第7章　Peeling 議論

この章では, M 推定量の収束のレートを経験過程[1]の連続性の係数から導出できることをみていく.

7.1　第一の方法

Θ を母数空間, $\mathbb{X}$ を標本空間とし, P を $\mathbb{X}$ 上の確率測度とする. P に対応する真の母数を θ^* と書くことにする. $\theta \in \Theta$ で添え字付られた可測関数 $\gamma_\theta : \mathbb{X} \longrightarrow \mathbb{R}$ を考える. $X, X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P}$ とし, $\widehat{\mathsf{P}}_n$ を $X_1, X_2, \ldots, X_n$ に基づく経験測度とする. すると M 推定量 $\widehat{\theta}_n$ は

$$\Theta \ni \theta \mapsto \widehat{\mathsf{P}}_n\gamma_\theta := \frac{1}{n}\sum \gamma_{j=1}^{n}\gamma_\theta(X_j)$$

を最大化するものである. 母数コントラスト

$$\Theta \ni \theta \mapsto \mathsf{P}\gamma_\theta := \int_{\mathbb{X}} \gamma_\theta \,\mathrm{d}\mathsf{P}$$

は, Θ 上のある距離 d に対して

$$\mathsf{P}\gamma_\theta - \mathsf{P}\gamma_{\theta^*} \precsim -d^2(\theta, \theta^*) \tag{7.1}$$

をみたすとする.

定理 7.1. (7.1) をみたすとする. 関数 $\phi_n : \mathbb{X} \longrightarrow \mathbb{R}$ は, ある $\alpha < 2$ に対して

$$\delta \mapsto \frac{\phi_n(\delta)}{\delta^\alpha}$$

は非増加で

$$\mathsf{E}\left[\sup_{\theta:\, d(\theta,\, \theta^*)<\delta}\left|\nu_n(\gamma_\theta - \gamma_{\theta^*})\right|\right] \precsim \phi_n(\delta) \tag{7.2}$$

をみたすとする. このとき, 任意の数列 $\{r_n\}_{n=1}^{\infty}$ で

$$\phi_n(1/r_n) \leq \frac{\sqrt{n}}{r_n^2} \tag{7.3}$$

[1] M 推定量を定義するコントラスト関数によって添え字付られた経験過程を考える.

をみたすものに対して

$$d\left(\widehat{\theta}_n,\, \theta^*\right) = O_P\left(\frac{1}{r_n}\right)$$

となる.

Proof. 議論を簡単にするために

$$\left\{\widehat{\theta}_n\right\} = \arg\max_{\theta\in\Theta} \widehat{\mathsf{P}}_n\gamma_\theta$$

と仮定する. 各 $n \in \mathbb{N}$ に対して, $\Theta \setminus \theta^*$ の「シェル」$\left\{S_{n,j};\, j \in \mathbb{Z}\right\}$ を

$$S_{n,j} := \left\{\theta \in \Theta;\, 2^{j-1} < r_n d(\theta,\, \theta^*) \le 2^j\right\} \qquad (j \in \mathbb{Z})$$

で定める. ある $M > 0$ が存在して, $r_n d\left(\widehat{\theta}_n,\, \theta^*\right) > 2^M$ ならば, ある $j > M$ が存在して

$$\widehat{\theta}_n \in S_{n,j} \;\Leftrightarrow\; \widehat{\theta}_n \text{ は } S_{n,j} \ni \theta \mapsto \widehat{\mathsf{P}}_n \text{ を最大化}$$

となる. したがって

$$\sup_{\theta\in S_{n,j}} \left(\widehat{\mathsf{P}}_n\gamma_\theta - \widehat{\mathsf{P}}_n\gamma_{\theta^*}\right) \ge 0$$

となる. このことを踏えると

$$\limsup_{n\to\infty} \mathsf{Pr}^*\left(r_n d\left(\widehat{\theta}_n,\, \theta^*\right) > 2^M\right) \overset{M\to\infty}{\longrightarrow} 0$$

