$$

Then for each $x \in C(f)$ it holds that

$$(nh_n^p)^{1/2} \sum_{m=1}^{n} a_m \beta_{mn} Z_m \xrightarrow[L]{} N(0, Bf(x)) \quad \text{as } n \to \infty, \tag{1.3.38}$$

where B is given in Theorem 1.3.2, $N(0,\sigma^2)$ stands for the normal random variable with mean 0 and variance σ^2 and " $\xrightarrow[L]{}$ " means convergence in law.

PROOF. Let $U_n = a_n \gamma_n^{-1} Z_n$, $S_n = \sum_{m=1}^{n} U_m$ and $s_n^2 = \mathrm{Var}(S_n)$ for all $n \geq 1$. Then it holds that $s_n^2 = \sum_{m=1}^{n} a_m^2 \gamma_m^{-2} EZ_m^2$. Let $b_n = \sum_{m=1}^{n} a_m^2 \gamma_m^{-2} h_m^{-p}$ for all $n \geq 1$. First we consider the case when $f(x) = 0$. Since

$$E[\{(nh_n^p)^{1/2} \sum_{m=1}^{n} a_m \beta_{mn} Z_m\}^2] = (b_n nh_n^p \gamma_n^2)(b_n^{-1} s_n^2),$$

it follows from (1.3.20) and (1.3.23) that

$$\lim_{n\to\infty} E[\{(nh_n^p)^{1/2} \sum_{m=1}^{n} a_m \beta_{mn} Z_m\}^2] = 0,$$

which implies that

$$(nh_n^p)^{1/2} \sum_{m=1}^{n} a_m \beta_{mn} Z_m \to 0 \quad \text{in probability as } n \to \infty.$$

Thus (1.3.38) holds.

In what follows we consider the case when $f(x) > 0$. It follows from (1.3.20) and (1.3.23) that

$$s_n^2 \sim Bf(x)(nh_n^p \gamma_n^2)^{-1} \quad \text{as } n \to \infty. \tag{1.3.39}$$

If the Lyapounov condition

$$s_n^{-3} \sum_{m=1}^{n} E|U_m|^3 \to 0 \quad \text{as } n \to \infty \tag{1.3.40}$$

holds, then it follows that

$$s_n^{-1} S_n \xrightarrow[L]{} N(0,1) \quad \text{as } n \to \infty. \tag{1.3.41}$$

Using the inequality $|a+b|^3 \leqq 4(|a|^3 + |b|^3)$ and the Hölder inequality we get

$$E|U_m|^3 = a_m^3 \gamma_m^{-3} E|Z_m|^3 \leqq 8 a_m^3 \gamma_m^{-3} E[|K_m(x, X_m)|^3]. \tag{1.3.42}$$

Since

$$E[|K_m(x,X_m)|^3] = h_m^{-2p} \int |K(y)|^3 f(x - h_m y)\, dy \leqq ||K||_\infty^2 ||f||_\infty \int |K(y)|\, dy\, h_m^{-2p}$$

for all $m \geqq 1$, it follows from (K1),(K2) and (1.3.37) that

$$E[|K_m(x,X_m)|^3] \leqq C_3 h_m^{-2p} \quad \text{for all } m \geqq 1, \tag{1.3.43}$$

where $C_3 = ||K||_\infty^2 ||f||_\infty \int |K(y)|\, dy < \infty$. From (1.3.42) and (1.3.43) we obtain

$$E|U_m|^3 \leqq C_4 m^{-3} \gamma_m^{-3} h_m^{-2p} \quad \text{for all } m \geqq 1. \tag{1.3.44}$$

Let $n_1 = n_0 + 1$ with n_0 being given in (H2). Thus, by using (1.2.3), (1.3.39),(1.3.44) and (H2) we have

$$s_n^{-3} \sum_{m=n_1}^{n} E|U_m|^3 \leqq C_5 \gamma_n^3 (nh_n^p)^{3/2} \sum_{m=n_1}^{n} m^{-3} \gamma_m^{-3} h_m^{-2p} \tag{1.3.45}$$

$$\leq C_6 n^{2-3a}(nh_n^p)^{-1/2} \sum_{m=n_1}^{n} m^{3(a-1)}(h_n/h_m)^{2p}$$

$$\leq C_6 n^{2-3a}(nh_n^p)^{-1/2} \sum_{m=n_1}^{n} m^{3(a-1)} \quad \text{for all } n \geq n_1.$$

It is easy to see that (H4) implies (H3). We now consider the following two cases for the value of a.

Case (i). $2/3 < a \leq 1$.

It follows from Lemma 1.2.2 that

$$\sum_{m=n_1}^{n} m^{3(a-1)} \sim (3a-2)^{-1} n^{3a-2} \quad \text{as } n \to \infty. \tag{1.3.46}$$

By making use of (H3), (1.3.45) and (1.3.46) we obtain

$$s_n^{-3} \sum_{m=n_1}^{n} E|U_m|^3 \to 0 \quad \text{as } n \to \infty. \tag{1.3.47}$$

Case (ii). $a = 2/3$.

Since $\sum_{m=n_1}^{n} m^{-1} < \log n$, we have (1.3.47) from (H4) and (1.3.45).

Hence (1.3.47) holds for $2/3 \leq a \leq 1$. From (1.2.3) it follows that

$$0 < nh_n^p \gamma_n^2 \leq C_7 n^{1-2a} h_n^p \quad \text{for all } n \geq 1,$$

which yields, together with $a \geq 2/3$ and (H1), that

$$\lim_{n\to\infty} nh_n^p \gamma_n^2 = 0. \tag{1.3.48}$$

By using the fact that $Bf(x) > 0$, (1.3.39) and (1.3.48) we get

$$\lim_{n\to\infty} s_n^2 = \infty, \tag{1.3.49}$$

which implies that

$$s_n^{-3} \sum_{m=1}^{n_1-1} E|U_m|^3 \to 0 \quad \text{as } n\to\infty. \tag{1.3.50}$$

Combining (1.3.47) and (1.3.50) we obtain (1.3.40). Since

$$(nh_n^p)^{1/2} \sum_{m=1}^{n} a_m \beta_{mn} Z_m = (nh_n^p)^{1/2} \gamma_n s_n \cdot s_n^{-1} S_n,$$

we have (1.3.38) from (1.3.39) and (1.3.41). Therefore the proof is complete.

The following theorem gives the asymptotic normality of f_n.

THEOREM 1.3.6. Let x be a point at which f is loc. Lip. λ. Suppose that $\{h_n\}$ satisfies (H1),(H2),(H4),(1.3.36) and (1.3.30) with a being given in (1.3.36). Furthermore assume the condition

$$\lim_{n\to\infty} nh_n^{2\lambda+p} = 0. \tag{1.3.51}$$

Then it holds that

$$(nh_n^p)^{1/2}(f_n(x) - f(x)) \xrightarrow[L]{} N(0,Bf(x)) \quad \text{as } n\to\infty, \tag{1.3.52}$$

where B is given in Theorem 1.3.2.

PROOF. Let Z_m and δ_m be given in (1.3.17) and (1.3.26), respectively. Then we can rewrite $f_n(x) - f(x)$ as

$$f_n(x) - f(x) \tag{1.3.53}$$

$$= \beta_{0n}(f_0(x) - f(x)) + \sum_{m=1}^{n} a_m \beta_{mn} Z_m + \sum_{m=1}^{n} a_m \beta_{mn} \delta_m$$

for all $n \geq 1$. Definition 1.2.6 implies (1.3.37). Thus by virtue of Lemma 1.3.5 (1.3.38) holds. If it holds that

$$(nh_n^p)^{1/2}\beta_{0n}(f_0(x) - f(x)) \to 0 \quad \text{as } n \to \infty \tag{1.3.54}$$

and

$$(nh_n^p)^{1/2} \sum_{m=1}^{n} a_m \beta_{mn} \delta_m \to 0 \quad \text{as } n \to \infty, \tag{1.3.55}$$

then by the use of (1.3.38), (1.3.53), (1.3.54), (1.3.55) and corollary of Chung [8] (page 93) we can obtain (1.3.52).

We next show (1.3.54) and (1.3.55). By (1.2.3), (K2) and (1.3.37) we get

$$(nh_n^p)^{1/2}\beta_{0n}|f_0(x) - f(x)| \leqq C_3 n^{1/2-a} h_n^{p/2}.$$

Hence (H1) together with $a \geqq 2/3$ implies (1.3.54). In view of (1.2.2), (1.3.30) and Lemma 1.3.3 we have

$$(nh_n^p)^{1/2} \left| \sum_{m=1}^{n} a_m \beta_{mn} \delta_m \right| \tag{1.3.56}$$

$$\leqq C_4 (nh_n^p)^{1/2} n^{-a} \sum_{m=1}^{n} m^{a-1} h_m^{\lambda} \leqq C_5 (nh_n^{2\lambda+p})^{1/2}.$$

From (1.3.51) and (1.3.56) we get (1.3.55). This completes the proof.

We shall give an example of $\{h_n\}$.

EXAMPLE.

Let $h_n = n^{-r/p}$ for $n \geqq 1$. If $1/2 < a \leqq 1$ and $0 < r < \min(ap/\lambda, 1)$, then $\{h_n\}$ satisfies all of (H1) $\sim$ (H5), (1.3.15) with $\beta = (2a+r-1)^{-1}$ and (1.3.30). If $1/2 < a \leqq 1$ and $p/(2\lambda+p) < r < \min(ap/\lambda, 1)$, then $\{h_n\}$ satisfies (1.3.51) in addition to (H1) $\sim$ (H5), (1.3.15) with $\beta = (2a+r-1)^{-1}$ and (1.3.30).

1.4. CHOICE OF THE COEFFICIENT

Let the estimators $\hat{f}_n$ and f_n with the kernel K satisfying (K1) $\sim$ (K3) be given by ($\hat{F}$) and (F), respectively. In this section we shall discuss the asymptotic rate of variances between $\hat{f}_n$ and f_n, and we shall give an optimal choice, in a certain sense, of the coefficient a in (A).

THEOREM 1.4.1. Let $\{h_n\}$ satisfy (H1) and (H2) with $n_0 = 1$. Suppose the following condition be satisfied:

For some a in (A) with $1/2 < a \leqq 1$ there exist two positive constants α with $\alpha \leqq 1$ and β such that

$$n^{1-2a} h_n^p \sum_{m=1}^{n} m^{2(a-1)} h_m^{-p} \to \beta \quad \text{as } n \to \infty \tag{1.4.1}$$

and

$$n^{-1} h_n^p \sum_{m=1}^{n} h_m^{-p} \to \alpha \quad \text{as } n \to \infty. \tag{1.4.2}$$

Suppose that f(x) is continuous on R^p. Then for each point x with $f(x) > 0$

$$\lim_{n\to\infty} \mathrm{Var}(f_n(x))/\mathrm{Var}(\hat{f}_n(x)) = a^2\beta/\alpha. \tag{1.4.3}$$

PROOF. From Theorem 3 of Yamato [38] it follows that

$$\lim_{n\to\infty} nh_n^p \mathrm{Var}(\hat{f}_n(x)) = \alpha f(x) \int K^2(y)dy \ (>0), \tag{1.4.4}$$

which together with (1.3.16) implies (1.4.3). This completes the proof.

COROLLARY 1.4.2. Let $\{h_n\}$ be given by

$$h_n = n^{-r/p} \quad \text{where } 0 < r < 1. \tag{1.4.5}$$

If $f(x)$ is continuous on R^p, then for each point x with $f(x) > 0$

$$\lim_{n\to\infty} \operatorname{Var}(f_n(x))/\operatorname{Var}(\hat{f}_n(x)) = a^2(1+r)/(2a+r-1) , \tag{1.4.6}$$

where $1/2 < a \leqq 1$. Furthermore, if for each fixed $0 < r < 1/2$ we put $a = 1 - r$, then

$$\lim_{n\to\infty} \operatorname{Var}(f_n(x))/\operatorname{Var}(\hat{f}_n(x)) = 1 - r^2 . \tag{1.4.7}$$

PROOF. As is seen in Example of Section 1.3, all the conditions of Theorem 1.4.1 are satisfied by taking $\alpha = (1+r)^{-1}$ and $\beta = (2a+r-1)^{-1}$, where $1/2 < a \leqq 1$. Hence (1.4.6) is a direct consequence of (1.4.3).

Now we shall discuss an optimal choice of the coefficient a. Let $\{h_n\}$ be given in (1.4.5). It is easy to see that $a^2(1+r)/(2a+r-1)$ in the right hand side of (1.4.6) achives its minimum at $a = 1-r$ as a function of a for each fixed $r \in (0,1)$. Thus, taking account of the condition that $1/2 < a \leqq 1$, for each fixed $0 < r < 1/2$, $a = 1-r$ gives the minimum value of the asymptotic rate of variances, $\lim_{n\to\infty} \operatorname{Var}(f_n(x))/\operatorname{Var}(\hat{f}_n(x))$. For each fixed $1/2 \leqq r < 1$, the closer the coefficient a is to 1/2, the better it is in the sense of making the asymptotic rate of variances smaller.

We next consider the speed of convergence of variance. By Theorem 1.3.2 and (1.4.4) we have

$$\mathrm{Var}(f_n(x)) = \mathrm{Var}(\hat{f}_n(x)) = O(n^{-1+r}) \quad \text{as } n \to \infty .$$

Thus the closer r is to 0, the better it is from the viewpoint of the speed of convergence of variance. If $2/3 \leqq a \leqq 1$ and $p/(2\lambda+p) < r < \min\{ap/\lambda , 1\}$ in (1.4.5), then the sequence $\{h_n\}$ satisfies all the conditions in the previous sections. Hence, in this case, $a = 2/3$ is the best and the closer r is to $\min\{ap/\lambda , 1\}$, the better it is from the viewpoint of the asymptotic rate of variances.

