

推定関数と推定量 — とくに推定尤度関数と最尤推定量

統計数理研究所 稲垣 宜生

1. 序

観測 X_1, \dots, X_n と parameter θ の関数 $\xi_n(\theta) = \xi_n(X_1, \dots, X_n; \theta)$ を使い, 推定方程式 $\xi_n(\theta) = 0$ を解いた $\theta = \hat{\theta}_n$ を θ の推定量を求める場合がある。ここでは, 可測関数 $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$ が,

$$(1) \quad \xi_n(\hat{\theta}_n) \rightarrow 0 \quad \text{in } P$$

なるときに, 推定関数 $\xi_n(\theta)$ による推定量と呼ぶ。このような推定量, とくに最尤推定量, の一致性や漸近正規性についての議論は多くなされている (たとえば Wilks [6] Section 12.5 を参照)。最近では Huber [2] が, 独立, 同一分布に従う観測に対してゆるい正則条件の下でこれらの性質を推定量が満たしていることを証明した。Markov 従属性をもつ観測に対しても (Rao [7] を参照), 独立非同分布に従う観測に対しても (Inagaki [4] を参照), Huber の方法が適用できることが示される。

我々の目的の一つは、もっと一般的に推定関数と推定量の漸近的關係を論じることである。すなわち、推定関数の parameter θ の代りに、そのある推定量 T_n を入れた統計量の分布 $\mathcal{L}(\xi_n(T_n))$ が

(2) $\mathcal{L}(\xi_n(T_n)) \rightarrow G$ in law, 但し G はある c.d.f. なる場合について論じる。もう一つの目的は、とくに $\xi_n(\theta)$ が推定尤度関数のときは、 $\xi_n(T_n)$ が T_n と最大推定量 $\hat{\theta}_n$ の漸近的な差を表わし、したがってその極限分布 G は、 T_n の漸近的効率の悪さを表わしてゐることを示すことである。

2. 推定関数と推定量の漸近的諸関係

はじめに、推定関数の基本的な性質を仮定する。しかし、これらの性質は“正則条件”から来るものである ([2], [4] (p) を参照)。

仮定

(3) $\frac{1}{\sqrt{n}} \xi_n(T_n) \rightarrow 0$ in P ならば $T_n \rightarrow \theta_0$ in P.

(4) $\mathcal{L}[\xi_n(\theta_0)] \rightarrow N_k(0, S)$, in law.

(5) 任意の正数 M に対し、 n を十分大きくすれば、

$$\sup_{\|t\| \leq M} \|\xi_n(\theta_0 + \frac{t}{\sqrt{n}}) - \xi_n(\theta_0) - \Lambda(\theta_0) \cdot t\| \rightarrow 0 \text{ in P,} \\ \text{as } n \rightarrow \infty.$$

ここで $\Lambda(\theta_0)$ は ある positive definite 行列。

(6) 任意の $\varepsilon > 0$ に対して, ある $\delta > 0$ が存在し, $n \in \mathbb{N}$ が十分大きくとれば,

$$P \left\{ \sup_{\|\tau - \theta_0\| \leq \delta} \frac{\|\xi_n(\tau) - \xi_n(\theta_0) - \Lambda(\theta_0) \sqrt{n}(\tau - \theta_0)\|}{1 + \|\Lambda(\theta_0) \sqrt{n}(\tau - \theta_0)\|} > \varepsilon \right\} \rightarrow 0$$

as $n \rightarrow \infty$.

注意: (3) 式は推定量の一致性を保障する。(4) 式は正則条件の下で中心極限定理により導かれる。(5), (6) 式は, ξ_n の“漸近的可微分性”を示すもので, (6) から (5) は導出されるけれども, 正則条件の下で (6) を示すためには, 先ず (5) を証明する必要がある。

定義: 分布族 $\{\mathcal{L}[\sqrt{n}(T_n - \theta_0)]\}$ が relatively compact であるとは, 任意の $\varepsilon > 0$ に対して, 存在して $M > 0$,

$$(7) \quad P\{\|\sqrt{n}(T_n - \theta_0)\| > M\} < \varepsilon \quad \text{for } \forall n.$$

(5), (6) から次の補題が明らかである。

補題 1.

