第10章　M 推定法

10.1　M 推定量とは

$(\Omega, \mathcal{A}, \mathsf{Pr})$ を確率空間とする. $\mathbb{M}$ を距離空間とし, $\mathcal{B}(\mathbb{M})$ を $\mathbb{M}$ の開集合を含む最小の σ 加法族とし, $X : \Omega \longrightarrow \mathbb{M}$ は分布 P をもつ確率要素とする. すなわち,

$$\mathsf{P}(B) = \mathsf{Pr}(\{\omega \in \Omega;\, X(\omega) \in B\}) \qquad (B \in \mathcal{B}(\mathbb{M}))$$

である.

$X_1, X_2, \ldots, X_n$ は X の独立複製とする. (Θ, d) を距離空間とし, 距離関数 d によって誘導されるノルムを $\|\cdot\|_d$ と書く. ある $\theta \in \Theta$ に対し,

$$\gamma_\theta : \mathbb{M} \longrightarrow \mathbb{R}$$

をある損失関数（可測）とする. すべての $\theta \in \Theta$ に対して

$$\mathsf{P}|\gamma_\theta| < \infty$$

とする.

未知の母数

$$\theta^* := \arg\min_{\theta \in \Theta} \mathsf{P}\gamma_\theta \tag{10.1}$$

を M 推定量

$$\widehat{\theta}_n := \widehat{\theta}_n(X_1, X_2, \ldots, X_n) := \operatorname*{argmin}_{\theta \in \Theta} \widehat{\mathsf{P}}_n \gamma_\theta$$

で推定することを考える.

仮定

- θ^* は一意的に存在.
- $\widehat{\theta}_n$ は存在（一意でないかもしれない）.

例 10.1. (i) 位置母数推定. $\mathbb{M} = \mathbb{R}$, $\Theta = \mathbb{R}$ とする.

(ia) $\gamma_\theta(x) = (x - \theta)^2$.

(ib) $\gamma_\theta(x) = |x - \theta|$.

(ii) 最尤推定. $\{p_\theta;\, \theta \in \Theta\}$ は σ 有限な測度 μ に支配されている密度関数の族とし

$$\gamma_\theta := -\log \mathsf{p}_\theta$$

とする.

もし

$$\frac{\mathrm{d}\mathsf{P}}{d\mu} = \mathsf{p}_{\theta^*} \qquad (\theta^* \in \Theta)$$

ならば, θ^* は

$$\Theta \ni \theta \mapsto \mathsf{P}\gamma_\theta = \int_{\mathcal{X}} \gamma_\theta(x) \mathsf{p}_\theta(x)\, \mathrm{d}\mu(x) \in [0,\, \infty)$$

を最小化する.

(iia) Poisson 分布: $\theta > 0$ に対して,

$$\mathsf{p}_\theta(x) = e^{-\theta} \frac{\theta^x}{x!} \qquad (x = 0,\, 1,\, 2,\, \ldots)$$

(iib) ロジィステック分布: $\theta \in \mathbb{R}$ に対して,

$$\mathsf{p}_\theta(x) = \frac{e^{\theta - x}}{(1 + e^{\theta - x})^2} \qquad (x \in \mathbb{R}).$$

10.2 M 推定量の一致性

各 $\theta \in \Theta$ に対して

$$\begin{aligned} R(\theta) &= \mathsf{P}\gamma_\theta = \int_{\mathcal{X}} \gamma_\theta(x)\, \mathrm{d}\mathsf{P}(x), \\ R_n(\theta) &= \widehat{\mathsf{P}}_n \gamma_\theta = \frac{1}{n} \sum_{i=1}^n \gamma_\theta(X_i) \end{aligned}$$

とする.

定義 10.2. θ^* はうまく分離されているとは, すべての $\eta > 0$ に対して

$$\inf\{R(\theta);\, d(\theta,\, \theta^*) > \eta\} > R(\theta^*)$$

が成り立つことである.

