

## STRONG CONSISTENCY AND OPTIMALITY OF A SEQUENTIAL DENSITY ESTIMATOR

Isogai, Eiichi

Department of Mathematics, Faculty of Science, Niigata University

<https://doi.org/10.5109/13140>

---

出版情報：統計数理研究. 19 (1/2), pp.55-69, 1980-03. Research Association of Statistical Sciences

バージョン：

権利関係：



# STRONG CONSISTENCY AND OPTIMALITY OF A SEQUENTIAL DENSITY ESTIMATOR

By

Eiichi ISOGAI\*

(Received December 25, 1979)

## 1. Introduction.

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 are several literatures on the problem of sequentially estimating the p.d.f.  $f(x)$ . Yamato [9] introduced the sequential estimator defined by

$$\hat{f}_n(x) = \left(1 - \frac{1}{n}\right) \hat{f}_{n-1}(x) + \frac{1}{n h_n^p} K\left(\frac{x - X_n}{h_n}\right).$$

Carroll [2] and Davies [4] consider this estimator  $\hat{f}_n$ . The following sequential estimator  $f_n^*$  was introduced by Wegman and Davies [8];

$$f_n^*(x) = \frac{n-1}{n} \left(\frac{h_{n-1}}{h_n}\right)^{1/2} f_{n-1}^*(x) + \frac{1}{n h_n} K\left(\frac{x - X_n}{h_n}\right),$$

where  $x \in R^1$ .

In this paper we shall propose a sequential estimator  $f_n(x)$  in a modified form of  $\hat{f}_n(x)$ , that is,  $\frac{1}{n}$  is replaced by  $\frac{a}{n}$  with  $\frac{1}{2} < a \leq 1$ . The sequential estimator  $f_n(x)$  defined in section 2 and the sequential estimator  $\hat{f}_n(x)$  can be rewritten, respectively, as

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

and

$$\hat{f}_n(x) = \sum_{m=1}^n \frac{1}{n} \frac{1}{h_m^p} K\left(\frac{x - X_m}{h_m}\right).$$

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

In this paper we shall consider the problem of estimating  $f(x)$  at a given point  $x$ . Yamato [9] showed the weak consistency of  $\hat{f}_n(x)$  and its asymptotic normality. Davies [4] showed the strong consistency of  $\hat{f}_n(x)$ . We shall show the strong con-

---

\* Department of Mathematics, Faculty of Science, Niigata University, Niigata, Japan.

sistency 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 sequence  $\{h_n\}$ .

In section 2 we shall define the sequential estimator  $f_n(x)$  in a certain form. Also, auxiliary results needed later will be given. In section 3 the estimator  $f_n(x)$  will be shown to be strongly consistent and its asymptotic normality will also be shown. Furthermore the square convergence rate of  $f_n(x)$  will be given. Section 4 is devoted to comparison between  $f_n(x)$  and  $\hat{f}_n(x)$  under the criterion of the limit of  $\text{Var}(f_n(x))/\text{Var}(\hat{f}_n(x))$ . 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. Thus the estimator  $f_n(x)$  is better than the estimator  $\hat{f}_n(x)$  under the above criterion when the special type of the sequence  $\{h_n\}$  is chosen. Furthermore an optimal choice of the coefficient  $a$  in the sequential estimator  $f_n(x)$  will be given for the special type of  $\{h_n\}$ .

## 2. Preliminaries and auxiliary results.

In this section we shall give a sequential 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

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

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

and

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

where  $\|y\| = (\sum_{i=1}^p y_i^2)^{1/2}$  for  $y = (y_1, \dots, y_p)$  and the domain of integral is  $R^p$  unless otherwise specified. Let  $\{a_n\}$  be a sequence of positive numbers defined by

$$(A) \quad a_n = \frac{a}{n} \quad \text{with} \quad \frac{1}{2} < a \leq 1 \quad \text{for all } n \geq 1.$$

Let  $\{h_n\}$  be a sequence of positive numbers. On this sequence we shall impose some of the following conditions:

$$(H1) \quad \lim_{n \rightarrow \infty} h_n = 0,$$

$$(H2) \quad h_{n_0} \geq h_{n_0+1} \geq \dots \quad \text{for some } n_0 \geq 1,$$

$$(H3) \quad \lim_{n \rightarrow \infty} n h_n^p = \infty,$$

$$(H4) \quad \lim_{n \rightarrow \infty} (n h_n^p)^{-1/2} \log n = 0,$$

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

Define  $K_n$  by

$$K_n(x, y) = \frac{1}{h_n^p} K\left(\frac{x-y}{h_n}\right) \quad \text{for all } x, y \in R^p, n=1, 2, 3, \dots$$

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

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

$$f_n(x) = (1-a_n)f_{n-1}(x) + a_n K_n(x, X_n) \quad \text{for all } x \in R^p, n=1, 2, 3, \dots,$$

where  $X_1, X_2, X_3, \dots$  is a sequence of independent identically distributed  $p$ -dimensional random vectors with the common (unknown) p.d.f.  $f$ , and the conditions (K1), (K2), (K3), (A) and (H1) are assumed to be satisfied. In what follows, for the estimator  $f_n$  we shall assume the conditions (K1)~(K3), (A) and (H1) without restating them repeatedly. Throughout this paper  $C_1, C_2, \dots$  denote 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  $K(x) \geq 0$  for all  $x \in R^p$ , then it is easy to see that  $f_n(x)$  ( $n=1, 2, \dots$ ) are actually probability density functions. If we put  $a=1$  in (A) then the estimator  $f_n(x)$  coincides the estimator  $\hat{f}_n(x)$ .