を示せばよい. しかし, 任意の $\eta > 0$ と $J > M$ $(J \in \mathbb{N})$ に対して,

$$\begin{aligned}
&\mathsf{Pr}^*\left(r_n d\left(\widehat{\theta}_n,\, \theta^*\right) > 2^M\right) \\
&\quad\le \mathsf{Pr}^*\left(r_n d\left(\widehat{\theta}_n,\, \theta^*\right) > 2^M,\, r_n d\left(\widehat{\theta}_n,\, \theta^*\right) \le 2^J\right) \\
&\qquad + \mathsf{Pr}^*\left(r_n d\left(\widehat{\theta}_n,\, \theta^*\right) > 2^J > \frac{\eta r_n}{2}\right) \\
&\quad\le \sum_{M<j\le J} \mathsf{Pr}^*\left(\sup_{\theta\in S_{n,j}} \left(\widehat{\mathsf{P}}_n\gamma_\theta - \widehat{\mathsf{P}}_n\gamma_{\theta^*}\right) \ge 0\right) \\
&\qquad + \mathsf{Pr}^*\left(2d\left(\widehat{\theta}_n,\, \theta^*\right) > 2^J > \eta\right)
\end{aligned}$$

を得る. η を十分小さくとり, $d(\theta,\, \theta^*) \le \eta$ に対して, 条件 (7.1) が成立すると仮定する. さらに, すべての $0 < \delta < \eta$ に対して, 条件 (7.2) が成立すると仮定する. すると $M < j \le J$ なる任意の j と $\forall\theta \in S_{n,j}$ に対して

$$\mathsf{P}\gamma_\theta - \mathsf{P}\gamma_{\theta^*} \precsim -d^2\left(\theta,\, \theta^*\right) \precsim -\frac{2^{2(j-1)}}{r_n^2}$$

となる[2]. 中心化された過程 $\widetilde{\nu}_n(\theta) := \widehat{P}_n\gamma_\theta - \mathsf{P}\gamma_\theta$ と書く. すると $\forall\theta \in S_{n,j}$ に対して,

$$
\begin{aligned}
&\widehat{\mathsf{P}}_n\gamma_\theta - \widehat{\mathsf{P}}\gamma_{\theta^*} \geq 0 \\
&\qquad \Rightarrow\ \left|\widetilde{\nu}_n(\theta) - \widetilde{\nu}_n(\theta^*)\right| \geq \left(\widehat{\mathsf{P}}_n\gamma_\theta - \widehat{\mathsf{P}}_n\gamma_{\theta^*}\right) - \left|\mathsf{P}\gamma_\theta - \mathsf{P}\gamma_{\theta^*}\right| \geq \frac{2^{2(j-1)}}{r_n^2}
\end{aligned}
$$

となることに注意する. Markov の不等式と条件 (7.2) から, $M < j \leq J$ に対して

$$
\begin{aligned}
\mathsf{Pr}^*&\left(\sup_{\theta\in S_{n,j}}\left(\widehat{\mathsf{P}}_n\gamma_\theta - \widehat{\mathsf{P}}_n\gamma_{\theta^*}\right) \geq 0\right) \\
&\leq \mathsf{Pr}^*\left(\sup_{\theta\in S_{n,j}}\left|\widetilde{\nu}_n(\theta) - \widetilde{\nu}_n(\theta^*)\right| \geq \frac{2^{2(j-1)}}{r_n^2}\right) \\
&\leq \mathsf{Pr}^*\left(\sup_{\theta\in S_{n,j}}\left|\nu_n(\theta) - \nu_n(\theta^*)\right| \geq \frac{\sqrt{n}2^{2(j-1)}}{r_n^2}\right) \\
&\leq \frac{r_n^2}{\sqrt{n}2^{2(j-1)}}\mathsf{E}\left[\sup_{\theta\in S_{n,j}}\left|\nu_n(\theta) - \nu_n(\theta^*)\right|\right] \\
&\leq \frac{r_n^2}{\sqrt{n}2^{2(j-1)}}\mathsf{E}\left[\sup_{d(\theta,\theta^*)\leq 2^j/r_n}\left|\nu_n(\theta) - \nu_n(\theta^*)\right|\right] \\
&\leq \frac{r_n^2}{\sqrt{n}2^{2(j-1)}}\phi(2^j/r_n)
\end{aligned}
$$