2. SEQUENTIAL DENSITY ESTIMATION

2.1. STATEMENT OF THE PROBLEM

Let f be a nonparametric probability density function on the real line R. There is a vast literature on nonparametric density estimation (see Devroye and Györfi [12]). In this chapter we shall treat the problem of sequential estimation of f(x) at a given point x. Let $X_1, X_2, X_3 \cdots$ be a sequence of independent and identically distributed random variables, with the common density function f, on a probability space $(\Omega, \mathcal{F}, P)$. The estimator is of the following form:

$$f_n(x) = n^{-1} \sum_{j=1}^{n} K_j(x, X_j) \quad \text{for } n \geqq 1.$$

Here,

$$K_n(x,y) = h_n^{-1} K((x-y)/h_n) \quad \text{for } x, y \in R \text{ and } n \geqq 1, \tag{2.1.1}$$

K is a given bounded probability density function on R satisfying

$$\lim_{|x| \to \infty} |x| K(x) = 0, \quad \int x K(x)\,dx = 0 \text{ and } \int x^2 K(x)\,dx < \infty$$

and

$$h_n = n^{-r} \quad \text{with } 1/5 < r < 1, \tag{2.1.2}$$

where the domain of integral is R throughout this chapter.

The problem of sequential estimation of f(x) by fixed-width confidence intervals has been studied by Carroll [6] and several authors (see Prakasa Rao [25]). For given α $(0 < \alpha < 1)$ and $d > 0$ Carroll [6] defined stopping rules N_d as the first n such that $nh_n \geqq (b/d)^2 f_n(x)$, with some $b > 0$, and then took $I_{N_d}(x) = [f_{N_d}(x) - d\,,\, f_{N_d}(x) + d]$ as a 2d-width confidence interval for f(x).

Stute [33] considered the following class of stopping rules: He constructs a sequence $\{I_n(x), n \geq 1\}$ of random intervals, defines N_d as the first n such that the length of $I_n(x) \leq 2d$ and takes $I_{N_d}(x)$ as a 2d-width confidence interval. The second class of stopping rules as considered in Carroll [6] is of this form. For both of the classes of stopping rules they showed that the probability $P\{f(x) \varepsilon I_{N_d}(x)\}$ converges to α as d tends to zero.

The aim of this chapter is to introduce a certain class of stopping rules N_d of the first type and derive the rate at which $P\{f(x) \varepsilon I_{N_d}(x)\}$ converges to α as d tends to zero. The main tool in the proof is the result of Rychlik [28] on the convergence rate in the random central limit theorem.

The chapter consists of four sections. In Section 2.2 the main theorem is presented, that is, the rate of convergence of $P\{f(x) \varepsilon I_{N_d}(x)\}$ tending to α is given. Section 2.3 contains several lemmas that will be used for proving the main theorem in Section 2.4. In Section 2.4 the main theorem will be proved.

2.2. MAIN RESULT

In this section we define stopping rules N_d and a class of fixed-width sequential confidence intervals $\{I_{n,d}(x), n \geq 1\}$, and derive the rate at which $P\{f(x) \varepsilon I_{N_d,d}(x)\}$ converges to α as d tends to zero.

Fix any $d > 0$ and define a confidence interval of fixed-width 2d as $I_{n,d}(x) = [f_n(x) - d, f_n(x) + d]$ for each $n \geq 1$. For given α

$(0 < \alpha < 1)$ choose $D = D_\alpha > 0$ such that $\Phi(D) - \Phi(-D) = \alpha$, where Φ is the standard normal distribution function. Define stopping rules $N_d = N_d(x)$ for each x as follows:

N_d = first $n \geq 1$ such that

$$(D^2B)^{-1}n^{1-r}d^2 \geq f_n(x) + n^{-1} \qquad (2.2.1)$$

where

$$B = (1+r)^{-1} \int K^2(y)dy. \qquad (2.2.2)$$

The following is our main theorem.

THEOREM. Assume that the density f has a bounded second derivative f'' on R. Then for each x with $f(x) > 0$ we obtain

$$P\{f(x) \varepsilon I_{N_d,d}(x)\} = \alpha + O(d^\eta) \quad \text{as } d \to 0,$$

where

$$\eta = \min\{r/2 \, , \, 2(1-r)/5(2-r) \, , \, (5r-1)/(1-r)\}.$$

REMARK 2.2.1. By the same argument for Theorem 3.3 of Isogai [15] it can be shown that for each x in the above theorem and each $d > 0$ $P\{N_d < \infty\} = 1$.

2.3. NOTATION AND PRELIMINARY LEMMAS

Throughout Sections 2.3 and 2.4 let an arbitrary point x with $f(x) > 0$ and the constant r in (2.1.2) be fixed. In this chapter we use the following notation unless otherwise stated:

$Z_n = K_n(x,X_n) - EK_n(x,X_n)$, $\delta_n = EK_n(x,X_n) - f(x)$,

$$S_n = \sum_{j=1}^{n} Z_j \ , \ \sigma_n^2 = \mathrm{Var}(Z_n) \ , \ s_n^2 = \sum_{j=1}^{n} \sigma_j^2 \ , \ B_n^3 = \sum_{j=1}^{n} E|Z_j|^3 ,$$

$$W_n = n^{-(1+r)/2} S_n/\sqrt{Bf(x)} \ , \ V_n = n^{-(1+r)/2} \sum_{j=1}^{n} \delta_j/\sqrt{Bf(x)} \ ,$$

where $K_n(x,y)$ and B are as defined in (2.1.1) and (2.2.2), respectively. When there is no confusion, we drop the argument x in $f_n(x)$, $f(x)$ and $I_{n,d}(x)$, and write them as f_n, f and $I_{n,d}$. Throughout this chapter $C, C_1, C_2, \cdots$ denote positive constants appropriately chosen for the context in which they appear, [a] stands for the greatest integer less than or equal to a, and the symbol " $\sim$ " means asymptotic equivalence.

LEMMA 2.3.1. Under the conditions of the preceding theorem, $s_n^2/(Bf(x)n^{1+r}) = 1 + O(n^{-r})$ as $n \to \infty$.

PROOF. It is easy to see that

$$|n^{1-r}\mathrm{Var}(f_n) - Bf|$$

$$\leqq |n^{-(1+r)} \sum_{j=1}^{n} EK_j^2(x,X_j) - Bf| + n^{-(1+r)} \sum_{j=1}^{n} \{EK_j(x,X_j)\}^2$$

$$= I + II, \text{ say.}$$

By the Taylor theorem and the conditions on f and K we get that

$$EK_n(x,X_n) = f(x) + O(n^{-2r}), \tag{2.3.1}$$

which implies that

$$|EK_n(x,X_n)| \leqq C_1 \quad \text{for all } n \geqq 1. \tag{2.3.2}$$

Hence we have that $II = O(n^{-r})$. It is obvious that

$$I \leq n^{-(1+r)} \sum_{j=1}^{n} j^r \left| \int K^2(y)\{f(x-j^{-r}y) - f(x)\}dy \right|$$

$$+ \left| n^{-(1+r)} \sum_{j=1}^{n} j^r - (1+r)^{-1} \right| f(x) \int K^2(y)\,dy$$

$$= III + IV, \text{ say.}$$

By the use of the Taylor theorem and the conditions on K we have $\int K^2(y)\{f(x-j^{-r}y) - f(x)\}dy = O(j^{-r})$, which yields that $III = O(n^{-r})$. Since $0 < n^{-(1+r)} \sum_{j=1}^{n} j^r - (1+r)^{-1} < n^{-1}$ for all $n \geq 1$, we get that $IV = O(n^{-1})$. Thus, taking account of the fact that $\mathrm{Var}(f_n) = s_n^2/n^2$, we obtain the statement of this lemma.

The following lemma was proved by Rychlik [28].

LEMMA 2.3.2. Let $\{X_n, n \geq 1\}$ be a sequence of independent random variables such that $EX_n = 0$, $\mathrm{Var}(X_n) = \alpha_n^2$ and $E|X_n|^3 = \beta_n^3 < \infty$, and let $\{N_n, n \geq 1\}$ be a sequence of positive integer-valued random variables. Set $S_n = \sum_{j=1}^{n} X_j$, $s_n^2 = \sum_{j=1}^{n} \alpha_j^2$ and $B_n^3 = \sum_{j=1}^{n} \beta_j^3$. If $B_n^3/s_n^3 \leq \sqrt{\varepsilon_n}$ and $P\{|s_{N_n}^2/\tau s_n^2 - 1| > \varepsilon_n\} = O(\sqrt{\varepsilon_n})$ for some constant $\tau > 0$ and a nonnegative sequence $\{\varepsilon_n, n \geq 1\}$ with $\varepsilon_n \to 0$ as $n \to \infty$, then

$$\sup_{-\infty<t<\infty} |P\{S_{N_n} \leq t s_{N_n}\} - \Phi(t)| = O(\sqrt{\varepsilon_n}).$$

The rate of approach of the distribution of S_{N_d}/s_{N_d} to the standard normal distribution is given by the next lemma.

LEMMA 2.3.3. Let $\eta = \min\{r/2, 2(1-r)/5(2-r)\}$. Then, under the conditions of the preceding theorem we have

$$\sup_{-\infty<t<\infty} |P\{S_{N_d} \leqq ts_{N_d}\} - \Phi(t)| = O(d^\eta) \quad \text{as } d \to 0.$$

PROOF. Let $\{d_k, k \geqq 1\}$ be any sequence of positive numbers converging to zero as k tends to infinity. It suffices to prove the lemma for this sequence. Throughout this proof let k be sufficiently large, so that d_k is sufficiently small. First we shall show that

$$B_n^3/s_n^3 = O(n^{(r-1)/2}). \tag{2.3.3}$$

By Hölder's inequality, the boundedness of K and (2.3.2) we get that $E|Z_n|^3 = O(n^{2r})$, which implies that

$$B_n^3 = O(n^{1+2r}). \tag{2.3.4}$$

Hence, according to (2.3.4) and Lemma 2.3.1 we have (2.3.3). Set

$$C_1 = D^2B, \ \beta = (1-r)/(1+r), \ \tau^\beta = C_1 f$$

and

$$n_k = [d_k^{-2/(1-r)}]. \tag{2.3.5}$$

Then, we shall show that

$$P\{|s_{N_k}^2/\tau s_{n_k}^2 - 1| > \zeta d_k^{2\eta}\} = O(d_k^\eta) \tag{2.3.6}$$

for any fixed $\zeta > 0$, where $N_k = N_{d_k}$. Let any $\zeta > 0$ be fixed. For notational simplicity we drop the subscript k of N_k, n_k and d_k throughout the proof of (2.3.6). By (2.2.1), Remark 2.2.1 and the fact that $f_n \geqq 0$ we get that $N^{2-r} \geqq C_1 d^{-2}$ almost surely. Hence, putting

$$g_1 = g_1(d) = [(C_1 d^{-2})^{1/(2-r)}], \tag{2.3.7}$$

we have that

$$P\{N \geqq g_1\} = 1. \tag{2.3.8}$$

Set

$$b(m) = b_r(m) = \begin{cases} m^{-2r} & \text{if } 1/5 < r < 1/2 \\ (1 + \log m)/m & \text{if } r = 1/2 \\ m^{-1} & \text{if } 1/2 < r < 1. \end{cases} \tag{2.3.9}$$

Since $m^{-1} \sum_{j=1}^{m} j^{-2r} \leqq C_2 b(m)$, it follows from (2.3.1) that

$$|m^{-1} \sum_{j=1}^{m} EK_j(x, X_j) - f| \leqq Cb(m) \quad \text{for } m \geqq 1. \tag{2.3.10}$$

By Lemma 2.3.1 we get

$$|s_m^2/Bfm^{1+r} - 1| \leqq C_3 m^{-r} \quad \text{for } m \geqq 1. \tag{2.3.11}$$

It is obvious that

$$P\{|s_N^2/\tau s_n^2 - 1| > \zeta d^{2\eta}\} \tag{2.3.12}$$

$$= P\{s_N^2 < (1 - \zeta d^{2\eta})\tau s_n^2\} + P\{s_N^2 > (1 + \zeta d^{2\eta})\tau s_n^2\}$$

$$= I + II, \text{ say.}$$

Set

$$\rho = \rho(d) = C_3(g_1^{-r} + n^{-r})/(1 - C_3 g_1^{-r}) \quad (> 0),$$

$$g_2 = g_2(d) = [\{\tau(1 + \rho)(1 - \zeta d^{2\eta})\}^{1/(1+r)} n]$$

and

$$A_1 = A_1(d) = C_1^{-1} g_2^{1-r} d^2 - f + Cb(g_1).$$

Then, it follows from (2.2.1), (2.3.8), (2.3.10), (2.3.11), the definition of f_n and the monotonicity of $b(m)$ that

$$I \leqq P\{g_1 \leqq N \leqq g_2\} \tag{2.3.13}$$

$$\leqq P\{j^{-1}S_j < A_1 \quad \text{for some } g_1 \leqq j \leqq g_2\}.$$

From (2.3.5),(2.3.7) and (2.3.9) we have that

$$b(g_1) = o(d^{2\eta}) \quad \text{and} \quad \rho = o(d^{2\eta}), \tag{2.3.14}$$

where $o(d)$ means that $o(d)/d \to 0$ as $d \to 0$. By using the inequality that $(1-t)^{\xi} < 1 - \xi t$ for $0 < t < 1$ and $0 < \xi < 1$ we obtain that

$$C_1^{-1} g_2^{1-r} d^2 - f \leqq f\{(1+\rho)^{\beta}(1 - \zeta d^{2\eta})^{\beta} - 1\}$$

$$\leqq - f\beta\zeta d^{2\eta}\{1 - (\beta\zeta)^{-1}\rho d^{-2\eta} + \rho\},$$

which, together with (2.3.14), yields that $\limsup_{d\to 0} A_1/d^{2\eta} \leqq - f\beta\zeta < 0$. From this we get that $A_1 \leqq - C_4 d^{2\eta}$ for d sufficiently small. Hence, by virtue of the Hájek-Rényi inequality (see Petrov [24], page 51) and (2.3.13) we have that

$$I \leqq P\{\max_{g_1 \leqq j \leqq g_2} j^{-1}|S_j| > C_4 d^{2\eta}\} \tag{2.3.15}$$

$$\leqq C_5 d^{-4\eta}\{g_1^{-2} \sum_{j=1}^{g_1} EK_j^2(x,X_j) + \sum_{j=g_1+1}^{g_2} j^{-2} EK_j^2(x,X_j)\}.$$

