

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, $\{X_n, n \geq 1\}$ はある確率空間 $(\Omega, \mathcal{F}, P)$ 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。

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

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 の Wolverton and Wagner [15] と Yamato [16] により提案された。

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

$f_n(x)$ の統計的性質についてはやはり Prakasa Rao [10] や

Devroye and Györfi [6] などにもとめられている。

ところで実際の場面では標本の大きさや確率変数に依ったりあるいは確率的にした方がよい場合がある。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] の考え方に基つくものであり、次のように定義される。

各 $d > 0$ に対して

$N(d) = \text{smallest integer } m \geq n_0 \text{ such that}$

$$m h_m \geq \left(\frac{b}{d}\right)^{\frac{1}{p}} \hat{f}_m(T_m)$$

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

$T_n \rightarrow x_0$ a.s. (almost surely) as $n \rightarrow \infty$ とする x_0 の推定

量であり。

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

$N(d) = \text{smallest integer } m \geq n_0 \text{ such that}$

$$U_m - L_m \leq 2d$$

となるとき $I_{N(d)} = [L_{N(d)}, U_{N(d)}]$ 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になる。

§2. 準備

$f(x)$ の推定量を次で与える。

$$f_0(x) \equiv 0$$

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

$$f \in \mathcal{L}$$

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

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

$K(x)$ は $\mathbb{R}^p$ 上で有界なボレル可測関数で次をみたす。

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

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

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

$\{h_n, n \geq 1\}$ は正数列で次をみたすとする。

$$(H1) \quad h_n \downarrow 0, \quad nh_n^p \uparrow \infty \quad \text{as } n \rightarrow \infty,$$

$$(H2) \quad \frac{n h_n^p}{\log n} \rightarrow \infty \quad \text{as } n \rightarrow \infty,$$

$$(H3) \quad \sqrt{h_n^p/n} \sum_{i=1}^n h_i^2 \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

(H4) ある正定数 β が存在して

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

(H5) $\forall \varepsilon > 0$ に対して $\exists \delta > 0$ such that

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

上の条件を満たす $K(x)$, $\{h_n\}$ の例を与えよう.

$K(x)$ の例

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

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

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

$\{h_n\}$ の例

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

$$\text{この場合 } \beta = \frac{1}{1+r} \text{ となる.}$$

以下では

$$(2.2) \quad B = \beta \int_{R^p} K^2(u) du \quad (\beta \text{ は (H4) に 依 る 定 数})$$

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

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

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 $\mathcal{F}_1$ 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

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

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

$$\sqrt{N(d)} h_{N(d)}^P \left(\int_{N(d)}^{(d)} - f(x) \right) \xrightarrow{L} N(0, B f(x)) \text{ (in law)} \\ \text{as } d \rightarrow 0$$

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

$\{x_n, n \geq 1\}$ は次をみたす確率変数列で有るとする：

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

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

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

各 $t > 0$ に対して

$N(t) = \text{smallest integer } m \geq 1 \text{ such that}$

$$x_m \leq \frac{b(m)}{t}$$

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

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

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

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

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

$N(d) = \text{smallest integer } m \geq 1 \text{ such that}$

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

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：

$$Z_n = K_n(x, X_n) - E K_n(x, X_n), \quad S_n = E K_n(x, X_n) - f(x),$$

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

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

また, $C, C_1, C_2, \dots$ 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

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

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

$$a_l = v(l) P\{N > l\} (\geq 0), \quad b_l = (E\{v(l) I(2 \leq N \leq l)\})^{\frac{1}{2}}$$

とおく. したがって $I(A)$ は A の indicator function を表す.補題 2.1 と N の定義を用いると $P\{N < \infty\} = 1$ よ, l が十分大なる l に対して $b_l > 0$. $T = \min\{N, l\}$ とおき, l が十分大なる l に対して考える.