を得る. $\delta \mapsto \phi_n(\delta)/\delta^\alpha$ は非増加なので, $c > 1$ に対して

$$
\frac{\phi_n(c\delta)}{(c\delta)^\alpha} \leq \frac{\phi_n(\delta)}{\delta^\alpha}
$$

となる. これを解くと

$$
\phi_n(c\delta) \leq c^\alpha\phi_n(\delta)
$$

となることに注意すると

$$
\phi_n(2^j/r_n) \leq 2^{j\alpha}\phi_n(1/r_n)
$$

となる. さらに, (7.3) から

$$
\frac{\phi_n(1/r_n)r_n^2}{\sqrt{n}} \leq 1
$$

となるので

$$
\frac{r_n^2}{\sqrt{n}2^{2(j-1)}}\phi(2^j/r_n) \leq \frac{2^{j\alpha}\phi_n(1/r_n)r_n^2}{\sqrt{n}2^{2(j-1)}} \leq 4 \times 2^{j\alpha-2j}
$$

[2] $\theta \in S_{n,j}$ なので, $2^{j-1} < r_n d(\theta, \theta^*)$ となることからわかる.

がわかる. 以上から

$$\sum_{M<j\leq J} \mathsf{Pr}^*\left(\sup_{\theta\in S_{n,\,j}} \left(\widehat{\mathsf{P}}_n\gamma_\theta - \widehat{\mathsf{P}}_n\gamma_{\theta^*}\right) \geq 0\right) \leq 4\sum_{j>M} 2^{j\alpha-2j} \xrightarrow{M\to\infty} 0$$

を得る. □

例 7.2. P を $\mathbb{R}$ 上の確率分布とし, Lebesgue 測度 m に関する p.d.f. p を持つとし

$$X_1,\, X_2,\, \ldots,\, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P}$$

とする. $\alpha > 0$ を固定し

$$m_\theta(x) := \mathbb{1}_{[\theta-\alpha,\,\theta+\alpha]}(x)$$

とする. すると

$$\widehat{\mathsf{P}}_n m_\theta = \frac{1}{n}\#\{X_j \in [\theta-\alpha,\, \theta+\alpha];\, j = 1,\, 2,\, \ldots,\, n\},$$
$$\mathsf{P}m_\theta = \mathsf{Pr}\big(|X-\theta| \leq \alpha\big) = \mathsf{F}_X(\theta+\alpha) - \mathsf{F}_X(\theta-\alpha)$$

となる. ただし, $\mathsf{F}_X = \int_{-\infty}^x \mathsf{p}(t)\,\mathrm{d}\mathsf{m}(t)$ である. ここで

$$\theta_0 = \arg\max_{\theta\in\mathbb{R}} \mathsf{P}m_\theta$$

と定める. このとき

$$\widehat{\theta}_n := \arg\max_{\theta\in\mathbb{R}} \widehat{\mathsf{P}}_n m_\theta \xrightarrow{\mathrm{P}} \theta_0 \qquad (n\to\infty)$$

となるかを調べよう. 上記が成立するならば

$$r_n\big(\widehat{\theta}_n - \theta_0\big) \begin{cases} = O_p(1) & (\text{ある } r_n \to \infty), \\ \rightsquigarrow \nu & (\nu \text{ は極限確率過程}) \end{cases}$$

を調べよう. そのために

$$\mathcal{G} = \{m_\theta;\, \theta\in\mathbb{R}\}$$

とおく. すると $\mathcal{G}$ の要素 g の $\mathrm{subgraph}(g) = \{(x,\, t)\in\mathbb{R}^2;\, g(x)\geq t\}$ に対して

$$\mathcal{D} := \{\mathrm{subgraph}(g);\, g\in\mathcal{G}\}$$

とおくと, $\mathcal{D}$ の VC 次元は

$$\mathrm{VC}^{\mathcal{D}} = \inf\{n\in\mathbb{N};\, \mathrm{VC}^{\mathcal{D}}(n) < 2^n\} = 2$$

