数理解析研究所講究録
第 623 巻 1987 年 26-42

Asymptotic properties of sequential density estimation

新潟大・理　磯貝英一　(Eiichi Isogai)

§1. はじめに

$f(x)$ は R^p 上のルベーグ測度に関するノンパラメトリックな確率密度関数で未知であるとし，$\{X_n, n \geqq 1\}$ はある確率空間 $(\Omega, \mathfrak{F}, P)$ 上で定義された独立で同一分布に従う確率ベクトル列で，共通の確率密度関数 $f(x)$ をもつとする。

$f(x)$ の推定問題は古くから研究されている。推定量の1つとして次のような Kernel を用いた Parzen - Rosenblatt Kernel estimator が Rosenblatt [11] と Parzen [9] によって提案された。

$$(1.1) \qquad g_n(x) = \frac{1}{n} \sum_{i=1}^{n} h_n^{-p} K\left(\frac{x - X_i}{h_n}\right)$$

その後，非常に多くの研究者によって色々な基準の下で $g_n(x)$ の統計的性質が研究されている。それらの結果は，たとえば Prakasa Rao [10] や Devroye and Györfi [6] にまとめられている。一方，大量のデータを扱う場合，経済的理由（メモリの

節約など）あるいは推定量の更新が容易である等の理由で(1.1)の推定量を修正した次のrecursive型の推定量Wolverton-Wagner-Yamato estimatorがWolverton and Wagner [15]とYamato [16]によって提案された。

$$(1.2) \qquad f_n(x) = \frac{1}{n}\sum_{i=1}^{n} h_i^{-p} K\left(\frac{x - X_i}{h_i}\right)$$

$f_n(x)$の統計的性質についてはやはりPrakasa Rao [10]やDevroye and Györfi [6]などにまとめられている。

ところで実際の場面では標本の大きさが確率変数になったりあるいは確率的にした方が良い場合がある。このことを考慮してCarroll [3]は$p=1$でx_0を既知又は未知とするとき，$f(x_0)$を推定する問題に対して2つの型の停止規則を導入しそれらの性質を調べた。第1の型はChow and Robbins [4]の考え方に基づくものであり，次のように定義される。

各$d>0$に対して

$$N(d) = \text{smallest integer } m \geq n_0 \text{ such that}$$
$$m h_m \geq \left(\frac{b}{d}\right)^2 \hat{f}_m(T_m)$$

ただし$b>0$はある定数，n_0は与えられた正の整数，$\hat{f}_m(x)$は(1.1)および(1.2)を含んだ$f(x)$の推定量，T_nは$T_n \to x_0$ a.s. (almost surely) as $n \to \infty$となるx_0の推定

量である。

Carroll [3] は $I_{N(d)} = [\hat{f}_{N(d)}(T_{N(d)}) - d, \hat{f}_{N(d)}(T_{N(d)}) + d]$ を $f(x_0)$ の信頼区間とし，d が十分小さいとき $N(d)$, $E\{N(d)\}$, $P\{f(x_0) \in I_{N(d)}\}$ などの性質を調べた。筆者 [7] も x_0 を既知とした場合第1の型の停止規則について論じた。次に，たとえば Sen and Ghosh [13] の考え方に基づく第2の型は次で定義される。

$L_n < U_n$ a.s. となる $d>0$ に無関係な，大きさ n の標本による統計量 L_n, U_n をみつけて

$N(d) =$ smallest integer $n \geqq n_0$ such that

$$U_n - L_n \leqq 2d$$

このとき $I_{N(d)} = [L_{N(d)}, U_{N(d)}]$ を信頼区間とする。Stute [14] は $p=1$ で x_0 が既知の場合第2の型の停止規則を扱っている。

本報告では推定量として (1.2) を用いて，第1の型の停止規則を導入し，x_0 が既知の場合について論じる。この場合 Carroll [3] の Lemma 4.3 における条件は自動的に成立することがわかる。

§2. 準備

$f(x)$ の推定量を次で与える。

$$f_0(x) \equiv 0$$

$$(2.1) \qquad f_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_i(x, X_i) \qquad n \geqq 1$$