Now, we introduce some notations. Let

$$\beta_{mn} = \prod_{k=m+1}^n (1-a_k) \quad \text{if } 0 \leq m < n$$

$$= 1 \quad \text{if } m=n \geq 0,$$

$$\gamma_0 = \gamma_1 = 1$$

and

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

It is clear that  $\gamma_n > 0$  for all  $n \geq 1$ ,  $\beta_{mn} \leq 1$  for all  $n \geq m \geq 0$  and  $\beta_{mn} = \gamma_n \gamma_m^{-1}$  for all  $n \geq m \geq 1$ .

Sacks [6] showed that

$$(2.0.1) \quad (1-\varepsilon'_m)m^a n^{-a} \leq \beta_{mn} \leq (1+\varepsilon'_m)m^a n^{-a}$$

for all  $n \geq m \geq 1$ , where  $\varepsilon'_m \rightarrow 0$  as  $m \rightarrow \infty$ .

By (2.0.1) and  $1-a_j > 0$  for all  $j \geq 2$ , there exist two positive constants  $C_1$  and  $C_2$  such that

$$(2.0.2) \quad C_1 m^a n^{-a} \leq \beta_{mn} \leq C_2 m^a n^{-a} \quad \text{for all } n \geq m \geq 1.$$

In particular, it holds that

$$(2.0.3) \quad C_1 n^{-a} \leq \gamma_n \leq C_2 n^{-a} \quad \text{for all } n \geq 1.$$

LEMMA 2.1. Let  $\{h_n\}$  satisfy (H1) and (H2). If for some  $a > \frac{1}{2}$  there exists a positive constant  $\beta$  such that

$$(2.1.1) \quad n^{1-2a} h_n^p \sum_{m=1}^n m^{2(a-1)} h_m^{-p} \longrightarrow \beta \quad \text{as } n \longrightarrow \infty,$$

then, for any positive integer  $m_0$ ,

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

where " $\varphi_n \sim \psi_n$  as  $n \rightarrow \infty$ " means that  $\varphi_n/\psi_n \rightarrow 1$  as  $n \rightarrow \infty$ .

PROOF. It suffices to show that

$$(2.1.2) \quad \beta^{-1} n h_n^p \sum_{m=m_0}^n m^{-2} \beta_{mn}^2 h_m^{-p} \rightarrow 1 \quad \text{as } n \rightarrow \infty.$$

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

$$(2.1.3) \quad \left(1 + \frac{2\varepsilon}{3}\right)(1 + \xi) < 1 + \varepsilon \quad \text{and} \quad \left(1 - \frac{\varepsilon}{3}\right)(1 - \xi) > 1 - \varepsilon.$$

By (2.0.1) 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

$$(2.1.4) \quad \left(1 - \frac{\varepsilon}{3}\right) m^{2a} n^{-2a} \leq \beta_{mn}^2 \leq \left(1 + \frac{\varepsilon}{3}\right) m^{2a} n^{-2a}$$

for all  $n \geq m \geq m_1$ . By the monotonicity of  $h_n$  and  $a > \frac{1}{2}$  we get

$$(2.1.5) \quad \sum_{m=1}^n m^{2(a-1)} h_m^{-p} \rightarrow \infty \quad \text{as } n \rightarrow \infty.$$

From (H1),  $a > \frac{1}{2}$ , (2.1.1) and (2.1.5) there exists a positive integer  $m_2 \geq m_1$  such that

$$(2.1.6) \quad 1 - \xi < \beta^{-1} n^{1-2a} h_n^p \sum_{m=m_1}^n m^{2(a-1)} h_m^{-p} < 1 + \xi$$

and

$$(2.1.7) \quad 0 \leq C_2^2 \sum_{m=m_0}^{m_1-1} m^{-2} \gamma_m^{-2} h_m^{-p} / \sum_{m=m_1}^n m^{2(a-1)} h_m^{-p} < \frac{\varepsilon}{3}$$

for all  $n \geq m_2$ . Combining (2.0.3), (2.1.3), (2.1.4), (2.1.6) and (2.1.7) we obtain

$$(2.1.8) \quad \begin{aligned} & \beta^{-1} n h_n^p \sum_{m=m_0}^n m^{-2} \beta_{mn}^2 h_m^{-p} \\ & \leq \beta^{-1} n^{1-2a} h_n^p \sum_{m=m_1}^n m^{2(a-1)} h_m^{-p} \left\{ \left( C_2^2 \sum_{m=m_0}^{m_1-1} m^{-2} \gamma_m^{-2} h_m^{-p} / \sum_{m=m_1}^n m^{2(a-1)} h_m^{-p} \right) + 1 + \frac{\varepsilon}{3} \right\} \\ & < (1 + \xi) \left( 1 + \frac{2\varepsilon}{3} \right) < 1 + \varepsilon \quad \text{for all } n \geq m_2. \end{aligned}$$