$\{\mathcal{L}[\sqrt{n}(T_n - \theta_0)]\}$ が relatively compact ならば,

$$(8) \quad \xi_n(T_n) - \xi_n(\theta_0) - \Lambda(\theta_0) \sqrt{n}(T_n - \theta_0) \rightarrow 0 \quad \text{in } P.$$

補題 2.

$T_n \rightarrow \theta_0$ in P ならば,

$$(9) \quad \frac{\xi_n(T_n) - \xi_n(\theta_0) - \Lambda(\theta_0) \sqrt{n}(T_n - \theta_0)}{1 + \|\Lambda(\theta_0) \sqrt{n}(T_n - \theta_0)\|} \rightarrow 0 \quad \text{in } P.$$

上の 2 つの補題より次の 2 つの定理が来る。

定理 1.

$\{\mathcal{L}[\sqrt{n}(T_n - \theta_0)]\}$ が relatively compact であるための必要十分条件は、 $\{\mathcal{L}[\xi_n(T_n)]\}$ が relatively compact であることである。

証明

(必要性) $\{\mathcal{L}[\sqrt{n}(T_n - \theta_0)]\}$ が relatively compact ならば、(8)式が成り立つ。 $\{\mathcal{L}[\xi_n(\theta_0)]\}$ も relatively compact であるから、結局 $\{\mathcal{L}[\xi_n(T_n)]\}$ も relatively compact となる。

(十分性) $\{\mathcal{L}[\xi_n(T_n)]\}$ が relatively compact ならば、

$\frac{1}{\sqrt{n}} \xi_n(T_n) \rightarrow 0$ in P であるから、仮定(3)により、 $T_n \rightarrow \theta_0$ in P 。ゆえに (9)式が成り立つ: for $\forall \varepsilon > 0$, n が十分大であれば、

$$P\left\{ \frac{\|\xi_n(T_n) - \xi_n(\theta_0) - \Lambda(\theta_0)\sqrt{n}(T_n - \theta_0)\|}{1 + \|\Lambda(\theta_0)\sqrt{n}(T_n - \theta_0)\|} > \varepsilon \right\} < \varepsilon$$

$$\therefore P\left\{ \|\Lambda(\theta_0)\sqrt{n}(T_n - \theta_0)\| > \frac{\varepsilon + \|\xi_n(T_n) - \xi_n(\theta_0)\|}{1 - \varepsilon} \right\} < \varepsilon.$$

このことから、 $\{\mathcal{L}[\xi_n(T_n) - \xi_n(\theta_0)]\}$ が relatively compact であることを使えば、 $\{\mathcal{L}[\Lambda(\theta_0)\sqrt{n}(T_n - \theta_0)]\}$ の relatively compact であることが示され、ゆえに $\{\mathcal{L}[\sqrt{n}(T_n - \theta_0)]\}$ の relatively compact であることが示される。 \triangleleft

定理 2.

推定量 $\hat{\theta}_n$ は条件(1)を併せ、 T_n は条件(2)を併せると

仮定する。このとき、

(i) $\hat{\theta}_n$ は漸近的に正規である：

$$(10) \quad \mathcal{L}[\sqrt{n}(\hat{\theta}_n - \theta_0)] \rightarrow N_k(0, \Lambda(\theta_0)^{-1} S(\Lambda(\theta_0)^{-1})'), \text{ in law.}$$

(ii) $\xi_n(T_n)$ と 差 $T_n - \hat{\theta}_n$ は漸近的に同値である：

$$(11) \quad \xi_n(T_n) - \Lambda(\theta_0)\sqrt{n}(T_n - \hat{\theta}_n) \rightarrow 0 \text{ in } P, \text{ したがって}$$

$$(12) \quad \mathcal{L}[\Lambda(\theta_0)\sqrt{n}(T_n - \hat{\theta}_n)] \rightarrow G \text{ in law.}$$