ただし

$$K_n(x, y) = h_n^{-p} K\left(\frac{x-y}{h_n}\right) \qquad x, y \in R^p$$

（註） (2.1) は次のように recursive に書ける。

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

本報告では f, K, h_n についての条件 A と呼ぶ以下の条件を仮定する：

$f(x)$ は R^p 上で有界，かつ R^p 上で有界で連続な２階偏導関数 $\frac{\partial^2 f(x)}{\partial x_i \partial x_j}$ $(i, j = 1, \cdots, p)$ が存在する；

$K(x)$ は R^p 上で有界なボレル可測関数で次をみたす．

$$K(x) \geqq 0, \quad \int_{R^p} K(x)\,dx = 1, \quad \int_{R^p} x_i K(x)\,dx = 1 \quad (i=1, \cdots, p)$$

$$\int_{R^p} \|x\|^2 K(x)\,dx < \infty, \qquad \|x\|^p K(x) \to 0 \quad \text{as} \quad \|x\| \to \infty$$

ここに $\|\cdot\|$ は R^p 上のユークリッドノルムを表わす；

$\{h_n, n \geqq 1\}$ は正数列で次をみたすとする．

(H 1) $\quad h_n \downarrow 0, \quad n h_n^p \uparrow \infty \quad$ as $\quad n \to \infty$, (H 2) $$\frac{n h_n^p}{\log n} \to \infty \quad \text{as} \quad n \to \infty,$$

(H 3) $$\sqrt{h_n^p / n} \sum_{i=1}^{n} h_i^2 \to 0 \quad \text{as} \quad n \to \infty,$$

(H 4) ある正定数 β が存在して

$$\frac{h_n^p}{n} \sum_{i=1}^{n} h_i^{-p} \to \beta \quad \text{as} \quad n \to \infty,$$

(H 5) $\forall \varepsilon > 0$ に対して $\exists \delta > 0$ such that

$$\left| \frac{n}{m} - 1 \right| < \delta \Rightarrow \left| \frac{h_n}{h_m} - 1 \right| < \varepsilon .$$

上の条件をみたす $K(x)$, $\{h_n\}$ の例を与える.

<u>$K(x)$ の例</u>

$$K(x) = \begin{cases} 2^{-p} & \text{if } -1 \leqq x_i \leqq 1 \text{ for all } i = 1, \cdots, p \\ 0 & \text{otherwise} \end{cases}$$

$$K(x) = (2\pi)^{-\frac{p}{2}} \exp\left\{ -\frac{1}{2} \|x\|^2 \right\}$$

$$K(x) = 2^{-p} \exp\left\{ -\sum_{i=1}^{p} |x_i| \right\}$$

<u>$\{h_n\}$ の例</u>

$$h_n = n^{-\frac{r}{p}} \qquad \frac{p}{p+4} < r < 1$$

この場合 $\beta = \dfrac{1}{1+r}$ となる.

以下では

$$B=\beta\int_{R^p}K^2(u)du \quad (\beta \text{は} (H4) \text{にある定数}) \tag{2.2}$$

とおく。

Devroye [5] の Theorem 2 より次の補題が得られる。

補題 2.1

(H3) ~ (H5) を除く f, K, h_n についての条件 A の下で，各 x に対して

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

が成り立つ。

次の補題は筆者 [8] の Theorem から得られる。

補題 2.2

各 $d>0$ に対して $N(d)$ は §1 の確率空間 $(\Omega, \mathcal{F}, P)$ 上で定義された正の整数値をとる確率変数で $n(d)$ は正整数で

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

をみたすとする。このとき (H2) を除く f, K, h_n についての条件 A の下で，$f(x)>0$ なる任意の x に対して

$$\sqrt{N(d)\, h_{N(d)}^{p}}\left(f_{N(d)}(x) - f(x)\right) \xrightarrow[L]{} N(0, B f(x)) \ \text{(in law)}$$
$$\text{as } d \to 0$$