In the same manner as above we have

$$(2.1.9) \quad \beta^{-1} n h_n^p \sum_{m=m_0}^n m^{-2} \beta_{mn}^2 h_m^{-p} > 1 - \varepsilon \quad \text{for all } n \geq m_2.$$

Thus the combined use of (2.1.8) and (2.1.9) yields (2.1.2), which completes the proof.

The next lemma is one of the results given by Sacks [6].

LEMMA 2.2. Let  $q > -1$ . Then, for any positive integer  $m_0$ ,

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

The following two propositions can be found in Watanabe [7].

PROPOSITION 2.3. *Let  $\{A_n\}$  be a sequence of non-negative numbers. Suppose that there exist two sequences of non-negative numbers  $\{b_n\}$  and  $\{d_n\}$  such that*

$$(2.3.1) \quad A_{n+1} \leq (1-b_{n+1})A_n + b_{n+1}d_{n+1} \quad \text{for all } n \geq 1,$$

$$(2.3.2) \quad \sum_{n=1}^{\infty} b_n = \infty \quad \text{and} \quad \lim_{n \rightarrow \infty} b_n = 0,$$

and

$$(2.3.3) \quad \lim_{n \rightarrow \infty} d_n = 0.$$

Then we have  $\lim_{n \rightarrow \infty} A_n = 0$ .

PROPOSITION 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  $\sigma$ -fields,  $\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{F}$  for all  $n \geq 1$ , where  $U_n$  and  $V_n$  are measurable with respect to  $\mathcal{F}_n$  for each  $n \geq 1$ . And let  $\{b_n\}$  be a sequence of real numbers. Suppose that the following conditions are satisfied:*

$$(2.4.1) \quad 0 \leq U_n \text{ a.s. for all } n \geq 1,$$

$$(2.4.2) \quad E[U_1] < \infty,$$

$$(2.4.3) \quad E[U_{n+1} | \mathcal{F}_n] \leq (1-b_{n+1})U_n + V_n \text{ a.s. for all } n \geq 1,$$

$$(2.4.4) \quad \sum_{n=1}^{\infty} E|V_n| < \infty$$

and

$$(2.4.5) \quad 0 \leq b_n \leq 1 (n=1, 2, \dots), \lim_{n \rightarrow \infty} b_n = 0 \quad \text{and} \quad \sum_{n=1}^{\infty} b_n = \infty,$$

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

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

Cacoullos [1] has given the following proposition.

PROPOSITION 2.5. *Let  $K(y)$  be a real-valued Borel measurable function on  $R^p$  satisfying (K1), (K2) and (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*

$$g_n(x) = \frac{1}{h_n^p} \int K\left(\frac{y}{h_n}\right) g(x-y) dy,$$

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

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

DEFINITION 2.6. A bounded real-valued function  $g$  defined on  $R^p$  is said to be locally Lipschitz of order  $\lambda$ ,  $0 < \lambda \leq 1$ , at  $x_0$  (abbreviated as loc. Lip.  $\lambda$  at  $x_0$ ) if there exist two positive constants  $L$  and  $\eta$  depending on  $x_0$  and  $\lambda$  such that  $\|y\| < \eta$  implies  $|g(x_0 + y) - g(x_0)| \leq L\|y\|^\lambda$ , where  $\|\cdot\|$  denotes the Euclidean norm.

### 3. Strong consistency and asymptotic normality.

In this section we shall show the strong consistency of the sequential estimator  $f_n$  defined by (F) of section 2 and we shall discuss the asymptotic normality of the estimator.

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

THEOREM 3.1. Let  $\{h_n\}$  satisfy (H1) and (H5). Then, for each  $x \in C(f)$ ,

$$(3.1.1) \quad \lim_{n \rightarrow \infty} f_n(x) = f(x) \text{ a. s.}$$

and

$$(3.1.2) \quad \lim_{n \rightarrow \infty} E(f_n(x) - f(x))^2 = 0.$$

PROOF. From the algorithm (F) it follows that

$$(3.1.3) \quad |Ef_n(x) - f(x)| \\ \leq (1 - a_n)|Ef_{n-1}(x) - f(x)| + a_n|EK_n(x, X_n) - f(x)|$$