定理 10.3.

$$\sup_{\theta\in\Theta}|R(\theta)-R_n(\theta)| \xrightarrow{\mathrm{P}} 0 \qquad (n\to\infty) \tag{10.2}$$

を仮定する．すなわち, $\{\gamma_\theta;\, \theta\in\Theta\}$ は GC 族であることを仮定する．このとき

$$R(\widehat{\theta}_n) \xrightarrow{\mathrm{P}} R(\theta^*) \qquad (n\to\infty).$$

さらに, θ^* がうまく分離されていれば

$$\widehat{\theta}_n \xrightarrow{\mathrm{P}} \theta^* \qquad (n\to\infty).$$

Proof. θ^* と $\widehat{\theta}_n$ の定義

$$\sup_{\theta\in\Theta}|R_n(\theta)-R(\theta)| \xrightarrow{\mathrm{P}} 0 \qquad (n\to\infty).$$

と (10.2) に注意すれば

$$R(\widehat{\theta}_n) \xrightarrow{\mathrm{P}} R(\theta^*) \qquad (n\to\infty) \tag{10.3}$$

がわかる．これは

$$\begin{aligned}
0 &\le R(\widehat{\theta}_n)-R(\theta^*)\\
&= R(\widehat{\theta}_n)-R(\theta^*)-\{R_n(\widehat{\theta}_n)-R_n(\theta^*)\}+\{R_n(\widehat{\theta}_n)-R_n(\theta^*)\}\\
&\le R(\widehat{\theta}_n)-R(\theta^*)-\{R_n(\widehat{\theta}_n)-R_n(\theta^*)\} \qquad (\because R_n(\widehat{\theta}_n)-R_n(\theta^*)\le 0)\\
&\le |R_n(\widehat{\theta}_n)-R(\widehat{\theta}_n)|+|R_n(\theta^*)-R(\theta^*)|\\
&\le 2\sup_{\theta\in\Theta}|R_n(\theta)-R(\theta)| \xrightarrow{\mathrm{P}} 0 \qquad (n\to\infty)
\end{aligned}$$

からわかる．さらに, θ^* がうまく分離されていれば

$$\widehat{\theta}_n \xrightarrow{\mathrm{P}} \theta^* \qquad (n\to\infty).$$

なぜならば, 任意の $\epsilon>0$ に対して, 存在する $\eta>0$ があって

$$d(\theta,\,\theta^*)>\eta \text{ ならば } R(\theta)-R(\theta^*)>\epsilon$$

となる．これの対偶をとれば

$$R(\theta)-R(\theta^*)\le\epsilon \text{ ならば } d(\theta,\,\theta^*)\le\eta$$

となるので

$$\mathsf{Pr}\left(d(\widehat{\theta}_n,\,\theta^*)\le\eta\right)\ge \mathsf{Pr}\left(R(\widehat{\theta}_n)-R(\theta^*)\le\epsilon\right)\xrightarrow{n\to\infty} 1$$

からわかる． □

注意 10.4. 仮定 (10.2) はやや厳しい条件である．なぜならば, Θ はコンパクトであることを要求しているのに近い．

補題 10.5. 距離空間 (Θ, d) はコンパクト．写像

$$\theta \mapsto \gamma_\theta$$

は連続[1] とする．さらに,

$$G(x) = \sup_{\theta\in\Theta} |\gamma_\theta(x)|$$

とし, $\mathsf{P}G < \infty$ とする．このとき

$$\sup_{\theta\in\Theta} |R_n(\theta) - R(\theta)| \xrightarrow{\mathrm{P}} 0 \qquad (n \to \infty).$$

Proof. 補題 **??** よりわかる． □

補題 10.6. X の分布を P とし, $X_1, X_2, \ldots, X_n\,(n \geq 1)$ を X の独立複製とする．Θ はノルム $\|\cdot\|_d$ を持つ凸集合とし, 写像

$$\Theta \ni \theta \mapsto \gamma_\theta \in \mathbb{R}$$

は狭義凸とする．さらに,

$$\begin{aligned}
\theta^* &= \arg\min_{\theta\in\Theta} \mathsf{P}\gamma_\theta = \arg\min_{\theta\in\Theta} \int_{\mathcal{X}} \gamma_\theta(x)\,\mathrm{d}\mathsf{P}(x),\\
\widehat{\theta}_n &= \widehat{\theta}_n(X_1, X_2, \ldots, X_n) = \arg\min_{\theta\in\Theta} \widehat{\mathsf{P}}_n\gamma_\theta = \arg\min_{\theta\in\Theta} \left\{\frac{1}{n}\sum_{i=1}^n \gamma_\theta(X_i)\right\}
\end{aligned}$$