が成り立つ。

次の補題は Chow and Robbins [4] で与えられている。

補題 2. 3

$\{x_n, n \geqq 1\}$ は次をみたす確率変数列であるとする：

$$x_n > 0 \ \text{a.s.}, \quad \lim_{n\to\infty} x_n = 1 \ \text{a.s.}$$

$\{b(n), n \geqq 1\}$ は任意の実数列で次をみたすとする：

$$b(n) > 0, \ \lim_{n\to\infty} b(n) = \infty, \quad \lim_{n\to\infty} \frac{b(n)}{b(n-1)} = 1 .$$

各 $t > 0$ に対して

$$N(t) = \text{smallest integer } m \geqq 1 \text{ such that}$$
$$x_m \leqq \frac{b(m)}{t}$$

として $N(t)$ を定義する。このとき，$N(t)$ は $t\ (\uparrow)$ の非減少関数で

$$\lim_{t\to\infty} N(t) = \infty \ \text{a.s.}, \ \lim_{t\to\infty} E\{N(t)\} = \infty, \ \lim_{t\to\infty} \frac{b(N(t))}{t} = 1 \ \text{a.s.}$$

が成り立つ。さらにもし $E\{\sup_{n\geqq 1} x_n\} < \infty$ ならば

$$\lim_{t\to\infty} \frac{E\{b(N(t))\}}{t} = 1$$

が成り立つ。

§3. 結果

任意に与えられた $\alpha \in (0,1)$ に対して $\Phi(D)-\Phi(-D)=1-\alpha$ となる $D=D_\alpha>0$ をえらぶ。ただし Φ は標準正規分布 $N(0,1)$ の分布関数である。以下では $f(x)>0$ となる任意の x を固定してすべての話を進める。

第1の型の停止規則 $N(d)=N(d,x,\alpha)$ を次で定義する：各 $d>0$ に対して

$N(d)=$ smallest integer $m \geqq 1$ such that

$$\frac{m h_m^p d^2}{D^2 B} \geqq f_m(x)+\frac{1}{m}$$

ただし B は (2.2) で定義された定数である。

以下では次の記号を用いる：

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

$$S_n=\sum_{i=1}^{n} Z_i,\quad v(n)=n h_n^p,\quad g(n)=n h_n^p/f(x),$$

$$y_n=\left(f_n(x)+\frac{1}{n}\right)\Big/f(x),\quad t=D^2 B d^{-2}.$$

また，$C, C_1, C_2, \cdots$ は適当な正の定数を表わすとする。

補題 3.1

(H4) を除く f, K, h_n についての条件 A の下で

$$E\{N(d)h_{N(d)}^p\}<\infty \qquad \text{for each } d>0$$

が成り立つ。

（証明）

$^\forall d>0$ を固定し，$N=N(d)$ とおく．また正整数 l に対して

$$a_l = v(l)P\{N>l\}\ (\geqq 0),\quad b_l = \left(E\{v(l)I(2\leqq N\leqq l)\}\right)^{\frac{1}{2}}$$

とおく。ただし $I(A)$ は A の indicator function を表わす。

補題 2.1 と N の定義を用いると $P\{N<\infty\}=1$

よって十分大きな l に対して $b_l>0$。

$T=\min\{N,l\}$ とおき，十分大きな l に対して考える。

(2.1) を用いて計算すると

$$(3.1)\quad f(x)E\{\sqrt{v(T)}\,y_T I(N\geqq 2)\}\leqq c_1 a_l + c_2 b_l^2 + c_3 b_l + c_4$$

が得られる。一方，N の定義を用いて計算すると

$$(3.2)\quad f(x)E\{\sqrt{v(T)}\,y_T I(N\geqq 2)\}\geqq \frac{\sqrt{v(l)}\,a_l}{x} + \frac{b_l^3}{4x}$$