for all  $n \geq 1$ . By Proposition 2.5 we get

$$(3.1.4) \quad \lim_{n \rightarrow \infty} |EK_n(x, X_n) - f(x)| = 0.$$

Thus, making use of (3.1.3), (3.1.4) and Proposition 2.3, we have

$$(3.1.5) \quad \lim_{n \rightarrow \infty} |Ef_n(x) - f(x)| = 0.$$

Now, by the algorithm (F) and the independence of  $\{X_n\}$  we obtain

$$(3.1.6) \quad E[(f_n(x) - Ef_n(x))^2 | X_1, \dots, X_{n-1}] \\ \leq (1 - a_n)(f_{n-1}(x) - Ef_{n-1}(x))^2 + a_n^2 \text{Var}[K_n(x, X_n)] \\ \leq (1 - a_n)(f_{n-1}(x) - Ef_{n-1}(x))^2 + a_n^2 E[K_n^2(x, X_n)] \text{ a. s.}$$

for all  $n \geq 2$ , where  $\text{Var}(X)$  denotes the variance of  $X$ . By Proposition 2.5 we have

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

which yields by the use of (K1) and (K2) that

$$(3.1.7) \quad h_n^2 E[K_n^2(x, X_n)] \leq C_3 \quad \text{for all } n \geq 1.$$

Hence, combining (3.1.6) and (3.1.7) we obtain

$$(3.1.8) \quad E[(f_n(x) - Ef_n(x))^2 | X_1, \dots, X_{n-1}] \\ \leq (1 - a_n)(f_{n-1}(x) - Ef_{n-1}(x))^2 + C_3 a_n^2 h_n^{-p} \text{ a. s.}$$

for all  $n \geq 2$ . From (H5) it follows that

$$(3.1.9) \quad \sum_{n=1}^{\infty} a_n^2 h_n^{-p} < \infty.$$

By (K2) we have

$$(3.1.10) \quad E(f_1(x) - Ef_1(x))^2 < \infty.$$

Thus, from (A), (3.1.8), (3.1.9), (3.1.10) and Proposition 2.4 we get

$$(3.1.11) \quad \lim_{n \rightarrow \infty} (f_n(x) - Ef_n(x))^2 = 0 \text{ a. s.}$$

and

$$(3.1.12) \quad \lim_{n \rightarrow \infty} E(f_n(x) - Ef_n(x))^2 = 0.$$

On the other hand

$$(3.1.13) \quad |f_n(x) - f(x)| \leq |f_n(x) - Ef_n(x)| + |Ef_n(x) - f(x)|$$

and

$$(3.1.14) \quad E(f_n(x) - f(x))^2 = E(f_n(x) - Ef_n(x))^2 + (Ef_n(x) - f(x))^2$$

for all  $n \geq 1$ . Hence we get (3.1.1) from (3.1.5), (3.1.11) and (3.1.13), and get (3.1.2) from (3.1.5), (3.1.12) and (3.1.14), which complete the proof.

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

**THEOREM 3.2.** *Let  $\{h_n\}$  satisfy (H1) and (H2).*

*Assume the following condition:*

(3.2.1) *For some  $a$  in (A) with  $\frac{1}{2} < a \leq 1$  there exists a positive constant  $\beta$  such that*

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

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

$$(3.2.2) \quad \lim_{n \rightarrow \infty} n h_n^p \text{Var}(f_n(x)) = Bf(x),$$

*where*

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

**PROOF.** Let

$$(3.2.3) \quad Z_m = K_m(x, X_m) - EK_m(x, X_m) \quad \text{for all } m \geq 1.$$

From the algorithm (F) we get

$$(3.2.4) \quad f_n(x) - Ef_n(x) = \sum_{m=1}^n a_m \beta_{m,n} Z_m \quad \text{for all } n \geq 1.$$

It follows from (H1) and Proposition 2.5 that

$$(3.2.5) \quad \lim_{n \rightarrow \infty} h_n^p E Z_n^2 = f(x) \int K^2(y) dy.$$

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 2.1 we obtain

$$(3.2.6) \quad b_n n h_n^p \gamma_n^2 \sim a^2 \beta (>0) \quad \text{as } n \longrightarrow \infty.$$

On the other hand, by (2.0.3) we have

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

which yields, together with  $a > \frac{1}{2}$  and (H1), that

$$(3.2.7) \quad \lim_{n \rightarrow \infty} n h_n^p \gamma_n^2 = 0.$$

Thus it follows from (3.2.6) and (3.2.7) that

$$(3.2.8) \quad b_n \uparrow \infty \quad \text{as } n \longrightarrow \infty.$$

By (3.2.5), (3.2.8) and the Toeplitz lemma (see Loève [5]) we have

$$(3.2.9) \quad b_n^{-1} \sum_{m=1}^n a_m^2 \gamma_m^{-2} E Z_m^2 \longrightarrow f(x) \int K^2(y) dy \quad \text{as } n \longrightarrow \infty.$$

From (3.2.4) we get

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

which yields

$$(3.2.10) \quad n h_n^p \text{Var}(f_n(x)) = (b_n n h_n^p \gamma_n^2) b_n^{-1} \sum_{m=1}^n a_m^2 \gamma_m^{-2} E Z_m^2.$$

Combining (3.2.6), (3.2.9) and (3.2.10) we obtain

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

which concludes the theorem.