となる. $\delta > 0$ に対して

$$\mathcal{G}^{\delta}(\theta_0) := \bigl\{m_\theta - m_{\theta_0};\, d(\theta,\,\theta_0) < \delta\bigr\}, \qquad \mathcal{G}^{\infty}(\theta_0) = \bigl\{m_\theta - m_{\theta_0};\, \theta \in \mathbb{R}\bigr\},$$

とおく. 定理 6.47 から, 任意の $\mathbb{R}$ 上の確率分布 Q に対して

$$N\bigl(\epsilon \mathsf{Q}G,\, \mathcal{G},\, \|\cdot\|_{L_1(\mathsf{Q})}\bigr) \le \max\Biggl\{A\epsilon^{-4},\, e^{\epsilon/4}\Biggr\}$$

となる. ただし, $A > 0$ は定数である. $G(x) = \sup_{\theta\in\mathbb{R}} |m_\theta \le 1$ なので

$$\mathsf{P}|m_\theta - m_{\theta_0}|^2 \le \mathsf{P}|m_\theta - m_{\theta_0}|$$

である. よって

$$\begin{aligned} N\bigl(\epsilon,\, \mathcal{G}^{\delta}(\theta_0),\, \|\cdot\|_{L_2(\mathsf{Q})}\bigr) &\le N\bigl(\epsilon,\, \mathcal{G}^{\infty}(\theta_0),\, \|\cdot\|_{L_2(\mathsf{Q})}\bigr)] \\ &\le N^2\bigl(\epsilon/2,\, \mathcal{G},\, \|\cdot\|_{L_2(\mathsf{Q})}\bigr) \\ &\le N^2\bigl(\epsilon 2/2,\, \mathcal{G},\, \|\cdot\|_{L_1(\mathsf{Q})}\bigr) \\ &\le N^2\bigl(\epsilon \mathsf{Q}G/2,\, \mathcal{G},\, \|\cdot\|_{L_1(\mathsf{Q})}\bigr) \\ &\le A^2\left(\frac{2}{\epsilon}\right)^8 \end{aligned}$$

となる. よって

$$\begin{aligned} \int_0^\infty \sqrt{\log N\bigl(\epsilon,\, \mathcal{G}^{\delta}(\theta_0),\, \|\cdot\|_{L_2(\mathsf{Q})}\bigr)}\, \mathrm{d}\epsilon &= \int_0^1 \sqrt{\log N\bigl(\epsilon,\, \mathcal{G}^{\delta}(\theta_0),\, \|\cdot\|_{L_2(\mathsf{Q})}\bigr)}\, \mathrm{d}\epsilon \\ &\le \int_0^1 \sqrt{8\log(2/\epsilon)}\, \mathrm{d}\epsilon + \sqrt{2\log A} \\ &< \infty \end{aligned}$$

となる. さらに

$$G^{\delta}(x) := \sup\{|m_\theta(x) - m_{\theta_0}(x);\, d(\theta,\, theta_0) < \delta\} \qquad (x \in \mathbb{R})$$

とおくと, $\delta < \alpha$ に対して

$$G^{\delta}(x) \le \mathbb{1}_{[\theta_0+\alpha-\delta,\, \theta_0+\alpha+\delta]}(x) + \mathbb{1}_{[\theta_0-\alpha-\delta,\, \theta_0-\alpha+\delta]}(x)$$

となる. よって

$$\begin{aligned} \mathsf{P}\bigl\{G^{\delta}\bigr\}^2 &\le \mathsf{Pr}(\theta_0 + \alpha - \delta \le X \le \theta_0 + \alpha + \delta) + \mathsf{Pr}(\theta_0 - \alpha - \delta \le X \le \theta_0 - \alpha + \delta) \\ &\le 4\|\mathsf{p}\|_\infty \delta \end{aligned}$$