From (2.3.2) and the boundedness of K we get that

$$g_1^{-2} \sum_{j=1}^{g_1} EK_j^2(x,X_j) \leqq C_6 g_1^{-2} \sum_{j=1}^{g_1} j^r \leqq C_7 g_1^{r-1} \tag{2.3.16}$$

and

$$\sum_{j=g_1+1}^{g_2} j^{-2} EK_j^2(x,X_j) \leqq C_7 g_1^{r-1},$$

which, together with (2.3.15), imply that $I \leqq C_8 d^{-4\eta} g_1^{r-1}$. Thus, by (2.3.7) and the definition of η we obtain

$$I = O(d^{\eta}). \tag{2.3.17}$$

Next, we shall estimate II in (2.3.12). Put

$$\rho = C_3(g_1^{-r} + n^{-r})/(1 + C_3 g_1^{-r}),$$

$$g_3 = [\{\tau(1-\rho)(1+\zeta d^{2\eta})\}^{1/(1+r)} n]$$

and

$$A_2 = C_1^{-1} g_3^{1-r} d^2 - f - (C+1) b(g_3).$$

Then, by the same argument for (2.3.13) we have that

$$II \leqq P\{N \geqq g_3 + 1\} \leqq P\{g_3^{-1} S_{g_3} > A_2\}. \tag{2.3.18}$$

By making use of (2.3.5) and the inequalities that $(1+t)^{\xi} \geqq 1 + \xi t/2$ for small t and $\xi > 0$ and that $(t-1)^{\xi} > t^{\xi} - 1$ for $t > 1$ and $0 < \xi < 1$ we obtain that

$$C_1^{-1} g_3^{1-r} d^2 - f$$

$$> C_1^{-1}\{\{\tau(1-\rho)(1+\zeta d^{2\eta})\}^{1/(1+r)} n - 1\}^{1-r} d^2 - f$$

$$> f\{(1-\rho)^{\beta}(1+\zeta d^{2\eta})^{\beta}(1-d^2) - 1\} - C_1^{-1} d^2$$

$$> f\{(1-\rho)(1+(\beta\zeta/2)d^{2\eta}) - 1\} - C_9 d^2$$

$$= (f\beta\zeta/2) d^{2\eta}\{1 - (2/\beta\zeta)\rho d^{-2\eta} - \rho\} - C_9 d^2.$$

Hence, in view of the fact that $b(g_3) = o(d^{2\eta})$ and $\rho = o(d^{2\eta})$ we have that $\liminf_{d \to 0} A_2/d^{2\eta} \geqq f\beta\zeta/2 > 0$, which yields that $A_2 \geqq C_{10} d^{2\eta}$ for d sufficiently small. Thus, according to Chebychev's inequality, (2.3.16),(2.3.18), the fact that $g_3^{r-1} = O(d^2)$ and the definition of η we obtain that $II = o(d^{\eta})$, which, together with (2.3.12) and (2.3.17), yields (2.3.6). Now, let

$$L_n = \begin{cases} N_k & \text{if } n = n_k \text{ for some } k \geqq 1 \\ [\tau^{1/(1+r)} n] & \text{otherwise} \end{cases}$$

and

$$\varepsilon_n = \begin{cases} \zeta d_k^{\eta} & \text{if } n = n_k \text{ for some } k \geqq 1 \\ \zeta n^{-\eta} & \text{otherwise,} \end{cases}$$

where τ and n_k are as defined in (2.3.5), and the positive constant ζ will be chosen later. Without loss of generality we may assume that $L_n \geqq 1$ for all $n \geqq 1$. Hence by Remark 2.2.1 $\{L_n, n \geqq 1\}$ is a sequence of positive integer-valued random variables. First we shall show that

$$|s_{L_n}^2/\tau s_n^2 - 1| < \varepsilon_n^2 \tag{2.3.19}$$

for n ($\neq n_k$) sufficiently large. Let $n \neq n_k$ for all $k \geqq 1$ be sufficiently large. By the use of (2.3.11) and the definition of η we get that

$$\begin{aligned} & n^{2\eta}(s_{L_n}^2/\tau s_n^2 - 1) \\ & < n^{2\eta}\{((1 + C_3 L_n^{-r})/(1 - C_3 n^{-r}))(L_n/\tau^{1/(1+r)} n)^{1+r} - 1\} \\ & \leqq n^{2\eta - r} C_3\{(n/L_n)^r + 1\}/(1 - C_3 n^{-r}) \leqq C_{11}, \end{aligned}$$

which implies that

$$s_{L_n}^2/\tau s_n^2 - 1 < C_{11} n^{-2\eta}. \tag{2.3.20}$$

Set $q_n = (1 - C_3 L_n^{-r})/(1 + C_3 n^{-r})$. By the use of (2.3.11) and the inequality that $(1 - t)^{\xi} > 1 - \xi t$ for $0 < t < 1$ and $\xi > 1$ we have that

$$n^{2\eta}(s_{L_n}^2/\tau s_n^2 - 1)$$

$$> n^{2\eta}\{q_n(1 - (\tau^{1/(1+r)}n)^{-1})^{1+r} - 1\}$$

$$> n^{2\eta}(q_n - 1) - (1 + r)\tau^{-1/(1+r)}q_n n^{2\eta-1}$$

$$\sim - C_3 (\tau^{-r/(1+r)} + 1)n^{2\eta-r} \geqq - C_3 (\tau^{-r/(1+r)} + 1),$$

which yields that

$$s_{L_n}^2/\tau s_n^2 - 1 > - C_{12}\, n^{-2\eta}. \qquad (2.3.21)$$

It follows from (2.3.3) and the definition of η that

$$B_m^3/s_m^3 \leqq C_{13}m^{-\eta} \quad \text{for } m \geqq 1. \qquad (2.3.22)$$

From (2.3.5) it is clear that

$$n_k^{-\eta} \leqq (C_{14}/C_{13})d_k^{\eta} \quad \text{for } k \geqq 1. \qquad (2.3.23)$$

Here, ζ is chosen as $\max\{C_{11}, C_{12}, C_{13}, C_{14}\} + 1$. Then the relations (2.3.20) and (2.3.21) imply (2.3.19). Thus, by virtue of (2.3.6) and (2.3.19) we have

$$P\{|s_{L_n}^2/\tau s_n^2 - 1| > \varepsilon_n^2\} = O(\varepsilon_n). \qquad (2.3.24)$$

From (2.3.22) and (2.3.23) we get

$$B_n^3/s_n^3 \leqq \varepsilon_n \quad \text{for } n \geqq 1. \qquad (2.3.25)$$

Therefore, since the conditions of Lemma 2.3.2 are satisfied by (2.3.24) and (2.3.25) we have that

$$\sup_{-\infty<t<\infty} |P\{S_{L_n} \leqq ts_{L_n}\} - \Phi(t)| = O(\varepsilon_n).$$

In particular, putting $n = n_k$ we obtain the result of Lemma 2.3.3. Thus the lemma was proved.

The next lemma is due to Michel and Pfanzagl [21].

LEMMA 2.3.4. Let $\{X_n, n \geqq 1\}$ and $\{Y_n, n \geqq 1\}$ be two sequences of random variables. Assume that

$$\sup_{-\infty<t<\infty} |P\{X_n \leqq t\} - \Phi(t)| = O(a_n)$$

and

$$P\{|Y_n - 1| > a_n\} = O(a_n).$$

Then

$$\sup_{-\infty<t<\infty} |P\{X_n \leqq tY_n\} - \Phi(t)| = O(a_n).$$

2.4. PROOF OF THE THEOREM

In this section we prove the theorem presented in Section 2.2. Throughout this section let $N = N_d$. It is easy to see that

$$|P\{f \varepsilon I_{N,d}\} - \alpha| \leqq 2 \sup_t |P\{f_N - f \leqq tD^{-1}d\} - \Phi(t)|. \qquad (2.4.1)$$

Set $C_1 = D^2B$. If it holds that as $d \to 0$

$$\sup_t |P\{N^{(1-r)/2}(f_N - f) \leqq t\sqrt{Bf}\} - \Phi(t)| = O(d^\eta) \qquad (2.4.2)$$

and

$$P\{|(d^2N^{1-r}/fC_1)^{1/2} - 1| > d^\eta\} = O(d^\eta), \qquad (2.4.3)$$

then by Lemma 2.3.4 we have that

$$\sup_t |P\{f_N - f \leqq tD^{-1}d\} - \Phi(t)| = O(d^\eta) \quad \text{as } d \to 0,$$

which, together with (2.4.1), yields the theorem. Hence it suffices to show (2.4.2) and (2.4.3). First we shall show (2.4.3). By the use of the inequality that

$$|y-1| > z \quad \text{if} \quad |\sqrt{y}-1| > \sqrt{z} \quad \text{for } y > 0 \text{ and } z > 0, \tag{2.4.4}$$

it is sufficient to show that

$$P\{|d^2N^{1-r}/fC_1 - 1| > d^{2\eta}\} = O(d^{\eta}). \tag{2.4.5}$$

Put $g_1 = [(C_1 d^{-2})^{1/(2-r)}]$, $g_2 = [\{fC_1(1-d^{2\eta})d^{-2}\}^{1/(1-r)}]$ and $g_3 = [\{fC_1(1+d^{2\eta})^{-2}\}^{1/(1-r)}]$. In view of (2.3.8) we easily get

$$\begin{aligned} &P\{|d^2N^{1-r}/fC_1 - 1| > d^{2\eta}\} \\ &\leqq P\{g_1 \leqq N \leqq g_2\} + P\{N \geqq g_3 + 1\} = I + II, \text{ say.} \end{aligned} \tag{2.4.6}$$

By means of arguments similar to I and II in (2.3.12) we can prove that $I + II = O(d^{\eta})$, which, together with (2.4.6), yields (2.4.5). Next, we shall show that

$$\sup_t |P\{W_N \leqq t\} - \Phi(t)| = O(d^{\eta}). \tag{2.4.7}$$

Suppose the following relation be valid:

$$P\{|s_N^2/BfN^{1+r} - 1| > d^{2\eta}\} = O(d^{\eta}). \tag{2.4.8}$$

Then from (2.4.4) and (2.4.8) we get that

$$P\{|s_N/(BfN^{1+r})^{1/2} - 1| > d^{\eta}\} = O(d^{\eta}). \tag{2.4.9}$$

It is easy to see that for a random variable Y and a positive number ε ($< 1/2$) $P\{|Y^{-1} - 1| > 2\varepsilon\} \leqq 2P\{|Y - 1| > \varepsilon\}$. By this fact and (2.4.9) we have

$$P\{|(BfN^{1+r})^{1/2}/s_N - 1| > 2d^{\eta}\} = O(d^{\eta}). \tag{2.4.10}$$

Hence, by virtue of (2.4.10), Lemmas 2.3.3. and 2.3.4 we obtain (2.4.7). Thus, in order to prove (2.4.7) it suffices to prove (2.4.8). Lemma 2.3.1 implies that

$$|s_n^2/Bfn^{1+r} - 1| \leqq C_2 n^{-r} \quad \text{for } n \geqq 1. \tag{2.4.11}$$

Since $C_2 g_1^{-r} < d^{2\eta}$ for d sufficiently small, it follows from (2.3.8) and (2.4.11) that

$$P\{|s_N^2/BfN^{1+r} - 1| > d^{2\eta}\}$$

$$= \sum_{n=g_1}^{\infty} P\{d^{2\eta} < |s_n^2/Bfn^{1+r} - 1| \leqq C_2 n^{-r}\ ,\ N=n\}$$

$$\leqq \sum_{n=g_1}^{\infty} P\{d^{2\eta} < C_2 g_1^{-r}\ ,\ N=n\} = 0 \quad \text{for d sufficiently small.}$$

Hence (2.4.8) holds. Last of all, we shall prove (2.4.2). Let I_d, $d > 0$, be the closed interval given by $I_d = [g_2, g_3]$. Since $(n^{1-r}/Bf)^{1/2}(f_n - f) = W_n + V_n$ and V_n is finite for each $n \geqq 1$ by (2.3.2), we get that

$$|P\{N^{(1-r)/2}(f_N - f) \leqq t\sqrt{Bf}\} - \Phi(t)| \tag{2.4.12}$$

$$\leqq |P\{W_N \leqq t - V_N\ ,\ N\varepsilon I_d\} - \Phi(t)| + P\{N\not\varepsilon I_d\}$$

for all $t\varepsilon R$. From (2.4.5) we have

$$P\{N\not\varepsilon I_d\} = O(d^{\eta}). \tag{2.4.13}$$

Set $M_d = \max\{|V_j| : j\varepsilon I_d\}$. Since

$$P\{W_N \leqq t - M_d\ ,\ N\varepsilon I_d\} \leqq P\{W_N \leqq t - V_N\ ,\ N\varepsilon I_d\} \leqq P\{W_N \leqq t + M_d\}$$

and $|\Phi(x) - \Phi(y)| \leqq |x - y|$ for $x,y\varepsilon R$, we obtain that

$$|P\{W_N \leqq t - V_N\ ,\ N\varepsilon I_d\} - \Phi(t)| \tag{2.4.14}$$

$$\leqq |P\{W_N \leqq t - M_d\} - \Phi(t - M_d)| + |\Phi(t - M_d) - \Phi(t)| + P\{N\not\varepsilon I_d\}$$

$$+ |P\{W_N \leqq t + M_d\} - \Phi(t + M_d)| + |\Phi(t + M_d) - \Phi(t)|$$

$$\leqq 2\{\sup_t |P\{W_N \leqq t\} - \Phi(t)| + M_d\} + P\{N\not\varepsilon I_d\} \quad \text{for all } t\varepsilon R.$$

Hence, according to (2.4.7),(2.4.13) and (2.4.14) we get that

$$\sup_t |P\{W_N \leqq t - V_N, N\varepsilon I_d\} - \Phi(t)| \leqq 2M_d + O(d^\eta). \tag{2.4.15}$$

By the use of the Taylor theorem and the conditions on f and K we have that $|\delta_n| \leqq C_3 n^{-2r}$. Hence, it holds that $|V_n| \leqq C_4 b(n)$, where

$$b(n) = \begin{cases} n^{-(5r-1)/2} & \text{if } 1/5 < r < 1/2 \\ n^{-3/4}(1+\log n) & \text{if } r = 1/2 \\ n^{-(1+r)/2} & \text{if } 1/2 < r < 1. \end{cases}$$

Since b(n) is nonincreasing for $n \geqq 2$, we get that $M_d \leqq C_4 b(g_2)$ for small d. It follows from the definition of η that $b(g_2) = O(d^\eta)$. Hence from (2.4.15) we have that

$$\sup_t |P\{W_N \leqq t - V_N, N\varepsilon I_d\} - \Phi(t)| = O(d^\eta). \tag{2.4.16}$$

Thus, combining (2.4.12), (2.4.13) and (2.4.16) we obtain (2.4.2). Therefore the proof of the theorem is complete.