LEMMA 3.3. Suppose that in addition to (K1)~(K3),  $K(y)$  satisfies the condition

$$(K4) \quad \int \|y\| |K(y)| dy < \infty.$$

Then, for each  $x$  at which  $f$  is loc. Lip.  $\lambda$ , there exists a positive constant  $C = C(x, \lambda)$  depending on  $x$  and  $\lambda$  such that

$$(3.3.1) \quad |EK_n(x, X_n) - f(x)| \leq C h_n^\lambda \quad \text{for all } n \geq 1.$$

PROOF. Let

$$(3.3.2) \quad \delta_n = EK_n(x, X_n) - f(x) \quad \text{for all } n \geq 1$$

and

$$(3.3.3) \quad \|f\|_\infty = \sup_{y \in \mathbb{R}^p} f(y).$$

The boundedness of the p.d.f.  $f$  implies  $\|f\|_\infty < \infty$ . By (K1) and Definition 2.6 we have

$$\begin{aligned}
 (3.3.4) \quad |\delta_n| &\leq \int |K(y)| |f(x - h_n y) - f(x)| dy \\
 &\leq C_3(x, \lambda) \int_{\{h_n \|y\| < \gamma\}} \|y\|^\lambda |K(y)| dy h_n^\lambda + 2\|f\|_\infty \int_{\{h_n \|y\| \geq \gamma\}} |K(y)| dy \\
 &\leq C_3(x, \lambda) \int \|y\|^\lambda |K(y)| dy h_n^\lambda + 2\|f\|_\infty \gamma^{-\lambda} h_n^\lambda \int \|y\|^\lambda |K(y)| dy \\
 &= (C_3(x, \lambda) + 2\|f\|_\infty \gamma^{-\lambda}) \int \|y\|^\lambda |K(y)| dy h_n^\lambda
 \end{aligned}$$

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

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

If  $0 < \lambda < 1$ , then it follows from (K1), (K4) and the Hölder inequality that

$$\begin{aligned}
 &\int \|y\|^\lambda |K(y)| dy \\
 &\leq \left[ \int \|y\| |K(y)| dy \right]^\lambda \left[ \int |K(y)| dy \right]^{1-\lambda} < \infty.
 \end{aligned}$$

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

we get

$$(3.3.5) \quad C(x, \lambda) < \infty.$$

Finally we get (3.3.1) from (3.3.4) and (3.3.5). The proof is complete.

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

**THEOREM 3.4.** *Let  $x$  be a point such that  $f(x) > 0$  and  $f$  is loc. Lip.  $\lambda$  at  $x$ . Let  $K(y)$  be given in Lemma 3.3. Suppose that  $\{h_n\}$  satisfies (H1), (H2), (H3) and (H5). In addition to (3.2.1), assume the following condition:*

$$(3.4.1) \quad \sum_{m=1}^n m^{a-1} h_m^\lambda = O(n^a h_n^\lambda) \quad \text{as } n \rightarrow \infty,$$

where  $a$  is given in (3.2.1).

Then there exists a positive constant  $C = C(x, \lambda)$  depending on  $x$  and  $\lambda$  such that

$$(3.4.2) \quad E(f_n(x) - f(x))^2 \leq C b_n \quad \text{for all } n \geq 1,$$

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

**PROOF.** By Theorem 3.2 there exists a positive constant  $C_3 = C_3(x)$  such that

$$(3.4.3) \quad \text{Var}(f_n(x)) \leq C_3 (nh_n^p)^{-1} \quad \text{for all } n \geq 1.$$

Let  $\delta_m$  be given in (3.3.2). From (2.0.2), (3.4.1) and Lemma 3.3 we get

$$\begin{aligned} \left| \sum_{m=1}^n a_m \beta_{mn} \delta_m \right| &\leq C_4(x, \lambda) n^{-a} \sum_{m=1}^n m^{a-1} h_m^\lambda \\ &\leq C_5(x, \lambda) h_n^\lambda \quad \text{for all } n \geq 1. \end{aligned}$$

Hence we have

$$(3.4.4) \quad \left( \sum_{m=1}^n a_m \beta_{mn} \delta_m \right)^2 \leq C_6(x, \lambda) h_n^{2\lambda} \quad \text{for all } n \geq 1.$$

Since  $|f_0(x) - f(x)| \leq \|K\|_\infty + \|f\|_\infty < \infty$ , where  $\|f\|_\infty$  is defined by (3.3.3), it follows from (2.0.3) that

$$(3.4.5) \quad \beta_{0n}^2 (f_0(x) - f(x))^2 \leq C_7 n^{-2a} \quad \text{for all } n \geq 1.$$

Since  $Ef_n(x) - f(x) = \beta_{0n}(f_0(x) - f(x)) + \sum_{m=1}^n a_m \beta_{mn} \delta_m$ , by (3.4.4) and (3.4.5) we have

$$(3.4.6) \quad (Ef_n(x) - f(x))^2 \leq C_8(x, \lambda) b_n \quad \text{for all } n \geq 1.$$

Let  $C = C(x, \lambda) = C_3 + C_8(x, \lambda) < \infty$ . Hence by (3.1.14), (3.4.3) and (3.4.6) we obtain (3.4.2). This completes the proof.