が得られる。(H1) より $c_1-\dfrac{\sqrt{v(l)}}{x}<0$

よって (3.1), (3.2) より

$$\frac{b_l}{4x}\leqq c_2+\frac{c_3}{b_l}+\frac{c_4}{b_l^2}$$

$b_0=\lim_{l\to\infty} b_l$ とおくと上の不等式で $l\to\infty$ として

$$b_0<\infty$$

一方 $T\uparrow N$ a.s. as $l\to\infty$ だから単調収束定理より

$$\infty > b_0^2 = E\{v(N)I(2\leqq N<\infty)\}$$

よって $E\{N h_N^p\}<\infty$ （証終）

さて $N(d)$ の性質についての定理を次で与える。

定理3.2

f, K, h_n についての条件 A の下で

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

$$P\{N(d)\uparrow\infty \text{ as } d\downarrow 0\}=1,$$

$$\lim_{d\to 0}\frac{N(d)h_{N(d)}^{p}d^{2}}{D^{2}Bf(x)}=1 \quad \text{a.s.},$$

(3. 3) $$\lim_{d\to 0}\frac{E\{N(d)h_{N(d)}^{p}\}d^{2}}{D^{2}Bf(x)}=1$$

が成り立つ。

(証明)

前半の3つの結果は補題2.1, 2.3, Nの定義より得られる。

$y_n'=\min\{1, y_n\}$ とおき

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

$$y_m'\leq\frac{g(m)}{t}$$

で $N'(d)$ を定義する。$g(n)$ の単調性を用いて

$$\frac{E\{g(N'(d))\}}{t}\leq\frac{E\{g(N(d))\}}{t}$$

よって $y_n', N'(d)$ に補題2.3を応用して

$$1 \leqq \varliminf_{d\to 0} \frac{E\{g(N(d))\}}{t}$$

を得る。一方 (2.1), N の定義などを用いて計算すると

$$\varlimsup_{d\to 0} \frac{E\{g(N(d))\}}{t} \leqq 1$$

従って (3.3) が成立する。

(証終)

補題 2.2 と定理 3.2 より次の定理が得られる。

定理 3.3

各 $d>0$ に対して正整数 $n(d)$ が存在して

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

が成り立つと仮定する。このとき，f, K, f_n についての条件 A の下で

$$\lim_{d\to 0} P\{f(x) \in I_{N(d),d}(x)\} = 1-\alpha$$

が成立する。ただし

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

さて §2 の例で与えた $\{f_n\}$ について考えてみる。すなわち

$$h_n = n^{-\frac{r}{p}} \qquad \frac{p}{p+4} < r < 1$$

補題 3.4

f, K についての条件 A の下で，ある定数 $d_0 > 0$ が存在して $\{N(d) d^{\frac{2}{1-r}}, 0 < d < d_0\}$ は一様に可積である。

（証明）

$C = D^2 B$ とおく。次をみたす $d_0 \in (0, \frac{1}{4})$ をえらぶ。

$$(3.4) \qquad \frac{1}{2} < \left(1 - d^{\frac{2}{1-r}}\right)^{r-1-\frac{2r}{p}} < 2$$

また，次をみたす整数 $M \geq 3$ をえらぶ。

$$(3.5) \qquad 2C(f(x)+1)\, m^{r-1} < \frac{1}{4} \quad \text{for all } m \geq M$$

${}^{\forall} d \in (0, d_0)$ と ${}^{\forall} m \geq M$ を固定する。

$$m(d) = [m\, d^{\frac{-2}{1-r}}], \quad A = P\{N(d) d^{\frac{2}{1-r}} > m\}$$

とおく。ただし $[a]$ は a を越えない最大整数を表わす。

テイラーの定理を用いて

$$|\delta_n| \leq C_1 h_n^2$$

が得られる。よって $N(d)$ の定義より

$$A = P\{N(d) > m(d)\}$$
$$\leqq P\{S_{m(d)} + m(d)f(x) + C_1 \sum_{i=1}^{m(d)} h_i^2 > \frac{m(d)^{2-r} d^2}{c} - 1\}$$

従って不等式

$$\sum_{i=1}^{n} h_i^2 \leqq C_2\, n^{1-\frac{2r}{p}} \log n \quad \text{for } n \geqq 3,$$
$$n^{r-1-\frac{2r}{p}} \log n \leqq C_2 \quad \text{for } n \geqq 1$$

と(3.4), (3.5)を用いて計算すると

(3.6) $$A \leqq P\{S_{m(d)} > C_3\, m(d)^{2-r} d^2\}$$

が得られる。計算により

$$|Z_i| \leqq C_4\, m(d)^r \quad \text{for } i = 1, \cdots, m(d)$$
$$\sum_{i=1}^{m(d)} E\{Z_i^2\} \leqq C_5\, m(d)^{1+r}$$