3. STOPPING RULES IN DENSITY ESTIMATION

3.1. STATEMENT OF THE PROBLEM

In this chapter we consider the problem of sequential estimation of a probability density function f at a given point x_0 on the p-dimensional Euclidean space R^p. In connection with this problem Carroll [6] proposed two classes of stopping rules. The first class is based on the idea of fixed-width interval estimation (see Chow and Robbins [7]). The idea for the second class is to find two statistics for $f(x_0)$ and to stop when the difference in those two statistics becomes at most 2d (see, e.g., Geertsema [13], and Sen and Ghosh [31]). Stute [33] treated the second class of stopping rules. In Chapter 2 that of the first type was discussed.

In this chapter we consider the following first class of stopping rules. For each $d > 0$ we define the stopping rule $N(d)$ as

$N(d)$ = smallest integer $n \geq 1$ such that

$$nh_n^p \geq (b/d)^2 f_n(x_0), \tag{3.1.1}$$

where h_n, b and $f_n(x_0)$ will be given in Section 3.2. The definition of this stopping rule $N(d)$ given by Carroll [6] is motivated for the construction of a 2d-width confidence interval for $f(x_0)$, $[f_{N(d)}(x_0) - d, f_{N(d)}(x_0) + d]$, with prescribed coverage probability. The discussion in the present chapter focusses on the limiting behavior of the moments of the stopping rule $N(d)$.

The purpose of this chapter is to show that for any given function g with certain properties

$$g(N(d)h_{N(d)}^p d^2)/g(b^2 f(x_0)) \to 1 \quad \text{a.s.} \quad \text{as } d \to 0$$

and

$$E\{g(N(d)h^p_{N(d)}d^2)\}/g(b^2f(x_0)) \to 1 \quad \text{as } d \to 0.$$

As an application of this result we obtain an approximation to the q-th moment $E\{(N(d))^q\}$ of $N(d)$ for each $q > 0$.

This chapter consists of four sections. In Section 3.2 we give assumptions. Section 3.3 presents the main result and its application. The proof of the theorem of Section 3.3 is given in Section 3.4.

3.2. ASSUMPTIONS

In this section we shall give assumptions used throughout this chapter.

Let f be a (unknown) probability density function on R^p with respect to the Lebesgue measure. Let $\{X_n, n \geqq 1\}$ be a sequence of independent and identically distributed random vectors, taking values in R^p with the common density f. For estimating f we use the recursive kernel estimators defined as follows:

$$f_0(x) \equiv 0$$

$$f_n(x) = n^{-1}\sum_{i=1}^{n} K_i(x,X_i) + n^{-1} \quad \text{for } n \geqq 1,$$

where

$$K_n(x,y) = h_n^{-p}K((x-y)/h_n) \quad \text{for } x,y \varepsilon R^p$$

Throughout this chapter we assume the following conditions on f, K and h_n, referred to as the "stated conditions":

f is continuous and bounded on R^p;

K is a bounded probability density function on R^p satisfying

$$\|u\|^p K(u) \to 0 \quad \text{as} \quad \|u\| \to \infty,$$

where $\|\cdot\|$ denotes the Euclidean norm on R^p;

$\{h_n, n \geq 1\}$ is a sequence of positive constants satisfying

$$h_n \downarrow 0 \quad \text{and} \quad nh_n^p \uparrow \infty \quad \text{as } n \to \infty, \tag{H1}$$

$$nh_n^p / \log\log n \to \infty \quad \text{as } n \to \infty, \tag{H2}$$

$$h_n / h_{n-1} \to 1 \quad \text{as } n \to \infty, \tag{H3}$$

for some constant $\beta \varepsilon (0,\infty)$

$$(h_n^p/n) \sum_{i=1}^{n} h_i^{-p} \to \beta \quad \text{as } n \to \infty. \tag{H4}$$

EXAMPLES.

After some calculations the following sequences $\{h_n, n \geq 1\}$ can be seen to satisfy all the conditions (H1) $\sim$ (H4).

(i) For each fixed integer $d \geq 1$

$$h_n = \{(n + n_0)^r (\log (n + n_0))^d\}^{-1/p} \quad \text{for } n \geq 1$$

with $0 < r < 1$, where $n_0 = [\exp\{d/(1-r)\}]$, and $[a]$ denotes the largest integer not greater than a;

(ii) $h_n = n^{-r/p}$ for $n \geq 1$ with $0 < r < 1$.

In both cases β in (H4) is equal to $(1+r)^{-1}$.

Let any $\alpha \varepsilon (0,1)$ be given. Set

$$b = \sqrt{B}\ \Phi^{-1}(1 - \alpha/2), \tag{3.2.1}$$

where

$B = \beta \int K^2(u)du$, β is as in (H4), the domain of integral is R^p

and Φ is the standard normal distribution function. Throughout this chapter let any x_0 with $f(x_0) > 0$ be fixed and we use the stopping rule N(d) defined by (3.1.1).

3.3. ASYMPTOTIC BEHAVIOR OF THE STOPPING RULE

In this section we shall give three results about the asymptotic behavior of the stopping rule N(d) defined by (3.1.1).

The following is our main result.

THEOREM 3.3.1. Let g be a continuous, strictly increasing function on $[0,\infty)$ satisfying

$$g(0) = 0, \quad g(x) \to \infty \quad \text{as } x \to \infty \quad \text{and}$$

$$\log g(x)/x \to 0 \quad \text{as } x \to \infty. \tag{3.3.1}$$

Then, under the stated conditions on f, K and h_n for any x_0 with $f(x_0) > 0$

$$g(N(d)h^p_{N(d)}d^2)/g(b^2 f(x_0)) \to 1 \quad \text{a.s.} \quad \text{as } d \to 0 \tag{3.3.2}$$

and

$$E\{g(N(d)h^p_{N(d)}d^2)\}/g(b^2 f(x_0)) \to 1 \quad \text{as } d \to 0, \tag{3.3.3}$$

where b is as in (3.2.1).

As applications of Theorem 3.3.1 we can obtain the following two corollaries.

Put $g(x) = x$ for $x \geqq 0$. Then we have

COROLLARY 3.3.2. Under the conditions of Theorem 3.3.1

$$N(d)h^p_{N(d)}/(b^2 f(x_0)d^{-2}) \to 1 \quad \text{a.s.} \quad \text{as } d \to 0$$

and

$$E\{N(d)h^p_{N(d)}\}/(b^2 f(x_0)d^{-2}) \to 1 \quad \text{as } d \to 0.$$

REMARK 3.3.1. This corollary with the result about the expectation corresponds to Lemma 4.1 of Carroll [6], though his and our assumptions are different. Also, by the same argument for Theorem 3.7 of Isogai [15], under additional assumptions on f, K and h_n we obtain

$$\lim_{d \to 0} P\{|f_{N(d)}(x_0) - f(x_0)| \leq d\} = 1 - \alpha.$$

Set $h_n = n^{-r/p}$ with $0 < r < 1$. For $q > 0$ put $g(x) = x^{q/(1-r)}$ on $[0,\infty)$. Then Theorem 3.3.1 gives

COROLLARY 3.3.3. Under the conditions of Theorem 3.3.1 for $q > 0$ and $r \in (0,1)$

$$(N(d))^q/(b^2 f(x_0)d^{-2})^{q/(1-r)} \to 1 \quad \text{a.s.} \quad \text{as } d \to 0$$

and

$$E\{(N(d))^q\}/(b^2 f(x_0)d^{-2})^{q/(1-r)} \to 1 \quad \text{as } d \to 0.$$

REMARK 3.3.2. This corollary with $q = 1$ corresponds to Lemma 4.3 of Carroll [6].

3.4. PROOF

In this section we shall prove Theorem 3.3.1 of Section 3.3. Throughout this section $C, C_1, C_2, \cdots$ denote appropriate positive constants.

The following lemma was given by Bennett [2].

LEMMA 3.4.1. Let $Y_1, \cdots, Y_n$ be a sequence of independent random variables satisfying

$$EY_i = 0 \quad \text{and} \quad |Y_i| \leqq C \quad \text{a.s.} \quad \text{for } 1 \leqq i \leqq n.$$

Set $S_n = \sum_{i=1}^{n} Y_i$ and $\sigma_n^2 = \mathrm{Var}(S_n)$. Then for any $\varepsilon > 0$

$$P\{|S_n| \geqq \varepsilon\} \leqq 2\exp\{-\varepsilon^2/2(\sigma_n^2 + C\varepsilon)\}.$$

Before proving the theorem we shall define two functions. Define a real-valued function $\eta(x)$ on $[0,\infty)$ as

$$\eta(x) = nh_n^p + \{(n+1)h_{n+1}^p - nh_n^p\}(x-n) \quad \text{for } n \leqq x < n+1,$$

where $n = 0,1,2, \cdots$ and $h_0 = h_1$. By (H1) it is easily seen that η is continuous and nondecreasing and $\eta(x) \to \infty$ as $x \to \infty$. Next, let a real-valued function $\phi(x)$ on $[0,\infty)$ be defined as

$$\phi(x) = \sup\{u \geqq 0 : \eta(u) \leqq x\} \quad \text{for } x \in [0,\infty).$$

Then by (H1) and the properties of η we can show that ϕ is strictly increasing and

$$\phi(x) \to \infty \quad \text{as } x \to \infty, \tag{3.4.1}$$

$$\eta(\phi(x)) = x \quad \text{for } x \in [0,\infty). \tag{3.4.2}$$

It follows from Theorem 2 of Devroye [10] that $f_n(x_0) \to f(x_0)$ a.s. as $n \to \infty$. Thus we can apply Lemma 1 of Chow and Robbins [7] to obtain that $N(d) < \infty$ a.s. for each $d > 0$ and

$$N(d)h^p_{N(d)}d^2 \to b^2 f(x_0) \quad (>0) \quad \text{a.s.} \quad \text{as } d \to 0,$$

which, together with the continuity of g, yields that

$$g(N(d)h^p_{N(d)}d^2) \to g(b^2 f(x_0)) \quad \text{a.s.} \quad \text{as } d \to 0. \tag{3.4.3}$$

Since $g(b^2 f(x_0)) > 0$ (3.3.2) is implied by (3.4.3). Now, if it holds that for $d_0 > 0$ chosen later

$$\sum_{m=1}^{\infty} \sup_{0<d<d_0} P\{g(N(d)h^p_{N(d)}d^2) > m\} < \infty, \tag{3.4.4}$$

by the use of (3.4.3) and Lemma 3.2 of Bickel and Yahav [3] we will obtain

$$E\{g(N(d)h^p_{N(d)}d^2)\} \to g(b^2 f(x_0)) \quad \text{as } d \to 0,$$

which implies (3.3.3) because of $g(b^2 f(x_0)) > 0$. Hence in order to prove the theorem it suffices to show (3.4.4). By the properties of g and (3.3.1) we can prove that the inverse function $g^{-1}(x)$ is continuous and strictly increasing on $[0,\infty)$ and that $g^{-1}(0) = 0$,

$$g^{-1}(x) \to \infty \quad \text{as } x \to \infty \tag{3.4.5}$$

and

$$g^{-1}(x)/\log x \to \infty \quad \text{as } x \to \infty. \tag{3.4.6}$$

It follows from Lemma 2 of Devroye [10] that $Ef_n(x_0) \to f(x_0)$ as $n \to \infty$. Hence there exists a positive integer n_0 such that

$$Ef_n(x_0) < 2f(x_0) \quad \text{for all } n \geqq n_0. \tag{3.4.7}$$

For $x \geqq 0$ and $d > 0$ define an integer-valued function $\psi(x,d)$ as

$$\psi(x,d) = [\phi(g^{-1}(x)d^{-2})],$$

where [a] denotes the largest integer not greater than a. Set $C_1 = g^{-1}(1)$ (>0). It is clear that

$$\inf_{m \geq 1} \psi(m,d) > \phi(C_1 d^{-2}) - 1 \quad \text{and} \quad \inf_{m \geq 1} \phi(g^{-1}(m) d^{-2}) \geq \phi(C_1 d^{-2}).$$

Thus, since by (3.4.1) $\phi(C_1 d^{-2}) \to \infty$ as $d \to 0$ there exists a constant d_0 $(\varepsilon(0,1))$ not depending on m such that for all $m \geq 1$

$$\inf_{0<d<d_0} \phi(g^{-1}(m) d^{-2}) > 2 \tag{3.4.8}$$

and

$$\inf_{0<d<d_0} \psi(m,d) > n_0. \tag{3.4.9}$$

(3.4.7) and (3.4.9) imply that

$$\sup_{0<d<d_0} Ef_{\psi(m,d)}(x_0) \leq 2\, f(x_0) \quad \text{for all } m \geq 1. \tag{3.4.10}$$

It follows from the monotonicity of h_n and the definition of η that

$$3\,\eta(n) \geq \eta(n+2) \quad \text{for all } n \geq 1. \tag{3.4.11}$$

Thus by the use of (3.4.2),(3.4.8),(3.4.11), the definition of ψ and the monotonicity of η we obtain that for $d\varepsilon(0,d_0)$ and $m \geq 1$

$$\eta(\psi(m,d)) \geq \eta([\phi(g^{-1}(m) d^{-2}) - 1]) \geq \eta([\phi(g^{-1}(m) d^{-2}) - 1] + 2)/3$$

$$\geq \eta(\phi(g^{-1}(m) d^{-2}))/3 = g^{-1}(m) d^{-2}/3,$$

which yields that

$$\inf_{0<d<d_0} \eta(\psi(m,d)) d^2 \geq g^{-1}(m)/3 \quad \text{for all } m \geq 1. \tag{3.4.12}$$

By (3.4.5) there exists an integer $m_0 > 1$ such that

$$g^{-1}(m) > 9\, b^2 f(x_0) \quad \text{for all } m \geq m_0. \tag{3.4.13}$$

Combining (3.4.12) and (3.4.13) we get that

$$\inf_{0<d<d_0} b^{-2} \eta(\psi(m,d)) d^2 > 3\, f(x_0) \quad \text{for all } m \geq m_0. \tag{3.4.14}$$

For notational simplicity let $N = N(d)$. According to (3.4.2), the definition of ψ and the monotonicity of η

$$N \leqq \psi(m,d) \quad \text{implies} \quad \eta(N) \leqq g^{-1}(m)d^{-2}. \tag{3.4.15}$$