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

LEMMA 3.5. *Let  $Z_m$  be given in (3.2.3). Suppose that  $\{h_n\}$  satisfies (H1), (H2) and (H4). Assume the following conditions:*

(3.5.1) *For some  $a$  in (A) with  $\frac{2}{3} \leq a \leq 1$  there exists a positive constant  $\beta$  such that*

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

and

$$(3.5.2) \quad \|f\|_\infty = \sup_{y \in R^p} f(y) < \infty.$$

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

$$(3.5.3) \quad (n h_n^p)^{1/2} \sum_{m=1}^n a_m \beta_{mn} Z_m \xrightarrow{\mathcal{L}} N(0, Bf(x)) \quad \text{as } n \longrightarrow \infty,$$

where  $B$  is given in Theorem 3.2,  $N(0, \sigma^2)$  stands for the normal random variable with mean 0 and variance  $\sigma^2$  and " $\xrightarrow{\mathcal{L}}$ " means convergence in law.

PROOF. Let  $U_n = a_n \gamma_n^{-1} Z_n$ ,  $S_n = \sum_{m=1}^n U_m$  and  $s_n^2 = \text{Var}(S_n)$  for all  $n \geq 1$ . Then it holds that  $s_n^2 = \sum_{m=1}^n a_m^2 \gamma_m^{-2} E Z_m^2$ .

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

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

it follows from (3.2.6) and (3.2.9) that

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

which implies that

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

Thus (3.5.3) holds.

In what follows we consider the case when  $f(x) > 0$ . It follows from (3.2.6) and (3.2.9) that

$$(3.5.4) \quad s_n^2 \sim Bf(x)(nh_n^p \gamma_n^2)^{-1} \quad \text{as } n \longrightarrow \infty.$$

If the Lyapounov condition

$$(3.5.5) \quad s_n^{-3} \sum_{m=1}^n E|U_m|^3 \longrightarrow 0 \quad \text{as } n \longrightarrow \infty$$

holds, then it follows that

$$(3.5.6) \quad s_n^{-1} S_n \xrightarrow{\mathcal{L}} N(0, 1) \quad \text{as } n \longrightarrow \infty.$$

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

$$(3.5.7) \quad E|U_m|^3 \leq a_m^3 \gamma_m^{-3} E|Z_m|^3 \leq 8a_m^3 \gamma_m^{-3} E[|K_m(x, X_m)|^3].$$

Since

$$\begin{aligned} E[|K_m(x, X_m)|^3] &= h_m^{-2p} \int |K(y)|^3 f(x - h_m y) dy \\ &\leq \|K\|_\infty^3 \|f\|_\infty \int |K(y)| dy h_m^{-2p} \quad \text{for all } m \geq 1, \end{aligned}$$

it follows from (K1), (K2) and (3.5.2) that

$$(3.5.8) \quad E[|K_m(x, X_m)|^3] \leq C_3 h_m^{-2p} \quad \text{for all } m \geq 1,$$

where

$$C_3 = \|K\|_\infty^3 \|f\|_\infty \int |K(y)| dy < \infty.$$

From (3.5.7) and (3.5.8) we obtain

$$(3.5.9) \quad E|U_m|^3 \leq C_4 m^{-3} \gamma_m^{-3} h_m^{-2p} \quad \text{for all } m \geq 1.$$

Let  $n_1 = n_0 + 1$  with  $n_0$  being given in (H2). Thus, using (2.0.3), (3.5.4), (3.5.9) and (H2) we have

$$\begin{aligned} (3.5.10) \quad s_n^{-3} \sum_{m=n_1}^n E|U_m|^3 &\leq C_4 \gamma_n^3 (nh_n^p)^{3/2} \sum_{m=n_1}^n m^{-3} \gamma_m^{-3} h_m^{-2p} \\ &\leq C_6 n^{2-3a} (nh_n^p)^{-1/2} \sum_{m=n_1}^n m^{3(a-1)} \left(\frac{h_n}{h_m}\right)^{2p} \end{aligned}$$

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

It is easy to see that (H4) implies (H3).

We now consider the following two cases for the value of  $a$ .

Case (i).  $\frac{2}{3} < a \leq 1$ .

It follows from Lemma 2.2 that

$$(3.5.11) \quad \sum_{m=n_1}^n m^{3(a-1)} \sim (3a-2)^{-1} n^{3a-2} \quad \text{as } n \rightarrow \infty.$$

By making use of (H3), (3.5.10) and (3.5.11) we obtain

$$(3.5.12) \quad s_n^{-3} \sum_{m=n_1}^n E|U_m|^3 \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Case (ii).  $a = \frac{2}{3}$ .

