数理解析研究所講究録
167巻 1972年 52-61

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 宣生

1．序

観測 $X_1, \cdots, X_n$ と parameter θ の関数 $\xi_n(\theta) = \xi_n(X_1 \cdots X_n; \theta)$ を使い，推定方程式 $\xi_n(\theta) = 0$ をみたす $\theta = \hat{\theta}_n$ をもって θ の推定量を求める場合がある。ここでは，可測関数 $\hat{\theta}_n = \hat{\theta}_n(X_1 \cdots X_n)$ が

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

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み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），独立非同一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1つは、もっと一般的に推定関数と推定量の漸近的関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、それのある推定量 T_n を入れた統計量の分布 $\mathcal{L}(\xi_n(T_n))$ が

(2) $\mathcal{L}(\xi_n(T_n)) \longrightarrow G$ in law, 但し G はある c.d.f.

なる場合について論じる。もう1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 と最尤推定量 $\hat{\theta}_n$ 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い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) $\frac{1}{\sqrt{n}}\xi_n(T_n) \to 0$ in P ならば $T_n \to \theta_0$ in P.

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

(5) 任意の正数 M に対し、n を十分大きくとれば,

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

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 $d>0$ が存在し、n を十分大きくとれば、

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

$$\text{as } n\longrightarrow\infty.$$

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

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

(5),(6)から次の補題が明らかである。

補題1.

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

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

補題2.

$T_n\to\theta_0$ in P ならば、

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

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)式が成立つ。$\{\mathcal{L}[\xi_n(\theta_0)]\}$ も relatively compact だから，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，$\frac{1}{\sqrt{n}}\xi_n(T_n)\longrightarrow 0$ in P であるから，仮定(3)により，$T_n\to\theta_0$ 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，

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

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

このことと，$\{\mathcal{L}[\xi_n(T_n)-\xi_n(\theta_0)]\}$ 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。◭

定理 2.

推定量 $\hat{\theta}_n$ は条件(1)をみたし，T_n は条件(2)をみたすと

仮定する。このとき,

(i) $\hat{\theta}_n$ は漸近的に正規である:

(10) $\mathcal{L}[\sqrt{n}(\hat{\theta}_n-\theta_0)] \longrightarrow N_k(0, \Lambda(\theta_0)^{-1}S(\Lambda(\theta_0)^{-1})')$, 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) $\xi_n(T_n)-\Lambda(\theta_0)\sqrt{n}(T_n-\hat{\theta}_n) \longrightarrow 0$ in P, したがって

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

証明

(i) 定理1によって. $\{\mathcal{L}[\sqrt{n}(\hat{\theta}_n-\theta_0)]\}$ も $\{\mathcal{L}[\sqrt{n}(T_n-\theta_0)]\}$ も relatively compact であるから. $\hat{\theta}_n$ に関しても T_n に関しても (8) 式が成り立つ。しかも $\xi_n(\hat{\theta}_n) \longrightarrow 0$ in P だから,

(13) $\xi_n(\theta_0)+\Lambda(\theta_0)\sqrt{n}(\hat{\theta}_n-\theta_0) \longrightarrow 0$ 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 (1) をみたし. したがって (10) 式が成立する。

3. 推定尤度関数と最尤推定量

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

仮定

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

$$\eta_i(\cdot,\theta) = \left[\frac{\partial}{\partial\theta}\log f_i(\cdot,\theta)\right]' = \left(\frac{\partial}{\partial\theta_1}\log f_i(\cdot,\theta),\cdots,\frac{\partial}{\partial\theta_k}\log f_i(\cdot,\theta)\right)'$$

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_{θ_0} の下で仮定 (3) — (6) がみたされている。

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

(16) 最尤推定量 $\hat{\theta}_n$ は，

$$\xi_n(\hat{\theta}_n) \longrightarrow 0 \quad \text{in } P_{\theta_0}$$

をみたしている。

このとき定理2によって，

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

が成り立つ。更に，

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

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、十分大きな n を考えると

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

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

(iii) 任意の $h \in R^k$ に対し、

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

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

補題4.

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

(i) $Y_n \longrightarrow 0$ in P_{θ_0} $\Longleftrightarrow$ $Y_n \longrightarrow 0$ in $\{P_{\theta_0+\frac{h}{\sqrt{n}}}\}$

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

$$\mathcal{L}[\xi_m(\theta_0), Y_m : P_{\theta_0}] \longrightarrow \mathcal{L}[\xi_0, Y], \text{ in law}$$

ならば、

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

すなわち、有界連続関数 u に対して、

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

とくに、

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

筆者[3]は"一様性をもつ"推定量の極限分布が、最尤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]はもっと一般的な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、推定関数を使ってもっと直観的に証明をしよう。

定理3.

任意の $h \in R^k$ に対して、

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

但し、L は h に無関係な確率分布.

$\Longleftrightarrow$ h に無関係な確率分布 G が存在して

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

更に、(19), (20) のどちらか一方が成り立つとき、

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

但し. $\tilde{G}(z) = 1 - G(-z)$.

証明

証明が長くなるので($\Leftarrow$)のみ示す。条件(20)と定理1から $\{\mathcal{L}[\Gamma\sqrt{n}(T_n-\theta_0) : P_{\theta_0}]\}$ 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

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

$\{\mathcal{L}[\xi_n(\theta_0), \xi_n(T_n); P_{\theta_0}]\}$ は relatively compact になるので，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) $\quad \mathcal{L}[\xi_m(\theta_0), \xi_m(T_m); P_{\theta_0}] \longrightarrow \mathcal{L}[\xi_0, \xi_T] \quad \text{in law}.$

ゆえに補題4 (ii) から．

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

したがって周辺分布は，(19), (20), (24) から 任意の $h \in R^k$ に対して

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

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

$\{\phi(\cdot\,; \Gamma h, \Gamma)\}_{h\in R^k}$ 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

$$\lim_{m\to\infty} \mathcal{L}[\Gamma\sqrt{m}(T_m - \theta_0 - \tfrac{h}{\sqrt{m}}); P_{\theta_0+\frac{h}{\sqrt{m}}}](z)$$
$$= \int_{-\xi_T+\xi_0-\Gamma h \le z} e^{h'\xi_0 - \frac{1}{2}h'\Gamma h}\,d\mathcal{L}[\xi_0, \xi_T]$$
$$= \int e^{h'\xi_0 - \frac{1}{2}h'\Gamma h}\,d\mathcal{L}(\xi_0) \int_{\xi_T \ge \xi_0 - z - \Gamma h} d\mathcal{L}[\xi_T | \xi_0]$$

ここで (25) を使えば

$$= \int \{1 - G(\xi_0 - z - \Gamma h)\}\,\phi(\xi_0; \Gamma h, \Gamma)\,d\xi_0$$
$$= \int \tilde{G}(z - x)\,\phi(x; 0, \Gamma)\,dx$$
$$= \tilde{G} * N_k(0, \Gamma)(z),$$

これは部分列$\{m\}$のとり方によらない，ゆえに

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

又極限分布はhに無関係である。すなわち(19),(21)が示された。

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 btween the likelihood estimating function and the maximum likelihood estimator,

5 LeCam,L.(1960). Locally asymptotically normal families of distributions, Univ.Califrnia 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).