Thus, taking account of the monotonicity of g^{-1}, the definition of η, (3.1.1),(3.4.10),(3.4.14) and (3.4.15) we obtain that for $d\varepsilon(0,d_0)$ and $m \geqq m_0$

$$P\{g(Nh_N^p d^2) > m\} = P\{\eta(N) > g^{-1}(m)d^{-2}\} \tag{3.4.16}$$

$$\leqq P\{N > \psi(m,d)\} \leqq P\{f_{\psi(m,d)}(x_0) > b^{-2}\eta(\psi(m,d))d^2\}$$

$$\leqq P\{f_{\psi(m,d)}(x_0) - Ef_{\psi(m,d)}(x_0) > f(x_0)\}$$

$$\leqq P\{|f_{\psi(m,d)}(x_0) - Ef_{\psi(m,d)}(x_0)| > f(x_0)\}.$$

Set $Z_i = K_i(x_0,X_i) - EK_i(x_0,X_i)$ for $1 \leqq i \leqq n$, $s_n^2 = \sum_{i=1}^{n} \mathrm{Var}(Z_i)$ and $\|q\|_\infty = \sup\{|q(x)| : x\varepsilon R^p\}$ for a real-valued function q on R^p. By the boundedness of K and the monotonicity of h_n

$$|Z_i| \leqq C_2 h_n^{-p} \quad \text{a.s.} \quad \text{for } 1 \leqq i \leqq n \tag{3.4.17}$$

According to the boundedness of f and K and the monotonicity of h_n we have

$$s_n^2 \leqq \sum_{i=1}^{n} EK_i^2(x_0,X_i) = \sum_{i=1}^{n} h_i^{-p} \int K^2(u) f(x_0 - h_i u)\,du \tag{3.4.18}$$

$$\leqq \|K\|_\infty \, \|f\|_\infty \sum_{i=1}^{n} h_i^{-p} \leqq C_3 n h_n^{-p}.$$

Since $f_n(x_0) - Ef_n(x_0) = n^{-1}\sum_{i=1}^{n} Z_i$, by the use of (3.4.17),(3.4.18) and Lemma 3.4.1 we have that for $d\varepsilon(0,d_0)$ and $m \geqq m_0$

$$P\{|f_{\psi(m,d)}(x_0) - Ef_{\psi(m,d)}(x_0)| > f(x_0)\} \tag{3.4.19}$$

$$\leqq 2\exp\{-C_4\eta(\psi(m,d))\}.$$

It follows from (3.4.12) and $d_0 < 1$ that

$$\inf_{0<d<d_0} \eta(\psi(m,d)) \geqq g^{-1}(m)/3 \quad \text{for } m \geqq 1. \tag{3.4.20}$$

Thus, by virtue of (3.4.16),(3.4.19) and (3.4.20) we obtain that for all $m \geqq m_0$

$$\sup_{0<d<d_0} P\{g(N(d)h^p_{N(d)}d^2) > m\} \leqq 2\exp\{-Cg^{-1}(m)\}. \tag{3.4.21}$$

From (3.4.6) there exists an integer $m_1 > m_0$ such that

$$Cg^{-1}(m) > 2\log m \quad \text{for all } m \geqq m_1. \tag{3.4.22}$$

Combining (3.4.21) and (3.4.22) we have

$$\sum_{m=1}^{\infty} \sup_{0<d<d_0} P\{g(N(d)h^p_{N(d)}d^2) > m\} \leqq 2(m_1 - 1) + 2\sum_{m=m_1}^{\infty} m^{-2} < \infty,$$

which implies (3.4.4). Therefore the proof of Theorem 3.3.1 was completed.

4. SEQUENTIAL ESTIMATION FOR A MULTIPLE REGRESSION FUNCTION

4.1. STATEMENT OF THE PROBLEM

Let $Z=(X,Y)$, $Z_1=(X_1,Y_1)$, $\cdots$, $Z_n=(X_n,Y_n)$ be independent and identically distributed $R^p\times R$-valued random vectors on a probability space (Ω,F,P) with a unknown joint probability density function (p.d.f.) $f^*(x,y)$ with respect to the Lebesgue measure. There have been many papers on the estimation of the nonparametric regression function $m(x) = E[Y|X=x]$ (of Y on X) by

$$m_n(x) = \sum_{i=1}^{n} W_{ni}(x)Y_i, \tag{4.1.1}$$

where $W_{ni}(x) = W_{ni}(x,X_1,\cdots,X_n)$ for each i $(1 \leqq i \leqq n)$ is a suitable real-valued Borel measurable function of $x,X_1,\cdots,X_n$.

Nadaraya [22] and Watson [35] proposed the estimator (4.1.1) with p=1 and

$$W_{ni}(x) = K((x-X_i)/h_n) \Big/ \sum_{j=1}^{n} K((x-X_j)/h_n),$$

where $K(x)$ is a suitable kernel function and $\{h_n\}$ is a sequence of window-widths tending to zero. Later on, many authors have studied its asymptotic properties (see Prakasa Rao [25] for example).

When data increase we may be faced with computational burdens in processing them. To decrease these burdens Ahmad and Lin [1] proposed a recursive version of (4.1.1) with

$$W_{ni}(x) = h_i^{-p}K((x-X_i)/h_i) \Big/ \sum_{j=1}^{n} h_j^{-p}K((x-X_j)/h_j),$$

or equivalently,

$$m_0(x) = f_0(x) \equiv 0$$

$$f_n(x) = (h_n/h_{n-1})^p f_{n-1}(x) + K((x - X_n)/h_n) \tag{4.1.2}$$

$$m_n(x) = m_{n-1}(x) + f_n^{-1}(x)\{Y_n - m_{n-1}(x)\}K((x - X_n)/h_n)$$

and they proved some pointwise results for these estimators. Devroye and Wagner [11] considered a still simpler recursive estimator than (4.1.2). The author [16] proposed a class of recursive estimators $m_n(x)$ defined in Section 4.2 and used in this chapter, which contains (4.1.2) as a special case. Stone [32] has investigated the estimator (4.1.1) and discussed sufficient conditions on $\{W_{ni}(x)\}$ for $m_n(x)$ to be consistent.

On the other hand, when one uses a recursive estimator in practical situations one may be required to terminate the computations to obtain the estimator with given accuracy. In this case the sample size is a random variable. Suppose that N_t for each $t \in (0,\infty)$ is a stopping rule. Recently, Samanta [30] has shown the asymptotic normality of $m_{N_t}(x)$ by using the estimator (4.1.2).

In this chapter we propose a class of stopping rules $N = N(\alpha,d,x)$ based on the idea of Chow and Robbins [7], construct a sequence of 2d-width sequential confidence intervals $I_{N,d}(x) = [m_N(x) - d\ ,\ m_N(x) + d]$ for $m(x)$ and show that the probability $P\{m(x) \in I_{N,d}(x)\}$ converges to α as d tends to zero.

In Section 4.2 we shall make some preparations and give several lemmas. In Section 4.3 we shall prove the asymptotic consistency of the sequential confidence intervals $I_{N,d}(x)$.

4.2. PRELIMINARIES AND LEMMAS

In this section we shall make some preparations for Section 4.3.

Let $K(x)$ be a given bounded p.d.f. on R^p with respect to the Lebesgue measure satisfying $\|u\|_p^p K(u) \to 0$ as $\|u\|_p \to \infty$, where $\|\cdot\|_p$ denotes the Euclidean norm on R^p. We shall impose either of the following conditions on $K(x)$:

$$\int_{R^p} u_i K(u_1,\cdots,u_p)\,du_1\cdots du_p = 0 \quad \text{for all } i=1,\cdots,p \tag{K1}$$

and

$$\int_{R^p} \|u\|_p^2 K(u)\,du < \infty. \tag{K2}$$

Let $\{h_n\}$ be a nonincreasing sequence of positive numbers converging to zero, on which some of the following conditions are imposed:

$$nh_n^p \uparrow \infty \quad \text{as } n \to \infty; \tag{H1}$$

For some a $(0 < a \leqq 1)$ (H2)

$$n^{1-2a}h_n^p \to 0 \quad \text{as } n \to \infty,$$

$$n^{1-2a}h_n^p \sum_{j=1}^{n} j^{2(a-1)} h_j^{-p} \to \beta \quad \text{as } n \to \infty \quad \text{for some constant } \beta > 0,$$

$$n^{3/2-3a}h_n^{3p/2} \sum_{j=1}^{n} j^{3(a-1)} h_j^{-2p} \to 0 \quad \text{as } n \to \infty, \text{ and}$$

$$n^{1/2-a}h_n^{p/2} \sum_{j=1}^{n} j^{a-1} h_j^{2} \to 0 \quad \text{as } n \to \infty;$$

For any ε (>0) there exists a constant $\delta > 0$ such that

$$|n/m - 1| < \delta \quad \text{implies} \quad |h_n/h_m - 1| < \varepsilon\ ; \tag{H3}$$

$$\sum_{n=1}^{\infty} (n^2 h_n^p)^{-1} < \infty ; \tag{H4}$$

$$n^{1+\eta} h_n^{p+4} \to 0 \quad \text{as } n \to \infty \quad \text{for some constant } \eta > 0. \tag{H5}$$

Throughout this chapter we use the following class of recursive estimators $m_n(x)$ of the regression function $m(x)$, which is proposed by the author [16]:

$$m_0(x) \equiv 0, \ f_0(x) \equiv c \quad \text{for an arbitrary constant } c > 0$$

$$f_n(x) = f_{n-1}(x) + a_n\{K_n(x,X_n) - f_{n-1}(x)\}$$

$$m_n(x) = m_{n-1}(x) + a_n G(f_n(x))\{Y_n - m_{n-1}(x)\}K_n(x,X_n)$$

for each $n \geqq 1$, where

$$a_n = a/n \quad \text{with } 0 < a \leqq 1, \tag{4.2.1}$$

$$G(y) = \begin{cases} y^{-1} & \text{if } y > 0 \\ 0 & \text{otherwise,} \end{cases}$$

and

$$K_n(x,s) = h_n^{-p} K((x-s)/h_n) \quad \text{for } x, s \in R^p.$$

In this chapter, if in some term its denominator is less than or equal to zero we define the value of the term to be zero. Let

$$f(x) = \int_R f^*(x,y)\,dy, \ q(x) = \int_R y f^*(x,y)\,dy, \ g(x) = \int_R y^2 f^*(x,y)\,dy,$$

$$\psi(x) = \int_R y^4 f^*(x,y)\,dy \ \text{and} \ v(x) = (g(x)/f(x)) - (q(x)/f(x))^2 \ (\geqq 0).$$

Clearly, $v(x) > 0$ is equivalent to $f(x) > 0$ and $v(x) > 0$. Throughout this chapter we assume that $f(x)$, $q(x)$, $g(x)$ and $\psi(x)$ are all finite on R^p and write $m(x) = q(x)/f(x)$. Also, let $C, C_1, C_2, \cdots$ denote appropriate positive constants.

Define sequences $\{q_n(x)\}$, $\{g_n(x)\}$ and $\{v_n(x)\}$ as follows:

$$q_0(x) = g_0(x) \equiv 0,$$

$$q_n(x) = q_{n-1}(x) + a_n\{Q_n(x, Z_n) - q_{n-1}(x)\}$$

$$g_n(x) = g_{n-1}(x) + a_n\{G_n(x, Z_n) - g_{n-1}(x)\}$$

$$v_n(x) = (g_n(x)/f_n(x)) - (q_n(x)/f_n(x))^2$$

for each $n \geqq 1$, where for $x \varepsilon R^p$, $z = (u, y) \varepsilon R^p \times R$ and $n \geqq 1$

$$Q_n(x, z) = yK_n(x, u) \quad \text{and} \quad G_n(x, z) = y^2 K_n(x, u).$$

For a_n in (4.2.1) set

$$\gamma_0 = \gamma_1 = 1, \ \gamma_n = \prod_{j=2}^{n} (1 - a_j) \quad \text{for } n \geqq 2 \quad \text{and}$$

$$\beta_{mn} = \begin{cases} \prod_{j=m+1}^{n} (1 - a_j) & \text{if } n > m \geqq 0 \\ 1 \quad \text{if } n = m \geqq 0. \end{cases}$$

Obviously, $\gamma_n \downarrow 0$ as $n \to \infty$, $\gamma_n > 0$ for all $n \geqq 0$ and

$$\beta_{mn} = \gamma_m^{-1} \gamma_n \quad \text{if } n \geqq m \geqq 1. \tag{4.2.2}$$

It is known that

$$\beta_{mn} \sim m^a n^{-a} \quad \text{as } n \geqq m \to \infty \quad \text{and} \tag{4.2.3}$$

$$C_1 n^{-a} \leqq \gamma_n \leqq C_2 n^{-a} \quad \text{for all } n \geqq 1, \tag{4.2.4}$$

where "$\phi_n \sim \psi_n$ as $n \to \infty$" means $\lim_{n\to\infty} \phi_n/\psi_n = 1$.

REMARK 4.2.1. We can write $f_n(x)$, $q_n(x)$ and $g_n(x)$ as follows:

$$f_n(x) = \sum_{j=1}^{n} a_j \beta_{jn} K_j(x, X_j) + \beta_{0n} c,$$

$$q_n(x) = \sum_{j=1}^{n} a_j \beta_{jn} Q_j(x, Z_j) \quad \text{and} \quad g_n(x) = \sum_{j=1}^{n} a_j \beta_{jn} G_j(x, Z_j) ,$$

where $\sum_{j=m}^{n} (\cdot) = 0$ if $n < m$.

For any real-valued function θ on R^p let $C(\theta)$ be the set of all continuity points of θ and $||\theta||_\infty = \sup\{|\theta(x)| : x \varepsilon R^p\}$. For any fixed $x \varepsilon R^p$ and $n \geqq 1$ let

$$U_n^{(0)} = K_n(x, X_n) - EK_n(x, X_n) , \; U_n^{(1)} = Q_n(x, Z_n) - EQ_n(x, Z_n) ,$$

$$W_n = a_n \gamma_n^{-1} (U_n^{(0)} , U_n^{(1)})' , \; B_n = (nh_n^p)^{1/2} \gamma_n \sum_{j=1}^{n} W_j \quad \text{and}$$

$$B_n^* = (nh_n^p)^{1/2} (f_n(x) - f(x) , q_n(x) - q(x))' ,$$

where the prime denotes transpose.
We shall summarize some results obtained in [15] and [16].