とする．ある $\epsilon > 0$ をとる．

- $\Theta_\epsilon = \Theta_\epsilon(\theta^*) := \{\theta \in \Theta;\ \|\theta - \theta^*\|_d \leq \epsilon\}$ はコンパクト,
- $x \in \mathcal{X}$ に対して, $G_\epsilon(x) := G_{\Theta_\epsilon}(x) := \sup_{\theta\in\Theta_\epsilon} |\gamma_\theta(x)|$ としたとき, $\mathsf{P}G_\epsilon = \int_{\mathbb{X}} G_\epsilon(x)\,\mathrm{d}\mathsf{P}(x) < \infty$,

を仮定する．このとき,

$$\widehat{\theta}_n \xrightarrow{\mathrm{P}} \theta^* \qquad (n \to \infty).$$

Proof. 補題 10.5 より

$$\sup_{\theta\in\Theta_\epsilon} |R_n(\theta) - R(\theta)| \xrightarrow{\mathrm{P}} 0 \qquad (n \to \infty).$$

[1]写像 $\theta \mapsto \gamma_\theta$ が $\theta = \theta^*$ で連続であるとは, 任意の $\epsilon > 0$ に対して, ある $\delta > 0$ が存在して, $|\theta - \theta^*| < \delta$ ならば, $\sqrt{\mathsf{P}(\gamma_\theta - \gamma_{\theta^*})^2} < \epsilon$ とであることである．

いま

$$\alpha := \frac{\epsilon}{\epsilon + \|\widehat{\theta}_n - \theta^*\|_d}, \qquad \tilde{\theta}_n := \alpha\widehat{\theta}_n + (1-\alpha)\theta^*$$

とする．Θ が凸集合であることから, $\tilde{\theta} \in \Theta$ であることに注意する．さらに

$$\begin{aligned}\|\tilde{\theta}_n - \theta^*\|_d &\le \alpha\|\widehat{\theta}_n - \theta^*\|_d + (1-\alpha)\|\widehat{\theta}_n - \theta^*\|_d \\ &= \|\widehat{\theta}_n - \theta^*\|_d = \epsilon.\end{aligned}$$

関数 $\Theta \ni \theta \mapsto \gamma_\theta \in \mathbb{R}$ は凸なので, 関数 $\Theta \ni \theta \mapsto \mathsf{P}\gamma_\theta \in \mathbb{R}$ も凸になることから,

$$\begin{aligned}R_n(\widehat{\theta}_n) &\le \alpha R_n(\widehat{\theta}_n) + (1-\alpha)R_n(\theta^*) \\ &\le \alpha R_n(\theta^*) + (1-\alpha)R_n(\theta^*) = R_n(\theta^*).\end{aligned}$$

すると

$$R(\tilde{\theta}_n) \xrightarrow{\mathrm{P}} R(\theta^*) \qquad (n \to \infty). \tag{10.4}$$

なぜならば

$$\begin{aligned}0 &\le R(\tilde{\theta}_n) - R(\theta^*) \\ &= R(\tilde{\theta}_n) - R(\theta^*) - \{R_n(\widehat{\theta}_n) - R_n(\theta^*)\} + \{R_n(\widehat{\theta}_n) - R_n(\theta^*)\} \\ &\le R(\tilde{\theta}_n) - R(\theta^*) - \{R_n(\widehat{\theta}_n) - R_n(\theta^*)\} \qquad (\because R_n(\widehat{\theta}_n) - R_n(\theta^*)\} < 0) \\ &\le |R_n(\tilde{\theta}_n) - R(\tilde{\theta}_n)| + |R_n(\theta^*) - R(\theta^*)| \\ &\le 2\sup_{\theta \in \Theta_\epsilon} |R_n(\theta) - R(\theta)| \xrightarrow{\mathrm{P}} 0 \qquad (n \to \infty)\end{aligned}$$