よって Bennett [1] の不等式を用いて

(3.7) $$A \leqq \exp\{-(C_3\, m(d)^{2-r} d^2)^2 / (C_6\, m(d)^{1+r} + C_7\, m(d)^2 d^2)\}$$

が得られる。

(3.8) $$\max\{m(d)^{1+r}, m(d)^2 d^2\} \leqq m^2 d^{-\frac{2(1+r)}{1-r}}$$

だから(3.4)より

(3.9) $$m(d)^{4-2r} d^4 > \frac{1}{2} m^{4-2r} d^{-\frac{4}{1-r}}$$

従って (3.7) ～ (3.9) より

$$A \leqq \exp\{-C_8\, m^{2(1-r)}\} \qquad \text{for all } d \in (0, d_0)$$

故に

$$\sup_{0<d<d_0} P\{N(d)\, d^{\frac{2}{1-r}} > m\}$$

$$\leqq \exp\{-C_8\, m^{2(1-r)}\} \qquad \text{for all } m \geqq M$$

よって

$$(3.10) \qquad \sum_{m=1}^{\infty} \sup_{0<d<d_0} P\{N(d)\, d^{\frac{2}{1-r}} > m\} < \infty$$

だから Bickel and Yahav [2] の Lemma 3.2 を用いると $\{N(d)\, d^{\frac{2}{1-r}},\ 0<d<d_0\}$ は一様に可積分になる。

（証終）

<u>定理 3.5</u>

f, K についての条件 A の下で，定理 3.2, 3.3 の結果が成り立ち，かつ

$$(3.11) \qquad \lim_{d\to 0} \frac{E\{N(d)\}\, d^{\frac{2}{1-r}}}{\left(D^2 B f(x)\right)^{\frac{1}{1-r}}} = 1$$

が成立する。ただし

$$B = \frac{1}{1+r} \int_{R^p} K^2(u)\, du$$

（証明）

定理3.2と補題3.4より

$$\lim_{d\to 0} E\{N(d)\}d^{\frac{2}{1-r}} = (D^2Bf(x))^{\frac{1}{1-r}}$$

よって（3.11）が成立する。他の結果は定理3.2, 3.3より得られる。

（証終）

（註）

Carroll [3] の Lemma 4.3 では（3.10）が成立するための1つの十分条件を仮定して（3.11）の結果を得ている。与えられた点 x が既知の場合に（1.1）の推定量 $g_n(x)$ に対して上の十分条件がみたされることを Schuster [12] は示した。補題3.4の証明中からわかるように点 x が既知の場合に，（1.2）の推定量 $f_n(x)$ に対して（3.10）は自動的に成立する。

References

[1] Bennett, G. (1962). Probability inequalities for the sum of independent random variables. J Amer. Statist. Assoc. 57, 33-45.

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

[3] Carroll, R.J. (1976). On sequential density estimation. Z. Wahrsch. Verw. Gebiete 36, 137-151.

[4] 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.

[5] Devroye, L. (1979). On the pointwise and the integral convergence of recursive kernel estimates of probability densities. Utilitas Mathematica 15, 113-128.

[6] Devroye, L., and Györfi, L. (1985). Nonparametric Density Estimation: The L_1 View. John Wiley & Sons, New York.

[7] Isogai, E. (1981). Stopping rules for sequential density estimation. Bull. Math. Statist. 19, 53-67.

[8] ________. (1987). On the asymptotic normality for non-parametric sequential density estimation. Bull. Inform. Cybernetics, in press.

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

[10] Prakasa Rao, B.L.S. (1983). Nonparametric Functional Estimation. Academic Press.

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

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

[13] 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.

[14] Stute, W. (1983). Sequential fixed-width confidence intervals for a nonparametric density function. Z. Warsch. Verw. Gebiete 62, 113-123.

[15] Wolverton, C.T., and Wagner, T.J. (1969). Asymptotically optimal discriminant functions for pattern classification. IEEE Trans. Inform. Theory IT-15, 258-265.

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