Since  $\sum_{m=n_1}^n m^{-1} < \log n$ , we have (3.5.12) from (H4) and (3.5.10). Hence (3.5.12) holds for  $\frac{2}{3} \leq a \leq 1$ . From (2.0.3) 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  $a \geq \frac{2}{3}$  and (H1), that

$$(3.5.13) \quad \lim_{n \rightarrow \infty} nh_n^p \gamma_n^2 = 0.$$

By the use of  $Bf(x) > 0$ , (3.5.4) and (3.5.13) we get

$$(3.5.14) \quad \lim_{n \rightarrow \infty} s_n^2 = \infty,$$

which implies that

$$(3.5.15) \quad s_n^{-3} \sum_{m=1}^{n_1-1} E|U_m|^3 \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Combining (3.5.12) and (3.5.15) we obtain (3.5.5). 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 (3.5.3) from (3.5.4) and (3.5.6). Therefore the proof is complete.

The following theorem gives the asymptotic normality of  $f_n$ .

**THEOREM 3.6.** *Let  $x$  be a point at which  $f$  is loc. Lip.  $\lambda$ . Suppose that  $\{h_n\}$  satisfies (H1), (H2), (H4), (3.5.1) and (3.4.1) with  $a$  being given in (3.5.1). Furthermore assume the condition*

$$(3.6.1) \quad \lim_{n \rightarrow \infty} nh_n^{2\lambda+p} = 0.$$

*Then it holds that*

$$(3.6.2) \quad (nh_n^p)^{1/2} (f_n(x) - f(x)) \xrightarrow{\mathcal{L}} N(0, Bf(x)) \quad \text{as } n \rightarrow \infty,$$

*where  $B$  is given in Theorem 3.2.*

PROOF. Let  $Z_m$  and  $\delta_m$  be given in (3.2.3) and (3.3.2), respectively. Then we can rewrite  $f_n(x) - f(x)$  as

$$(3.6.3) \quad f_n(x) - f(x) = \beta_{0n}(f_0(x) - f(x)) + \sum_{m=1}^n a_m \beta_{mn} Z_m + \sum_{m=1}^n a_m \beta_{mn} \delta_m$$

for all  $n \geq 1$ . Definition 2.6 implies (3.5.2). Thus by virtue of Lemma 3.5 (3.5.3) holds. If it holds that

$$(3.6.4) \quad (n h_n^p)^{1/2} \beta_{0n}(f_0(x) - f(x)) \longrightarrow 0 \quad \text{as } n \longrightarrow \infty$$

and

$$(3.6.5) \quad (n h_n^p)^{1/2} \sum_{m=1}^n a_m \beta_{mn} \delta_m \longrightarrow 0 \quad \text{as } n \longrightarrow \infty,$$

then by the use of (3.5.3), (3.6.3), (3.6.4), (3.6.5) and Corollary of Chung [3] (page 93) we can obtain (3.6.2).

We next show (3.6.4) and (3.6.5). By (2.0.3), (K2) and (3.5.2) we get

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

Hence (H1) together with  $a \geq \frac{2}{3}$  implies (3.6.4). In view of (2.0.2), (3.4.1) and Lemma 3.3 we have

$$(3.6.6) \quad \begin{aligned} (n h_n^p)^{1/2} \left| \sum_{m=1}^n a_m \beta_{mn} \delta_m \right| \\ \leq C_4(x, \lambda) (n h_n^p)^{1/2} n^{-a} \sum_{m=1}^n m^{a-1} h_m^\lambda \\ \leq C_5(x, \lambda) (n h_n^{2\lambda+p})^{1/2}, \end{aligned}$$

where  $C_5(x, \lambda)$  is a positive constant depending on  $x$  and  $\lambda$ . From (3.6.1) and (3.6.6) we get (3.6.5). This completes the proof.

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

EXAMPLE.

Let  $h_n = n^{-r/p}$  for  $n \geq 1$ . If  $\frac{1}{2} < a \leq 1$  and  $0 < r < \min\left(\frac{ap}{\lambda}, 1\right)$ , then  $\{h_n\}$  satisfies all of (H1)~(H5), (3.2.1) with  $\beta = (2a + r - 1)^{-1}$  and (3.4.1). If  $\frac{1}{2} < a \leq 1$  and  $\frac{p}{2\lambda + p} < r < \min\left(\frac{ap}{\lambda}, 1\right)$ , then  $\{h_n\}$  satisfies (3.6.1) in addition to (H1)~(H5), (3.2.1) with  $\beta = (2a + r - 1)^{-1}$  and (3.4.1).

#### 4. Asymptotic rate of variances and optimal coefficient.

##### 4.1. Asymptotic rate of variances.

Let the estimators  $\hat{f}_n$  and  $f_n$  with the kernel  $K$  satisfying (K1), (K2) and (K3) be given in sections 1 and 2, respectively. In this section we shall discuss the asymptotic rate of variances between  $f_n$  and  $\hat{f}_n$ , and we shall give an optimal choice, in a certain sense, of the coefficient  $a$  in (A).