証明

(i) 定理 1 によつて、 $\{\mathcal{L}[\sqrt{n}(\hat{\theta}_n - \theta_0)]\}$ も $\{\mathcal{L}[\sqrt{n}(T_n - \theta_0)]\}$ も relatively compact であるから、 $\hat{\theta}_n$ を閉しても T_n を閉しても

(8) 式が成り立つ。しかも $\xi_n(\hat{\theta}_n) \rightarrow 0$ in P である、

$$(13) \quad \xi_n(\theta_0) + \Lambda(\theta_0)\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow 0 \text{ in } P.$$

これより (i) が出る。

(ii) 更に (8) と (13) から (11) を得る。これと (2) より (12) を得る。

系

$\{\mathcal{L}[\sqrt{n}(T_n - \theta_0)]\}$ が relatively compact のとき、

$$T_n^* = T_n - \Lambda(T_n)^{-1} \frac{1}{\sqrt{n}} \xi_n(T_n)$$

とおけば、 T_n^* は条件 (i) を満たし、したがって (10) 式が成立する。

3. 推定尤度関数と最大推定量

$(\mathcal{X}, \mathcal{Q}, P_0)$ は確率空間で, $\theta \in \Theta (\subset \mathbb{R}^k)$ とする。 X_i , $i=1, 2, \dots$ は独立な観測で, 確率測度 P_0 の下で, それぞれ密度関数 $f_i(\cdot, \theta)$, $i=1, 2, \dots$ を持つ分布に従っている。

仮定

(14) $f_i(x, \theta)$, $i=1, 2, \dots$ は, θ は無関係な共通の support をもち, θ_r , $r=1, \dots, k$ に関して偏微分可能である。その微分係数ベクトルを

$$\begin{aligned} \eta_i(\cdot, \theta) &= \left[\frac{\partial}{\partial \theta} \log f_i(\cdot, \theta) \right]' \\ &= \left(\frac{\partial}{\partial \theta_1} \log f_i(\cdot, \theta), \dots, \frac{\partial}{\partial \theta_k} \log f_i(\cdot, \theta) \right) \end{aligned}$$

とし, 推定関数を

$$\xi_n(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \eta_i(\cdot, \theta)$$

とするとき, P_0 の下で仮定 (3) — (6) が成り立つとしている。

(15) $-\Lambda(\theta_0) = S$ ($= \Gamma$ とおく) である。

(16) 最大推定量 $\hat{\theta}_n$ は,

$$\xi_n(\hat{\theta}_n) \rightarrow 0 \text{ in } P_0.$$

を成り立たせる。

このとき定理 2.4.2 より,

(17) $\mathcal{L}[\sqrt{n}(\hat{\theta}_n - \theta_0)] \rightarrow N_k(0, \Gamma^{-1})$ in law

が成り立つ。更には,

$$\chi_n(h) = \sum_{i=1}^n \log \left\{ f_i(X_i, \theta_0 + \frac{h}{\sqrt{n}}) / f_i(X_i, \theta_0) \right\}$$

とおけば、次のことが明らかである。

補題3.

(i) 任意の $h \in R^k$ に対し、十分大きい n を考えると

$$\chi_n(h) - h' \xi_n(\theta_0) + \frac{1}{2} h' \Gamma h \rightarrow 0 \text{ in } P_{\theta_0}.$$

(ii) $\mathcal{L}[\xi_n(\theta_0); P_{\theta_0}] \rightarrow N_k(0, \Gamma)$ in law.

(iii) 任意の $h \in R^k$ に対し、

$$\mathcal{L}[\chi_n(h); P_{\theta_0}] \rightarrow N_1(-\frac{1}{2} h' \Gamma h, h' \Gamma h), \text{ in law.}$$

この補題から、 $\{P_{n, \theta_0}^x\}$ と $\{P_{n, \theta_0 + \frac{h}{\sqrt{n}}}\}$ は、Contiguous
であるから次の補題が来る。"Contiguous"の概念は LeCam [5]
に於る。

補題4.