LEMMA 4.2.1 ([16]). Let $\{d_n\}$ be a sequence of positive numbers converging to zero. Let $k(x)$ be a bounded, integrable, real-valued Borel measurable function on R^p satisfying $||u||_p^p |k(u)| \to 0$ as $||u||_p \to \infty$. Let $\theta(x)$ be an integrable, real-valued Borel measurable function on R^p. Then, for each point $x \varepsilon C(\theta)$ we have

$$\int_{R^p} d_n^{-p} k((x - u)/d_n) \theta(u) du \to \theta(x) \int_{R^p} k(u) du \quad \text{as } n \to \infty \quad \text{and}$$

$$\sup_{n \geqq 1} \int_{R^p} d_n^{-p} |k((x - u)/d_n)| |\theta(u)| du \leqq C ,$$

where C may depend on x.

LEMMA 4.2.2 ([16]). Assume $E[Y^2] < \infty$. Let (H4) be satisfied, and let x be a point with $f(x) > 0$, belonging to the set $C(f) \cap C(q) \cap C(g)$. Then

$$\lim_{n\to\infty} f_n(x) = f(x) \quad \text{a.s.}, \quad \lim_{n\to\infty} q_n(x) = q(x) \quad \text{a.s.} \quad \text{and}$$

$$\lim_{n\to\infty} m_n(x) = m(x) \quad \text{a.s.}$$

LEMMA 4.2.3 ([15]). Let $\{y_n\}$ be a sequence of random variables on a probability space $(\Omega, \mathcal{F}, P)$. Suppose that there exists a null set A such that each $\omega \varepsilon A^c$

$$y_n(\omega) \geqq 0 \quad \text{for all } n \geqq 1, \quad \lim_{n\to\infty} y_n(\omega) = 1 \quad \text{and}$$

$$y_n(\omega) > 0 \quad \text{for all } n \geqq m \quad \text{if } y_m(\omega) > 0 \text{ for some } m = m(\omega).$$

Let $\{b(n)\}$ be a sequence of positive numbers satisfying

$$\lim_{n\to\infty} b(n) = \infty \quad \text{and} \quad \lim_{n\to\infty} b(n)/b(n-1) = 1.$$

For each $t \varepsilon (0,\infty)$ define N_t as the smallest integer $n \geqq 1$ such that $b(n)/t \geqq y_n > 0$. Then

$$P\{N_t < +\infty\} = 1 \quad \text{for each } t \varepsilon (0,\infty), \quad P\{N_t \uparrow \infty \text{ as } t \to \infty\} = 1 \quad \text{and}$$

$$P\{b(N_t)/t \to 1 \text{ as } t \to \infty\} = 1 .$$

The following lemma gives the asymptotic normality of B_n.

LEMMA 4.2.4. Assume $E[Y^4] < \infty$. Let (H2) be satisfied. Consider a point $x \varepsilon C(f) \cap C(q) \cap C(g)$ with $v(x) > 0$. If either $x \varepsilon C(\psi)$ or $||\psi||_\infty < \infty$ holds, then

$$B_n \xrightarrow[L]{} N(0,\Gamma) \quad \text{as } n \to \infty \quad \text{(in law)},$$

where the covariance matrix $\Gamma = \Gamma(x)$ is given by

$$\Gamma = a^2\beta \int_{R^p} K^2(u)du \begin{pmatrix} f(x) & q(x) \\ q(x) & g(x) \end{pmatrix}.$$

PROOF. By the Cramér-Wold theorem (see Billingsley [4], page 49), it suffices to show that for $D' = (d_0, d_1)\varepsilon R^2$

$$D'B_n \xrightarrow[L]{} N(0, D'\Gamma D) \quad \text{as } n\to\infty. \tag{4.2.5}$$

We may assume $D \neq 0$. Let σ_n^2 be the variance of $D'B_n$. It holds that

$$D'B_n/\sigma_n \xrightarrow[L]{} N(0,1) \quad \text{as } n\to\infty \tag{4.2.6}$$

if we verify Lyapounov's condition

$$(nh_n^p)^{3/2}\gamma_n^3 \sum_{j=1}^n E|D'W_j|^3/\sigma_n^3 \to 0 \quad \text{as } n\to\infty. \tag{4.2.7}$$

It is easy to see that

$$\sigma_n^2 = nh_n^p\gamma_n^2 \sum_{j=1}^n a_j^2\gamma_j^{-2}\{\sum_{t=0}^1 d_t^2 E[(U_j^{(t)})^2] + 2d_0d_1E[U_j^{(0)}U_j^{(1)}]\}. \tag{4.2.8}$$

By Lemma 4.2.1 we get that as $n\to\infty$

$$h_n^pE[(U_n^{(0)})^2] \to f(x)\int K^2(u)du, \tag{4.2.9}$$

$$h_n^pE[(U_n^{(1)})^2] \to g(x)\int K^2(u)du$$

and

$$h_n^pE[U_n^{(0)}U_n^{(1)}] \to q(x)\int K^2(u)du,$$

where the domain of integral is R^p. According to (H2) and (4.2.4) we have that $\sum_{j=1}^n a_j^2\gamma_j^{-2}h_j^{-p} \uparrow \infty$ as $n\to\infty$, which, together with (4.2.9) and the Toeplitz lemma (see Loève [20], page 238), yields that

$$\left(\sum_{j=1}^{n} a_j^2\gamma_j^{-2}h_j^{-p}\right)^{-1} \sum_{j=1}^{n} a_j^2\gamma_j^{-2}E[(U_j^{(0)})^2] \to f(x)\int K^2(u)du \quad \text{as } n\to\infty.$$

We note that Lemma 1.2.1 can be proved under the condition (H2). Thus by making use of Lemma 1.2.1 and the above result we obtain

$$nh_n^p\gamma_n^2 \sum_{j=1}^{n} a_j^2\gamma_j^{-2}E[(U_j^{(0)})^2] \to a^2\beta f(x)\int K^2(u)du \quad \text{as } n\to\infty. \qquad (4.2.10)$$

By the same argument for (4.2.10) we have that as $n\to\infty$

$$nh_n^p\gamma_n^2 \sum_{j=1}^{n} a_j^2\gamma_j^{-2}E[(U_j^{(1)})^2] \to a^2\beta g(x)\int K^2(u)du \qquad (4.2.11)$$

and

$$nh_n^p\gamma_n^2 \sum_{j=1}^{n} a_j^2\gamma_j^{-2}E[U_j^{(0)}U_j^{(1)}] \to a^2\beta q(x)\int K^2(u)du\,. \qquad (4.2.12)$$

Combining (4.2.8), (4.2.10) $\sim$ (4.2.12) we get

$$\sigma_n^2 \to D'\Gamma D \quad \text{as } n\to\infty\,. \qquad (4.2.13)$$

As $v(x) > 0$ and $f(x) > 0$ we have

$$D'\Gamma D > 0\,. \qquad (4.2.14)$$

It can be easily shown that

$$E|D'W_j|^3 \leqq 4\left(\sum_{t=0}^{1}|d_t|^3\right)a_j^3\gamma_j^{-3}\max\{E|U_j^{(0)}|^3,\ E|U_j^{(1)}|^3\}. \qquad (4.2.15)$$

It follows from Lemma 4.2.1 that

$$E|U_j^{(0)}|^3 \leqq C_1h_j^{-2p} \quad \text{for all } j\geqq 1. \qquad (4.2.16)$$

By Hölder's inequality we get that for each $j\geqq 1$

$$E[|U_j^{(1)}|^3] \leqq 8E[|Q_j(x,Z_j)|^3]$$

$$= 8\int_{R^p\times R}\{|y|^3K_j^2(x,u)(f^*(u,y))^{3/4}\}\{K_j(x,u)(f^*(u,y))^{1/4}\}dudy$$

$$\leqq 8 [\int_{R^p \times R} y^4 \{K_j(x,u)\}^{8/3} f^*(u,y)\, du dy]^{3/4}$$

$$\times [\int_{R^p \times R} K_j^4(x,u) f^*(u,y)\, du dy]^{1/4} = 8h_j^{-2p} \times I_1 \times I_2 ,$$

where

$$I_1 = [\int_{R^p} h_j^{-p} \{K((x-u)/h_j)\}^{8/3} \psi(u)\, du]^{3/4} \quad \text{and}$$

$$I_2 = [\int_{R^p} h_j^{-p} K^4((x-u)/h_j) f(u)\, du]^{1/4} .$$

As $x \varepsilon C(f)$, it follows from Lemma 4.2.1 that $I_2 \leqq C_2$. If $x \varepsilon C(\psi)$ then by Lemma 4.2.1 we get $I_1 \leqq C_3$, and if $\|\psi\|_\infty < \infty$ then we have

$$I_1 \leqq \{ \|\psi\|_\infty (\|K\|_\infty)^{5/3} \}^{3/4} < \infty .$$

In any case, $I_1 \times I_2$ is bounded by a constant and therefore, we obtain

$$E[|U_j^{(1)}|^3] \leqq C_4 h_j^{-2p} \quad \text{for all } j \geqq 1. \tag{4.2.17}$$

Combining (4.2.15) $\sim$ (4.2.17) we have

$$E|D'W_j|^3 \leqq C_5 a_j^3 \gamma_j^{-3} h_j^{-2p} \quad \text{for all } j \geqq 1. \tag{4.2.18}$$

Since by (4.2.4) and (4.2.18)

$$(nh_n^p)^{3/2} \gamma_n^3 \sum_{j=1}^n E|D'W_j|^3 \leqq C_6 n^{3/2-3a} h_n^{3p/2} \sum_{j=1}^n j^{3(a-1)} h_j^{-2p} ,$$

it follows from (H2) that

$$(nh_n^p)^{3/2} \gamma_n^3 \sum_{j=1}^n E|D'W_j|^3 \to 0 \quad \text{as } n \to \infty ,$$

which, together with (4.2.13) and (4.2.14), implies (4.2.7). By virtue of (4.2.6) and (4.2.13) we obtain (4.2.5). Thus the lemma was proved.

LEMMA 4.2.5. Let $\{h_n\}$ be a nonincreasing sequence of positive numbers converging to zero and satisfy (H1) and (H2). Let $\{V_n\}$ be a sequence of independent random variables on a probability space $(\Omega, \mathcal{F}, P)$ satisfying

$$EV_n = 0 \quad \text{and} \quad h_n^p EV_n^2 \leqq C \quad \text{for all } n \geqq 1.$$

For any $d \in (0,\infty)$ let $N(d)$ be a positive integer-valued random variable on $(\Omega, \mathcal{F}, P)$ and $n(d)$ a positive integer with $\lim_{d\to 0} n(d) = \infty$. Set

$$T_n = \sum_{j=1}^{n} a_j \beta_{jn} V_j \quad \text{for each } n \geqq 1 ,$$

where a_n is given in (4.2.1). If

$$N(d)/n(d) \xrightarrow[P]{} 1 \quad \text{as } d \to 0 \text{ (in probability)} ,$$

then

$$(N(d) h_{N(d)}^p)^{1/2} (T_{N(d)} - T_{n(d)}) \xrightarrow[P]{} 0 \quad \text{as } d \to 0 .$$

PROOF. For simplicity let $N = N(d)$ and $n = n(d)$. Let any positive numbers ε and ξ be fixed. For any ρ (> 0) set

$$M_1 = [(1-\rho)n] \quad \text{and} \quad M_2 = [(1+\rho)n] ,$$

where $[b]$ denotes the largest integer not greater than b. Since as $d \to 0$

$$2\rho/(1-\rho) \to 0 \quad \text{and} \quad \{(1-\rho)/(1+\rho)\}^a \to 1 ,$$

there exists a positive constant $\rho = \rho(\varepsilon,\xi) < 1/2$ such that

$$2\rho/(1-\rho) < \varepsilon^2\xi/2 \quad \text{and} \quad (1 - \{(1-\rho)/(1+\rho)\}^a)^2 < \varepsilon^2\xi/2 . \tag{4.2.19}$$

As $\lim_{d\to 0} n = \infty$ we get $M_i \to \infty$ as $d \to 0$ for i=1,2 and $M_1/M_2 \sim (1-\rho)/(1+\rho)$

as $d \to 0$. Hence by (4.2.3) and (4.2.19) we have

$$(1 - \beta_{M_1 M_2})^2 < \varepsilon^2 \xi \quad \text{for } d \text{ sufficiently small.} \tag{4.2.20}$$

Also, since $(M_2 - M_1)/M_1 \sim 2\rho/(1-\rho)$ as $d \to 0$, it follows from (4.2.19) that

$$(M_2 - M_1)/M_1 < \varepsilon^2 \xi \quad \text{for } d \text{ sufficiently small.} \tag{4.2.21}$$

It is clear that

$$M_2/M_1 \leqq C_1 \quad \text{for all } d > 0\,. \tag{4.2.22}$$

Now, we shall prove the lemma. By assumption we get

$$P\{|N/n - 1| \geqq \rho\} < \xi \quad \text{for } d \text{ sufficiently small.} \tag{4.2.23}$$

Put

$$I = P\{(Nh_N^p)^{1/2} |T_N - T_n| \geqq \varepsilon\,, \quad |N - n| < \rho n\}\,.$$

Then

$$I \leqq P\{(ih_i^p)^{1/2} |T_i - T_n| \geqq \varepsilon \quad \text{for some } i \varepsilon (M_1, M_2]\} \tag{4.2.24}$$

$$\leqq P\{(ih_i^p)^{1/2} (|T_i - T_{M_1}| + |T_n - T_{M_1}|) \geqq \varepsilon \quad \text{for some } i \varepsilon (M_1, M_2]\}$$

$$\leqq P\{\max_{M_1 < i \leqq M_2} (ih_i^p)^{1/2} \times 2 \max_{M_1 < i \leqq M_2} |T_i - T_{M_1}| \geqq \varepsilon\}$$

$$= P\{\max_{M_1 < i \leqq M_2} (ih_i^p)^{1/2} \times \max_{M_1 < i \leqq M_2} |T_i - T_{M_1}| \geqq \varepsilon/2\}\,.$$