**THEOREM 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  $\frac{1}{2} < a \leq 1$  there exist two positive constants  $\alpha$  with  $\alpha \leq 1$  and  $\beta$  such that*

$$(4.1.1) \quad n^{1-2a} h_n^p \sum_{m=1}^n m^{2(a-1)} h_m^{-p} \longrightarrow \beta \quad \text{as } n \longrightarrow \infty$$

and

$$(4.1.2) \quad n^{-1} h_n^p \sum_{m=1}^n h_m^{-p} \longrightarrow \alpha \quad \text{as } n \longrightarrow \infty.$$

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

$$(4.1.3) \quad \lim_{n \rightarrow \infty} \text{Var}(f_n(x)) / \text{Var}(\hat{f}_n(x)) = a^2 \beta / \alpha.$$

**PROOF.** From Theorem 3 of Yamato [9] it follows that

$$(4.1.4) \quad \lim_{n \rightarrow \infty} n h_n^p \text{Var}(\hat{f}_n(x)) = \alpha f(x) \int K^2(y) dy (> 0),$$

which together with (3.2.2) implies (4.1.3). This completes the proof.

**COROLLARY 4.2.** *Let  $\{h_n\}$  be given by*

$$(4.2.1) \quad h_n = n^{-r/p} \quad \text{where } 0 < r < 1.$$

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

$$(4.2.2) \quad \lim_{n \rightarrow \infty} \text{Var}(f_n(x)) / \text{Var}(\hat{f}_n(x)) = a^2(1+r)/(2a+r-1),$$

where  $\frac{1}{2} < a \leq 1$ .

*Furthermore, if  $0 < r < \frac{1}{2}$  and  $a = 1 - r$  then*

$$(4.2.3) \quad \lim_{n \rightarrow \infty} \text{Var}(f_n(x)) / \text{Var}(\hat{f}_n(x)) = 1 - r^2.$$

**PROOF.** As is seen in Example of section 3, all conditions of Theorem 4.1 are satisfied by taking  $\alpha = (1+r)^{-1}$  and  $\beta = (2a+r-1)^{-1}$ , where  $\frac{1}{2} < a \leq 1$ . Hence (4.2.2) is a direct consequence of (4.1.3).

#### 4.2. Optimal choice of $a$ .

Let  $\{h_n\}$  be given in (4.2.1). It is easy to see that  $a^2(1+r)/(2a+r-1)$  in the right hand side of (4.2.2) achieves its minimum at  $a = 1 - r$  as a function of  $a$  for each fixed  $r$ . Thus, for each  $0 < r < \frac{1}{2}$ ,  $a = 1 - r$  gives the minimum value of the asymptotic rate of variances,  $\lim_{n \rightarrow \infty} \text{Var}(f_n(x)) / \text{Var}(\hat{f}_n(x))$ . For each  $\frac{1}{2} \leq r < 1$ , the closer the coefficient  $a$  is to  $\frac{1}{2}$ , 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 3.2 and (4.1.4) we have

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

Thus the closer  $r$  is to 0, the better it is from the viewpoint of the speed of convergence of variance. If  $\frac{2}{3} \leq a \leq 1$  and  $p(2\lambda + p)^{-1} < r < \min(ap\lambda^{-1}, 1)$  in (4.2.1), then the sequence  $\{h_n\}$  satisfies all conditions in the previous sections. Hence, in this case,  $a = \frac{2}{3}$  is the best and the closer  $r$  is to  $\min(ap\lambda^{-1}, 1)$ , the better it is from the viewpoint of the asymptotic rate of variances.

### Acknowledgment.

This work was done while the author was at Kyushu University on leave of absence from Niigata University. The author would like to express his hearty thanks to Professor N. Furukawa for his encouragement and critical readings of the original manuscript.

### References

- [1] CACOULOS, T., *Estimation of a multivariate density*. Ann. Inst. Statist. Math., 18 (1966), 179-189.
- [2] CARROLL, R.J., *On sequential density estimation*. Z. Wahrscheinlichkeitstheorie und verw. Gebiete., 36 (1976), 137-151.
- [3] CHUNG, K.L., *A course in probability theory*. 2nd Edition. Academic Press, 1974.
- [4] DAVIES, H.I., *Strong consistency of a sequential estimator of a probability density function*. Bull. Math. Statist., 15, No. 3~4 (1973), 49-54.
- [5] LOÈVE, M., *Probability theory*. 3rd Edition. D. Van Nostrand, Princeton, 1963.
- [6] SACKS, J., *Asymptotic distribution of stochastic approximation procedures*. Ann. Math. Statist., 29 (1958), 373-405.
- [7] WATANABE, M., *On convergences of asymptotically optimal discriminant functions for pattern classification problems*. Bull. Math. Statist., 16, No. 1~2 (1974), 23-34.
- [8] WEGMAN, E.J. and DAVIES, H.I., *Remarks on some recursive estimators of a probability density*. Ann. Statist., 7 (1979), 316-327.
- [9] YAMATO, H., *Sequential estimation of a continuous probability density function and mode*. Bull. Math. Statist., 14, No. 3~4 (1971), 1-12.