確率変数 $Y_n = Y_n(X_1, \dots, X_n)$ と任意の $h \in R^k$ に対し、

(i) $Y_n \rightarrow 0$ in $P_{\theta_0} \iff Y_n \rightarrow 0$ in $\{P_{\theta_0 + \frac{h}{\sqrt{n}}}\}$

(ii) 部分列 $\{m\} \subset \{n\}$ に対し、確率分布 $\mathcal{L}[\xi_0, \Gamma]$ が存在し、

$$\mathcal{L}[\xi_m(\theta_0), Y_m; P_{\theta_0}] \rightarrow \mathcal{L}[\xi_0, \Gamma], \text{ in law}$$

ならば、

$$\mathcal{L}[\xi_m(\theta_0), Y_m; P_{\theta_0 + \frac{h}{\sqrt{m}}}] \rightarrow e^{h' \xi_0 - \frac{1}{2} h' \Gamma h} \mathcal{L}[\xi_0, \Gamma] \text{ in law.}$$

すなわち、有界連続関数 u に対し、

$$\int u(z, y) d\mathcal{L}[\xi_m(\theta_0), Y_m; P_{\theta_0 + \frac{h}{\sqrt{m}}}] (z, y) \rightarrow \int u(z, y) e^{h' z - \frac{1}{2} h' \Gamma h} d\mathcal{L}[\xi_0, \Gamma] (z, y).$$

とくに,

$$(18) \quad \mathcal{L}[\xi_n(\theta_0); P_{\theta_0 + \frac{h}{\sqrt{n}}}] \longrightarrow N_k(\Gamma h, \Gamma) \quad \text{in law.}$$

筆者[3]は, "一様性をもつ" 推定量の極限分布が, 最も推定量の極限分布とある分布との Convolution で表現出来ることを示した。Hájek [1]は, より一般的に条件の下でこのことを示している。しかし, いずれの証明も複雑で, 技工的であるので, 推定関数を用いて, より直観的に証明をしよう。

定理3.

任意の $h \in \mathbb{R}^k$ に対して,

$$(19) \quad \mathcal{L}[\Gamma \sqrt{n}(\tau_n - \theta_0 - \frac{h}{\sqrt{n}}); P_{\theta_0 + \frac{h}{\sqrt{n}}}] \longrightarrow L \quad \text{in law}$$

但し, L は h に無関係な確率分布.

\iff h に無関係な確率分布 G が存在して

$$(20) \quad \mathcal{L}[\xi_n(\tau_n); P_{\theta_0 + \frac{h}{\sqrt{n}}}] \longrightarrow G \quad \text{in law.}$$

更に, (19), (20) のどちらか一方が成り立つとき,

$$(21) \quad L = \tilde{G} * N_k(0, \Gamma)$$

但し, $\tilde{G}(z) = 1 - G(-z)$.

証明

証明が長くなるので (\Leftarrow) のみを示す。条件(20)と定理1から $\{\mathcal{L}[\Gamma \sqrt{n}(\tau_n - \theta_0); P_{\theta_0}]\}$ は relatively compact である。したがって, (8)式が P_{θ_0} の下で成り立つ。ゆえに補題4.(i)から

61)

$$(22) \quad \xi_n(T_n) - \xi_n(\theta_0) + \Gamma \sqrt{n}(T_n - \theta_0) \rightarrow 0 \text{ in } \{P_{\theta_0 + \frac{h}{\sqrt{n}}}\}.$$

$\{\xi_n(\theta_0), \xi_n(T_n); P_{\theta_0}\}$ は relatively compact に収束するので、
任意の部分列 $\{n'\} \subset \{n\}$ に対して、部分列 $\{m\} \subset \{n'\}$ と確率分布
 $\mathcal{L}[\xi_0, \xi_T]$ が存在して

$$(23) \quad \mathcal{L}[\xi_m(\theta_0), \xi_m(T_m); P_{\theta_0}] \rightarrow \mathcal{L}[\xi_0, \xi_T] \text{ in law.}$$