By (4.2.2) and the definition of T_n we have that for $i \varepsilon (M_1, M_2]$

$$|T_i - T_{M_1}| \leqq \Big|\sum_{j=1}^{M_1} a_j (\beta_{ji} - \beta_{jM_1}) V_j\Big| + \Big|\sum_{j=M_1+1}^{i} a_j \beta_{ji} V_i\Big|$$

$$\leqq (\gamma_{M_1} - \gamma_i) \Big|\sum_{j=1}^{M_1} a_j \gamma_j^{-1} V_j\Big| + \gamma_i \Big|\sum_{j=M_1+1}^{i} a_j \gamma_j^{-1} V_j\Big|\,,$$

which, together with (4.2.24) and the monotonicity of h_n, nh_n^p and γ_n, implies that

$$I \leqq P\{(M_2h_{M_1}^p)^{1/2}(\gamma_{M_1} - \gamma_{M_2})|\sum_{j=1}^{M_1} a_j\gamma_j^{-1}V_j| \tag{4.2.25}$$

$$+ (M_2h_{M_2}^p)^{1/2} \max_{M_1<i\leqq M_2} \gamma_i|\sum_{j=M_1+1}^{i} a_j\gamma_j^{-1}V_j| \geqq \varepsilon/2\} \leqq J_1 + J_2,$$

where

$$J_1 = P\{(M_2h_{M_1}^p)^{1/2}(\gamma_{M_1} - \gamma_{M_2})|\sum_{j=1}^{M_1} a_j\gamma_j^{-1}V_j| \geqq \varepsilon/4\} \quad \text{and}$$

$$J_2 = P\{(M_2h_{M_2}^p)^{1/2} \max_{M_1<i\leqq M_2} \gamma_i|\sum_{j=M_1+1}^{i} a_j\gamma_j^{-1}V_j| \geqq \varepsilon/4\} .$$

From (H2), (4.2.4) and assumption we get

$$ih_i^p \sum_{j=1}^{i} a_j\beta_{ji}^2 EV_j^2 \leqq C_2 i^{1-2a} h_i^p \sum_{j=1}^{i} j^{2(a-1)} h_j^{-p} \leqq C_3 \quad \text{for all } i \geqq 1 .$$

Hence by Chebychev's inequality, (4.2.2),(4.2.20),(4.2.22) and the above fact we obtain

$$J_1 \leqq C_4\varepsilon^{-2}(1 - \beta_{M_1M_2})^2(M_2/M_1)M_1h_{M_1}^p \sum_{j=1}^{M_1} a_j^2\beta_{jM_1}^2 EV_j^2 \tag{4.2.26}$$

$$\leqq C_5\xi \quad \text{for d sufficiently small.}$$

From assumption, the Hájek-Rényi inequality (see Petrov [24], page 51), the monotonicity of h_n, (4.2.21) and (4.2.22) we have that for d sufficiently small

$$J_2 \leqq C_6\varepsilon^{-2}M_2h_{M_2}^p \sum_{j=M_1+1}^{M_2} a_j^2 EV_j^2 \leqq C_7\varepsilon^{-2}M_2 \sum_{j=M_1+1}^{M_2} j^{-2} \tag{4.2.27}$$

$$\leqq C_7\varepsilon^{-2}M_2M_1^{-2}(M_2 - M_1) \leqq C_8\xi .$$

Combining (4.2.23) and (4.2.25) $\sim$ (4.2.27) we obtain

$$P\{(Nh_N^p)^{1/2}|T_N - T_n| \geqq \varepsilon\} \leqq I + P\{|N-n| \geqq \rho n\} \leqq C_9\xi$$

for d sufficiently small, which concludes the proof of the lemma.

4.3. FIXED-WIDTH SEQUENTIAL CONFIDENCE INTERVALS

In this section we shall propose sequential confidence intervals $I_{n,d}(x)$ for $m(x)$ with prescribed width 2d and coverage probability α. Then the asymptotic consistency of these confidence intervals will be shown, that is, $P\{m(x) \varepsilon I_{N(d),d}(x)\} \to \alpha$ as $d \to 0$, where $\{N(d)\}$ is the class of stopping rules defined below.

Let any α $(0 < \alpha < 1)$ be given. Define $D = D(\alpha) > 0$ by $\Phi(D) - \Phi(-D) = \alpha$, where Φ is the standard normal distribution function. Let d be any positive number, and let any $x \varepsilon R^p$ be fixed. We define the stopping rule $N(d) = N(\alpha,d,x)$ as the smallest integer $n \geqq 1$ such that

$$(D^2B)^{-1}d^2nh_n^p \geqq v_n(x)/f_n(x) > 0 ,$$

where

$$B = a^2\beta\int_{R^p} K^2(u)du \ (>0) \text{ with } \beta \text{ being given in (H2)} . \qquad (4.3.1)$$

Define the confidence interval $I_{n,d}(x)$ as

$$I_{n,d}(x) = [m_n(x) - d , m_n(x) + d] .$$

Also, define $n(d) = n(\alpha,d,x)$ as the smallest integer $n \geqq 1$ such that

$$(D^2B)^{-1}d^2nh_n^p \geqq v(x)/f(x) > 0 .$$

Let $\sigma^2(x) = Bv(x)/f(x)$ with B being given in (4.3.1).

The following lemma states the asymptotic properties of N(d).

LEMMA 4.3.1. Assume $E[Y^4] < \infty$. Let (H1), (H3) and (H4) be satisfied. Consider a point $x \in C(f) \cap C(q) \cap C(g)$ with $v(x) > 0$. If either $x \in C(\psi)$ or $\|\psi\|_\infty < \infty$ holds then we have

$$P\{N(d) < +\infty\} = 1 \quad \text{for each } d > 0 \text{ , } P\{N(d) \uparrow \infty \text{ as } d \to 0\} = 1 \text{ ,}$$

$$N(d) h^p_{N(d)} d^2/(D^2 \sigma^2(x)) \to 1 \quad \text{a.s.} \quad \text{as } d \to 0 \text{ , } \quad \text{and}$$

$$N(d) h^p_{N(d)}/(n(d) h^p_{n(d)}) \to 1 \quad \text{a.s.} \quad \text{as } d \to 0 \text{ .}$$

PROOF. From the definition of N(d) we get

$$N(d) = \text{smallest integer } n \geqq 1 \text{ such that } b(n)/t \geqq y_n > 0 \text{ ,} \qquad (4.3.2)$$

where

$$y_n = \frac{v_n(x)/f_n(x)}{v(x)/f(x)} \text{ , } b(n) = \frac{nh^p_n}{v(x)/f(x)} \text{ , and } t = D^2 B d^{-2} \text{ .}$$

Clearly,

$$b(n) > 0 \quad \text{for all } n \geqq 1 \text{ , } \lim_{n\to\infty} b(n) = \infty \quad \text{and} \qquad (4.3.3)$$

$$\lim_{n\to\infty} b(n)/b(n-1) = 1 \text{ .}$$

We shall show that

$$v_n(x) \geqq 0 \quad \text{for all } n \geqq 1 \quad \text{on } \Omega \qquad (4.3.4)$$

and that for any fixed $\omega \in \Omega$

$$v_n(x) > 0 \quad \text{for all } n \geqq m \quad \text{if } v_m(x) > 0 \text{ for some } m = m(\omega) \geqq 1. \quad (4.3.5)$$

For simplicity we omit ω. By the definition of $v_n(x)$ it suffices to consider the case where $f_n(x) > 0$.

That $v_n(x) \geqq$ (resp. $>$) 0 for all $n \geqq 1$ is equivalent to

$$g_n(x)f_n(x) - q_n^2(x) \geqq (\text{resp.} >)\ 0 \quad \text{for all } n \geqq 1. \tag{4.3.6}$$

Let any $\omega \varepsilon \Omega$ be fixed. First we shall prove (4.3.4). By the definitions of $f_n(x)$, $q_n(x)$ and $g_n(x)$ we have

$$A_{n+1} = (1-a_n)^2 A_n + a_n(1-a_n)D_n \quad \text{for each } n \geqq 1, \tag{4.3.7}$$

where

$$A_n = g_{n-1}(x)f_{n-1}(x) - q_{n-1}^2(x) \qquad \text{and}$$

$$D_n = G_n(x,Z_n)f_{n-1}(x) + K_n(x,X_n)g_{n-1}(x) - 2Q_n(x,Z_n)q_{n-1}(x).$$

By Remark 4.2.1 we get

$$D_n = \sum_{j=1}^{n-1} a_j \beta_{j\ n-1} \{G_n(x,Z_n)K_j(x,X_j) + K_n(x,X_n)G_j(x,Z_j) \tag{4.3.8}$$

$$- 2Q_n(x,Z_n)Q_j(x,Z_j)\} + G_n(x,Z_n)\beta_{0\ n-1}c$$

for each $n \geqq 1$. From the definitions of $K_n(x,X_n)$, $Q_n(x,Z_n)$ and $G_n(x,Z_n)$ we have

$$G_n(x,Z_n)\beta_{0\ n-1}c \geqq 0 \quad \text{for each } n \geqq 1 \qquad \text{and}$$

$$G_n(x,Z_n)K_j(x,X_j) + K_n(x,X_n)G_j(x,Z_j) - 2Q_n(x,Z_n)Q_j(x,Z_j) \geqq 0$$

for each $j = 1,\cdots,n-1$ with $n \geqq 2$, which, together with (4.3.8), yields that

$$D_n \geqq 0 \quad \text{for each } n \geqq 1. \tag{4.3.9}$$

It follows from (4.3.7) and (4.3.9) that

$$A_{n+1} \geqq (1-a_n)^2 A_n \quad \text{for each } n \geqq 1. \tag{4.3.10}$$

As $A_1 = 0$, by (4.3.10) and induction we have $A_n \geqq 0$ for each $n \geqq 1$, which, together with (4.3.6), gives (4.3.4). Next we shall prove (4.3.5). Assume that $v_m(x) > 0$ for some $m \geqq 1$. From (4.3.6) we get $A_{m+1} > 0$. Suppose that $A_n > 0$ for some $n > m+1$. Then by (4.2.1) and

(4.3.10) we have $A_{n+1} > 0$. Hence by induction we obtain $A_n > 0$ for all $n \geqq m+1$, which is equivalent to $v_n(x) > 0$ for all $n \geqq m$. Thus (4.3.5) was proved. Next we shall show that

$$g_n(x) \to g(x) \quad \text{a.s.} \quad \text{as } n \to \infty . \tag{4.3.11}$$

Since $EG_n(x,Z_n) \to g(x)$ as $n \to \infty$ by Lemma 4.2.1, the definition of $G_n(x,Z_n)$ and $x \varepsilon C(g)$, it follows from Remark 4.2.1 and the Toeplitz lemma that

$$Eg_n(x,Z_n) \to g(x) \quad \text{as } n \to \infty . \tag{4.3.12}$$

Lemma 4.2.1 and (H4) give

$$\sum_{n=1}^{\infty} a_n^2 EG_n^2(x,Z_n) < \infty ,$$

which, together with Kolmogorov's convergence theorem, Kronecker's lemma, Remark 4.2.1 and (4.2.2), implies that

$$g_n(x) - Eg_n(x) \to 0 \quad \text{a.s.} \quad \text{as } n \to \infty . \tag{4.3.13}$$

Thus, according to (4.3.12) and (4.3.13) we obtain (4.3.11).

Lemma 4.2.2 and (4.3.11) give

$$v_n(x) \to v(x) \quad \text{a.s.} \quad \text{as } n \to \infty . \tag{4.3.14}$$

From Lemma 4.2.2, (4.3.4), (4.3.5), (4.3.14) and the property of $f_n(x)$, there exists a null set A such that for each $\omega \varepsilon A^c$

$$y_n(\omega) \geqq 0 \quad \text{for all } n \geqq 1 , \quad \lim_{n\to\infty} y_n(\omega) = 1 \quad \text{and}$$

$$y_n(\omega) > 0 \quad \text{for all } n \geqq m \quad \text{if } y_m(\omega) > 0 \quad \text{for some } m=m(\omega) \geqq 1 ,$$

which, together with (4.3.2) and (4.3.3), permits us to apply Lemma 4.2.3 to obtain the first three assertions of Lemma 4.3.1. Replacing $N(d)$, $v_n(x)$ and $f_n(x)$ by $n(d)$, $v(x)$ and $f(x)$, respectively, we have that as $d \to 0$ $n(d) \to \infty$ and $n(d)h_{n(d)}^p d^2/(D^2\sigma^2(x)) \to 1$, which, together with the third assertion, implies the last assertion. This completes the proof.

REMARK 4.3.1. By the use of Lemmas 4.2.2 and 4.3.1, and Theorem 1 of Richter [26] we have that under all the conditions of Lemma 4.3.1

$$m_{N(d)}(x) \to m(x) \quad \text{a.s.} \quad \text{as } d \to 0 .$$

The following theorem is one of main theorems.