(2.1) を用いて計算すると

$$(3.1) \quad f(x) E\{\sqrt{v(T)} y_T I(N \geq 2)\} \leq C_1 a_l + C_2 b_l^2 + C_3 b_l + C_4$$

が得られる. 一方, N の定義を用いて計算すると

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

が得られる. (H1) より $C_1 - \frac{\sqrt{v(l)}}{x} < 0$

よ, (3.1), (3.2) より

$$\frac{b_l}{4x} \leq C_2 + \frac{C_3}{b_l} + \frac{C_4}{b_l^2}$$

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

$$b_0 < \infty$$

一方 $T \uparrow N$ a.s. as $l \rightarrow \infty$ 故に 単調収束定理より

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

よ, $E\{N h_N^p\} < \infty$

(証明終)

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

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

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

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

$$(3.3) \quad \lim_{d \rightarrow 0} \frac{E\{N(d) h_{N(d)}^p\} d^2}{D^2 B f(x)} = 1$$

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。

$$y'_n = \min\{1, y_n\} \quad \text{と置き}$$

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

$$y'_m \leq \frac{g(m)}{\pi}$$

で $N'(d)$ を定義する。 $g(m)$ の単調性を用いて

$$\frac{E\{g(N'(d))\}}{\pi} \leq \frac{E\{g(N(d))\}}{\pi}$$

よって $y'_n, N'(d)$ に補題 2.3 を応用して

$$1 \leq \lim_{d \rightarrow 0} \frac{E\{g(N(d))\}}{\pi}$$

を得る。一方 (2.1), N の定義などを用いて計算すると

$$\lim_{d \rightarrow 0} \frac{E\{g(N(d))\}}{\pi} \leq 1$$

従って (3.3) が成立する。

(証明終)

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

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

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

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

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 $\{h_n\}$ について考えてみる。

すなわち

$$h_n = n^{-\frac{r}{p}} \quad \frac{p}{p+q} < r < 1$$

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 $\{N(d)d^{\frac{2}{1-r}}, 0 < d < d_0\}$ は一様に可積分である。

(証明)

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

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

また, 次をみたす整数 $M \geq 3$ をとる。

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

$\forall d \in (0, d_0)$ 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

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

従, 2 不等式

$$\sum_{i=1}^n h_i^2 \leq C_2 n^{1-\frac{2t}{p}} \log n \quad \text{for } n \geq 3,$$

$$n^{1-\frac{2t}{p}} \log n \leq C_2 \quad \text{for } n \geq 1$$

と (3.4), (3.5) を用いて計算すると

$$(3.6) \quad A \leq P\{S_{m(d)} > C_3 m(d)^{2-t} d^2\}$$

が得られる, 計算により

$$|z_i| \leq C_4 m(d)^r \quad \text{for } i=1, \dots, m(d)$$

$$\sum_{i=1}^{m(d)} E\{z_i^2\} \leq C_5 m(d)^{1+t}$$

よ, 2 Bennett [1] の不等式を用いて

$$(3.7) \quad A \leq \exp\left\{-\frac{(C_3 m(d)^{2-t} d^2)^2}{(C_6 m(d)^{1+t} + C_7 m(d)^2 d^2)}\right\}$$

が得られる.

$$(3.8) \quad \max\{m(d)^{1+t}, m(d)^2 d^2\} \leq m^2 d^{-\frac{2(1+t)}{1-t}}$$

従って (3.4) より

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

従って (3.7) ~ (3.9) より

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

故に

$$\begin{aligned} & \sup_{0 < d < d_0} P \{ N(d) d^{\frac{2}{1-r}} > m \} \\ & \leq \exp \{ -C_8 m^{2(1-r)} \} \quad \text{for all } m \geq M \end{aligned}$$

よって

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

従って 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。

(証明終)

定理 3.5

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

$$(3.11) \quad \lim_{d \rightarrow 0} \frac{E \{ N(d) \} d^{\frac{2}{1-r}}}{(D^2 B f(x))^{\frac{1}{1-r}}} = 1$$

が成立する。 $f \in L$

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

(証明)

定理 3.2 と補題 3.4 より

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

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。

(証明終)

(言註)

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 が既知の場合には (1.1) の推定量 $g_n(x)$ に対して上の十分条件がみたさゆることを 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 nonparametric 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.