ゆえに補題4(ii)から.

$$(24) \quad \mathcal{L}[\xi_m(\theta_0), \xi_m(T_m); P_{\theta_0 + \frac{h}{\sqrt{m}}}] \rightarrow e^{h'\xi_0 - \frac{1}{2}h'\Gamma h} \mathcal{L}[\xi_0, \xi_T] \text{ in law.}$$

$L \in \mathcal{C}^2$ が τ 周辺分布は (19), (20), (24) から 任意の $h \in \mathbb{R}^k$ に対して

$$\begin{aligned} G(z) &= \lim_{m \rightarrow \infty} \mathcal{L}[\xi_m(T_m); P_{\theta_0 + \frac{h}{\sqrt{m}}}] (z) \\ &= \int \mathcal{L}[\xi_T | \xi_0] (z) \phi(\xi_0; \Gamma h, \Gamma) d\xi_0 \end{aligned}$$

但し $\phi(\cdot; \mu, \Sigma)$ は $N_k(\mu, \Sigma)$ の密度関数

$\{\phi(\cdot; \Gamma h, \Gamma)\}_{h \in \mathbb{R}^k}$ は complete measures であるから

$$(25) \quad G(z) = \mathcal{L}[\xi_T | \xi_0] (z) \quad \text{a.e. } \xi_0 \text{ and any } z.$$

一方 (22) から 同様に

$$\begin{aligned} & \lim_{m \rightarrow \infty} \mathcal{L}[\Gamma \sqrt{m}(T_m - \theta_0 - \frac{h}{\sqrt{m}}); P_{\theta_0 + \frac{h}{\sqrt{m}}}] (z) \\ &= \int_{-\xi_T + \xi_0 - \Gamma h \leq z} e^{h'\xi_0 - \frac{1}{2}h'\Gamma h} d\mathcal{L}[\xi_0, \xi_T] \\ &= \int e^{h'\xi_0 - \frac{1}{2}h'\Gamma h} d\mathcal{L}(\xi_0) \int_{\xi_T \geq \xi_0 - z - \Gamma h} d\mathcal{L}[\xi_T | \xi_0] \\ & \quad \text{ここで (25) を使えば} \\ &= \int \{1 - G(\xi_0 - z - \Gamma h)\} \phi(\xi_0; \Gamma h, \Gamma) d\xi_0 \\ &= \int \tilde{G}(z-x) \phi(x; 0, \Gamma) dx \\ &= \tilde{G} * N_k(0, \Gamma) (z), \end{aligned}$$

これは部分列 $\{m\}$ のとり方によらぬ、ゆえに

$$\lim_{n \rightarrow \infty} \mathcal{L} \left[\Gamma \sqrt{n} \left(T_n - \theta_0 - \frac{h}{\sqrt{n}} \right); P_{\theta_0 + \frac{h}{\sqrt{n}}} \right] (z) = \tilde{G} * N_{\theta_0}(0, \Gamma)(z)$$

又極限分布は R に無関係である。丁は $U(9)$, (21) が示され下。 \triangleleft

参考文献

- 1 Hajék, J. (1970). A characterization of limiting distributions of regular estimates, Z. Wahrsch. verw. Geb., 14, 323-330.
- 2 Huber, P. J. (1967). The behavior of maximum likelihood estimators under nonstandard conditions, Proc. Fifth Berkeley Symp. Math. Statist. Prob., 1221-233.
- 3 Inagaki, N. (1970). On the limiting distribution of a sequence of estimators with uniformity property, Ann. Inst. Statist. Math., 22, 1-13.
- 4 _____ Asymptotic relations between the likelihood estimating function and the maximum likelihood estimator,
- 5 LeCam, L. (1960). Locally asymptotically normal families of distributions, Univ. California Publ. Statist. 3, 37-98.
- 6 Wilks, S. (1962). Mathematical Statistics, John Wiley Sons, Inc., New York.
- 7 Rao, B. L. S. Prakasa (1972). Maximum likelihood estimation for markov processes, Ann. Inst. Statist. Math., (to appear).