となる. したがって

$$\|G^{\delta}\|_{L_2(\mathsf{P})} \le 2\sqrt{\|\mathsf{p}\|_\infty}\,\delta^{1/2}$$

となる. すると

$$\mathsf{E}\left[\sup_{g\in\mathcal{G}^\delta(\theta_0)}\left|(\widehat{\mathsf{P}}_n-\mathsf{P})g\right|\right]\precsim\|G^\delta\|_{L_2(\mathsf{P})}\int_0^\infty\sqrt{\log N\big(\epsilon,\mathcal{G}^\delta(\theta_0),\|\cdot\|_{L_2(\mathsf{Q})}\big)}\,\mathrm{d}\epsilon$$
$$\precsim\sqrt{\delta}=:\phi(\delta)$$

となる. (7.1) を満足するためには, p は unimodal であればよい. 実際

$$\mathsf{P}m_\theta-\mathsf{P}m_{\theta_0}=\frac{1}{2}\left(\overset{\bullet}{\mathsf{p}}(\theta_0+\alpha)-\overset{\bullet}{\mathsf{p}}(\theta_0-\alpha)(\theta-\theta_0)^2+o(|\theta-\theta_0|^2)\right)$$

であり, $\overset{\bullet}{\mathsf{p}}(\theta_0-\alpha)>0$ かつ $\overset{\bullet}{\mathsf{p}}(\theta_0+\alpha)<0$ であるので

$$\mathsf{P}m_\theta-\mathsf{P}m_{\theta_0}\precsim-(\theta-\theta_0)^2$$

がわかる. $\phi(\delta)=\phi_n(\delta)=\sqrt{\delta}$ としたとき, $r_n=n^{1/3}$ とすると

$$\phi_n\left(\frac{1}{r_n}\right)=n^{-1/6},$$
$$\frac{\sqrt{n}}{r_n^2}=n^{1/2}n^{-2/3}=n^{-1/6}$$

となるので

$$n^{1/3}(\widehat{\theta}_n-\theta_0)=O_p(1)$$

がわかる. □

系 7.3. $\Theta\subset\mathbb{R}^d$ を開部分集合とし, 各 $\boldsymbol{\theta}\in\Theta$ に対して, 関数

$$\mathbb{R}\ni x\mapsto m_{\boldsymbol{\theta}}(x)\in\mathbb{R}$$

は可測とし, $\boldsymbol{\theta}_0:=\arg\max_{\boldsymbol{\theta}\in\Theta}\mathsf{P}m_{\boldsymbol{\theta}}$ とする. さらに, $\mathsf{P}\,\overset{\bullet}{m}^2<\infty$ なる関数 $\overset{\bullet}{m}\colon\mathbb{R}\longrightarrow\mathbb{R}$ が存在して, $\boldsymbol{\theta}_0$ のある近傍に含まれる任意の $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ に対して

$$\left|m_{\boldsymbol{\theta}_1}(x)-m_{\boldsymbol{\theta}_2}(x)\right|\leq\overset{\bullet}{m}(x)|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2|_{2,d}$$

をみたすとする. さらに, 関数 $\Theta\ni\boldsymbol{\theta}\mapsto\mathsf{P}m_{\boldsymbol{\theta}}$ は $\boldsymbol{\theta}_0$ のある近傍で 2 次までの Taylor 展開をもち, Hessian 行列式は退化しないとする. $\widehat{\boldsymbol{\theta}}_n=\arg\max_{\boldsymbol{\theta}\in\Theta}\widehat{\mathsf{P}}_n m_{\boldsymbol{\theta}}$ は

$$\widehat{\mathsf{P}}_n m_{\widehat{\boldsymbol{\theta}}_n}\geq\widehat{\mathsf{P}}_n m_{\boldsymbol{\theta}_0}-O_p(n^{-1})\quad\text{かつ}\quad\widehat{\boldsymbol{\theta}}_n\xrightarrow{\mathrm{P}}\boldsymbol{\theta}_0\,(n\to\infty)$$

ならば

$$\sqrt{n}\big(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0\big)=O_p(1)$$

となる.

Proof. □

7.2 第二の方法