THEOREM 4.3.2. Assume $E[Y^4] < \infty$. Let (K1),(K2) and (H1) $\sim$ (H5) be satisfied. Suppose that there exist bounded, continuous second partial derivatives $\partial^2 f(x)/\partial x_i \partial x_j$ and $\partial^2 q(x)/\partial x_i \partial x_j$ on R^p for $i,j=1,\cdots,p$. Consider a point $x \varepsilon C(g)$ with $v(x) > 0$. Assume $x \varepsilon C(\psi)$ or $\|\psi\|_\infty < \infty$. If

$$N(d)/n(d) \xrightarrow[P]{} 1 \quad \text{as } d \to 0 \tag{4.3.15}$$

then we obtain

$$(N(d)h_{N(d)}^p)^{1/2}(m_{N(d)}(x) - m(x)) \xrightarrow[L]{} N(0,\sigma^2(x)) \quad \text{as } d \to 0 .$$

PROOF. For simplicity put $N = N(d)$ and $n = n(d)$. It follows from Lemma 4.3.1 that $n \to \infty$ as $d \to 0$. First we shall show that

$$B_N \xrightarrow[L]{} N(0,\Gamma) \quad \text{as } d \to 0 , \tag{4.3.16}$$

where Γ is given in Lemma 4.2.4. From Lemma 4.2.4 and the Cramér-

Wold theorem we get

$$D'B_n \xrightarrow[L]{} N(0,D'\Gamma D) \quad \text{as } d\to 0 \quad \text{for any } D'\varepsilon R^2 . \tag{4.3.17}$$

Since $D'B_N = D'B_n + (D'B_N - D'B_n)$ for any $D'\varepsilon R^2$, in order to prove (4.3.16) it suffices from (4.3.17) and the Cramér-Wold theorem to show that

$$D'B_N - D'B_n \xrightarrow[P]{} 0 \quad \text{as } d\to 0 \quad \text{for any } D'\varepsilon R^2 . \tag{4.3.18}$$

Let any $D' = (d_0, d_1)\varepsilon R^2$ be fixed. For $i \geqq 1$ set

$$S_i^{(t)} = \sum_{j=1}^{i} a_j \beta_{ji} U_j^{(t)} \quad \text{for } t=0,1.$$

It is clear that

$$D'B_N - D'B_n \tag{4.3.19}$$

$$= \sum_{t=0}^{1} d_t (Nh_N^p)^{1/2} (S_N^{(t)} - S_n^{(t)}) + \{(Nh_N^p/(nh_n^p))^{1/2} - 1\}D'B_n .$$

Put

$$\xi^{(t)}(x) = (f(x))^{1-t}(g(x))^t \quad \text{for } t=0,1 .$$

It follows from assumption and Lemma 4.2.1 that for $t=0,1$

$$h_i^p E[(U_i^{(t)})^2] \leqq \int_{R^p} h_i^{-p} K^2((x-u)/h_i)\xi^{(t)}(u)du \leqq C_1 \quad \text{for all } i \geqq 1 .$$

Thus by the use of Lemma 4.2.5 we have

$$\sum_{t=0}^{1} d_t (Nh_N^p)^{1/2} (S_N^{(t)} - S_n^{(t)}) \xrightarrow[P]{} 0 \quad \text{as } d\to 0 . \tag{4.3.20}$$

Let any ε (> 0) be fixed. From (4.3.15) and (H3) we get that for δ in (H3)

$$P\{|h_N/h_n - 1| \geqq \varepsilon\} \leqq P\{|N/n - 1| \geqq \delta\} \to 0 \quad \text{as } d\to 0 ,$$

which implies that $h_N/h_n \xrightarrow[P]{} 1$ as $d\to 0$. Therefore, (4.3.15) gives

$$\{Nh_N^p/(nh_n^p)\}^{1/2} - 1 \xrightarrow[P]{} 0 \quad \text{as } d \to 0 ,$$

which, together with (4.3.17), yields that

$$\{(Nh_N^p/(nh_n^p))^{1/2} - 1\}D'B_n \xrightarrow[P]{} 0 \quad \text{as } d \to 0 . \qquad (4.3.21)$$

From (4.3.19) ∿ (4.3.21) we obtain (4.3.18). Set

$$d_{j1} = EK_j(x,X_j) - f(x) , \quad d_{j2} = EQ_j(x,Z_j) - q(x) ,$$

$$D_j' = (d_{j1}, d_{j2}) \quad \text{for } j \geqq 1 \quad \text{and} \quad D_0' = (c - f(x), - q(x)) .$$

Then

$$B_i^* - B_i = (ih_i^p)^{1/2}\gamma_i \sum_{j=1}^{i} a_j\gamma_j^{-1}D_j + (ih_i^p)^{1/2}\beta_{0i}D_0 \quad \text{on } \Omega . \qquad (4.3.22)$$

It follows from (4.2.4) and (H2) that $(ih_i^p)^{1/2}\beta_{0i} \to 0$ as $i \to \infty$. Thus, if we show that for each $t = 1,2$

$$(ih_i^p)^{1/2}\gamma_i \sum_{j=1}^{i} a_j\gamma_j^{-1}|d_{jt}| \to 0 \quad \text{as } i \to \infty , \qquad (4.3.23)$$

then it follows from (4.3.22) that

$$||B_i^* - B_i||_2 \to 0 \quad \text{as } i \to \infty \quad \text{on } \Omega. \qquad (4.3.24)$$

We shall show (4.3.23). By the Taylor theorem, (K1),(K2) and the boundedness of $\partial^2 f(x)/\partial x_i \partial x_j$ and $\partial^2 q(x)/\partial x_i \partial x_j$, we get that for each $t = 1,2$ $|d_{jt}| \leqq C_2 h_j^2$ for all $j \geqq 1$, which, together with (4.2.4), yields that for each $t = 1,2$

$$(ih_i^p)^{1/2}\gamma_i \sum_{j=1}^{i} a_j\gamma_j^{-1}|d_{jt}| \leqq C_3 i^{1/2-a} h_i^{p/2} \sum_{j=1}^{i} j^{a-1}h_j^2 .$$

Hence by (H2) we obtain (4.3.23). By Lemma 4.3.1 and (4.3.24) we get that $||B_N^* - B_N||_2 \to 0$ a.s. as $d \to 0$, which, together with (4.3.16), implies that

$$B_N^* \xrightarrow[L]{} N(0,\Gamma) \quad \text{as } d \to 0 . \qquad (4.3.25)$$

Define a function $T(u,v)$ on R^2 as

$$T(u,v) = \begin{cases} v/u & \text{if } u \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Let $L' = (-q(x)/f^2(x)\ ,\ f^{-1}(x))$. By the Taylor theorem we get

$$(Nh_N^p)^{1/2}\{T(f_N(x)\ ,\ q_N(x)) - T(f(x)\ ,\ q(x))\} \tag{4.3.26}$$

$$= L'B_N^* + \varepsilon_N \|B_N^*\|_2 \quad \text{on } [N < +\infty]\ ,$$

where

$$\varepsilon_i \to 0 \quad \text{if } \|(f_i(x)\ ,\ q_i(x))' - (f(x)\ ,\ q(x))'\|_2 \to 0\ .$$

Suppose $N < \infty$. Define N_1 = smallest integer $i \geqq 1$ such that $f_i(x) > 0$. It is clear that $N_1 \leqq N$. From the nonnegativeness of $K(y)$ we can easily get

$$f_i(x) > 0 \quad \text{for } i \geqq N_1 \tag{4.3.27}$$

and

$$K_i(x,X_i) = 0 \quad \text{for } N_1 > i \geqq 1\ . \tag{4.3.28}$$

It follows from (4.3.27) and the definition of N_1 that

$$f_i(x)m_i(x) = \sum_{j=N_1}^{i} a_j\beta_{ji}Q_j(x,Z_j) \quad \text{for } i \geqq N_1\ . \tag{4.3.29}$$

Since by (4.3.28) $Q_j(x,Z_j) = 0$ for $N_1 > j \geqq 1$, using Remark 4.2.1 and (4.3.29) we obtain $f_i(x)m_i(x) = q_i(x)$ for $i \geqq N_1$. Thus by (4.3.27) we get $m_i(x) = q_i(x)/f_i(x)$ for $i \geqq N_1$, which yields $m_N(x) = q_N(x)/f_N(x)$ on $[N < +\infty]$. Hence by the definitions of $T(u,v)$ and N we obtain

$$T(f_N(x)\ ,\ q_N(x)) = m_N(x) \quad \text{on } [N < +\infty]\ . \tag{4.3.30}$$

Since $m(x) = q(x)/f(x)$ it follows from (4.3.26) and (4.3.30) that

$$(Nh_N^p)^{1/2}(m_N(x) - m(x)) = L'B_N^* + \varepsilon_N \|B_N^*\|_2 \quad \text{on } [N < +\infty]\ . \tag{4.3.31}$$

Lemma 4.2.2 gives that $\varepsilon_i \to 0$ a.s. as $i \to \infty$, which, together with Lemma 4.3.1, yields that

$$\varepsilon_N \to 0 \quad \text{a.s.} \quad \text{as } d \to 0 . \tag{4.3.32}$$

Combining (4.3.25) and (4.3.32) we get

$$\varepsilon_N \, \|B_N^*\|_2 \xrightarrow[P]{} 0 \quad \text{as } d \to 0 . \tag{4.3.33}$$

Therefore, according to (4.3.25),(4.3.31),(4.3.33) and Lemma 4.3.1 we obtain

$$(Nh_N^p)^{1/2}(m_N(x) - m(x)) \xrightarrow[L]{} N(0, L'\Gamma L) \quad \text{as } d \to 0 ,$$

which concludes the proof of Theorem 4.3.2.

We are now in the position to give our main result.

THEOREM 4.3.3. Under all the conditions of Theorem 4.3.2 we have

$$P\{m(x) \varepsilon I_{N(d),d}(x)\} \to \alpha \quad \text{as } d \to 0 .$$

PROOF. Put $N = N(d)$. By Lemma 4.3.1 and Theorem 4.3.2 we have

$$Dd^{-1}(m_N(x) - m(x))$$

$$= (D^2\sigma^2(x)/(Nh_N^p d^2))^{1/2}(Nh_N^p/\sigma^2(x))^{1/2}(m_N(x) - m(x))$$

$$\xrightarrow[L]{} N(0,1) \quad \text{as } d \to 0 .$$

Thus we obtain

$$P\{m(x) \varepsilon I_{N,d}(x)\}$$

$$= P\{|Dd^{-1}(m_N(x) - m(x))| \leqq D\} \to \Phi(D) - \Phi(-D) = \alpha \text{ as } d \to 0 .$$

This completes the proof.

COROLLARY 4.3.4. Assume $E[Y^4] < \infty$. Let (K1) and (K2) be satisfied, and let $\|\psi\|_\infty < \infty$. Suppose that there exist bounded, continuous second partial derivatives $\partial^2 f(x)/\partial x_i \partial x_j$ and $\partial^2 q(x)/\partial x_i \partial x_j$ on R^p for $i,j = 1,\cdots,p$ and that $g(x)$ is continuous on R^p. Set

$$h_n = n^{-r/p} \quad \text{with } p/(p+4) < r < 1 .$$

Let a in (4.2.1) satisfy $1 \geq a > (1-r)/2$. Then, for each point x with $v(x) > 0$ we obtain

$$P\{m(x) \varepsilon I_{N(d),d}(x)\} \to \alpha \quad \text{as } d \to 0 .$$

PROOF. We can easily verify (H1) $\sim$ (H5) with $\beta = (2a + r - 1)^{-1}$. Lemma 4.3.1 gives (4.3.15). Thus, since all the conditions of Theorem 4.3.2 are fulfilled, we obtain Corollary 4.3.4 by Theorem 4.3.3. This completes the proof.

ACKNOWLEDGEMENTS

The author would like to thank Professor M. Akahira, University of Tsukuba, who has helped me prepare this thesis. The author would also like to thank Professor K. Tanaka, Niigata University, for his hearty encouragements.

REFERENCES

[1] Ahmad, I. A. and Lin, P. (1976). Nonparametric sequential estimation of a multiple regression function. Bull. Math. Statist. 17, 63-75.

[2] Bennett, G. (1962). Probability inequalities for the sum of independent random variables. J. Amer. Statist. Assoc. 57, 33-45.

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

[4] Billingsley, P. (1968). Convergence of Probability Measures. John Wiley and Sons, Inc., New York.

[5] Cacoullos, T. (1966). Estimation of a multivariate density. Ann. Inst. Statist. Math. 18, 179-189.

[6] Carroll, R. J. (1976). On sequential density estimation. Z. Wahrsch. Verw. Gebiete 36, 137-151.

[7] Chow, Y. S. and Robbins, H. (1965). On the asymptotic theory of fixed-width sequential confidence intervals for the mean. Ann. Math. Statist. 36, 457-462.

[8] Chung, K. L. (1974). A Course in Probability Theory. 2nd ed. Academic Press.

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

[10] Devroye, L. P. (1979). On the pointwise and the integral convergence of recursive kernel estimates of probability densities. Utilitas Mathematica 15, 113-128.

[11] Devroye, L. P. and Wagner, T. J. (1980). On the L_1 convergence of kernel estimators of regression functions with applications in discrimination. Z. Wahrsch. Verw. Gebiete 51, 15-25.

[12] Devroye, L. and Györfi, L. (1985). Nonparametric Density Estimation: The L_1 View. John Wiley and Sons, Inc., New York.

[13] Geertsema, J. C. (1970). Sequential confidence intervals based on rank tests. Ann. Math. Statist. 41, 1016-1026.

[14] Isogai, E. (1980). Strong consistency and optimality of a sequential density estimator. Bull. Math. Statist. 19, 55-69.

[15] Isogai, E. (1981). Stopping rules for sequential density estimation. Bull. Math. Statist. 19, 53-67.

[16] Isogai, E. (1983). A class of nonparametric recursive estimators of a multiple regression function. Bull. Inform. Cybernetics 20, 33-44.

[17] Isogai, E. (1986). Asymptotic consistency of fixed-width sequential confidence intervals for a multiple regression function. Ann. Inst. Statist. Math. 38, 69-83.

[18] Isogai, E. (1987). The convergence rate of fixed-width sequential confidence intervals for a probability density function. Sequential Analysis 6, 55-69.

[19] Isogai, E. (1988). A note on sequential density estimation. Sequential Analysis 7, 11-21.

[20] Loève, M. (1963). Probability Theory. 3rd ed. D. Van Nostrand, Princeton.

[21] Michel, R. and Pfanzagl, J. (1971). The accuracy of the normal approximation for minimum contrast estimates. Z. Wahrsch. Verw. Gebiete 18, 73-84.

[22] Nadaraya, E. A. (1964). On estimating regression. Theor. Prob. Appl. 9, 141-142.

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

[24] Petrov, V. V. (1975). Sums of Independent Random Variables. Springer-Verlag.

[25] Prakasa Rao, B. L. S. (1983). Nonparametric Functional Estimation. Academic Press.

[26] Richter, W. (1965). Limit theorems for sequences of random variables with sequences of random indeces. Theor. Prob. Appl. 10, 74-84.

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

[28] Rychlik, Z. (1978). The order of approximation in the random central limit theorem. Lect. Notes in Math. No. 656. Springer-Verlag.

[29] Sacks, J. (1958). Asymptotic distribution of stochastic approximation procedures. Ann. Math. Statist. 29, 373-405.

[30] Samanta, M. (1984). On sequential estimation of the regression function. Bull. Inform. Cybernetics 21, 19-27.

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

[32] Stone, C. J. (1977). Consistent nonparametric regression. Ann. Statist. 5, 595-645.

[33] Stute, W. (1983). Sequential fixed-width confidence intervals for a nonparametric density function. Z. Wahrsch. Verw. Gebiete 62, 113-123.

[34] Watanabe, M. (1974). On convergences of asymptotically optimal discriminant functions for pattern classification problems. Bull. Math. Statist. 16, 23-34.

[35] Watson, G. S. (1964). Smooth regression analysis. Sankhyā Ser. A, 26, 359-372.

[36] Wegman, E. J. and Davies, H. I. (1979). Remarks on some recursive estimators of a probability density. Ann. Statist. 7, 316-327.

[37] Wolverton, C. T. and Wagner, T. J. (1969). Asymptotically optimal discriminant functions for pattern classification. IEEE Trans. Inform. Theory IT-15, 258-265.

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