よりわかる．写像 $\Theta \ni \theta \mapsto \gamma_\theta \in [0, \infty)$ の狭義凸性と θ^* は一意に存在することに注意する, (8.6) より,

$$\|\tilde{\theta}_n - \theta^*\|_d \xrightarrow{\mathrm{P}} 0 \qquad (n \to \infty).$$

しかし, $0 \le \alpha \le 1$ なので,

$$\|\widehat{\theta}_n - \theta^*\| = \alpha\|\tilde{\theta}_n - \theta^*\|_d \xrightarrow{\mathrm{P}} 0 \qquad (n \to \infty).$$

□

例 10.7. X は実数値確率変数とし, $\Theta = \mathbb{R}$ とする．X の分布を P と書く．ある $r \ge 1$ に対して

$$\gamma_\theta(x) = |x - \theta|^r \qquad (x \in \mathbb{R}), \qquad \mathsf{E}[|X - \theta|^r] < \infty$$

とする. このとき, 任意の $\epsilon > 0$ に対して

$$G_\epsilon(x) := G_{\epsilon,\theta^*}(x) := \sup_{|\theta-\theta^*|\le\epsilon} |\gamma_\theta(x)| \le 2^{r-1}\{|x-\theta^*|^r + \epsilon^r\}$$

なので

$$G_\epsilon \in L^1(\mathsf{P}).$$

したがって, θ^* が一意ならば,

$$\widehat{\theta}_n \xrightarrow{\mathrm{P}} \theta^* \qquad (n\to\infty).$$

以上の議論を下記のように拡張できる. $p \in \mathbb{N}\,(p\ge 2)$ とする. $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$, $\boldsymbol{X}$ は $\mathbb{R}^p$ 値確率変数列ベクトル (共変量) で, Y は実数値確率変数 (応答変数) とする. $(\boldsymbol{X}, Y)$ の独立複製を考える. $r \ge 1$ とし, 損失関数を

$$\gamma_{\boldsymbol{\theta}}(\boldsymbol{x}, y) = |y - \boldsymbol{x}^\top\boldsymbol{\theta}|^r$$

とする. $\mathsf{E}[|Y - \boldsymbol{X}^\top\boldsymbol{\theta}|^r] < \infty$ であり

$$\boldsymbol{\theta}_0 = \arg\min_{\boldsymbol{\theta}} \mathsf{P}\gamma_{\boldsymbol{\theta}}, \qquad \widehat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}} \widehat{\mathsf{P}}_n\gamma_{\boldsymbol{\theta}},$$

とすれば,

$$\widehat{\boldsymbol{\theta}}_n \xrightarrow{\mathrm{P}} \boldsymbol{\theta}_0 \qquad (n\to\infty)$$

である. ただし, P_n は $(\boldsymbol{X}, Y)$ の独立複製 $(\boldsymbol{X}_i, Y_i)\,(i = 1, 2, \ldots, n)$

例 10.8. X は閉区間 $[0, 1]$ を値域にもつ確率変数で, Y は実数値確率変数で

$$Y = \theta^*(X) + \xi$$

とする. ただし, $\theta^* : [0, 1] \longrightarrow \mathbb{R}$ は可測関数で, ξ は $N(0, \sigma^2)$ に従うとする. ここで, $\sigma^2\,(\sigma > 0)$ は未知とする. 可測関数 $\theta : [0, 1] \longrightarrow \mathbb{R}$ とし, 損失関数

$$\gamma_\theta(x, y) = (y - \theta(x))^2 \qquad (x \in [0, 1], y \in \mathbb{R})$$

を考える.

$$\Theta = \Big\{\theta : [0, 1] \longrightarrow \mathbb{R} \quad ; \quad \theta \text{ は } m \text{ 階連続微分可能}$$
$$\text{（その } m \text{ 階導関数を } \theta^{(m)} \text{ とかく）で} \quad \int_0^1 |\theta^{(m)}(x)|^2\,dx < 1\Big\}$$

とし, $\theta^* \in \Theta$ を固定する. 集合 Θ は一様ノルム

$$\|\theta\|_\infty := \sup_{x\in[0,1]} |\theta(x)|$$

をもつとする. このとき, 任意の $\epsilon > 0$ に対して,

$$\Theta_\epsilon := \Theta_{\epsilon,\theta^*} := \left\{\theta\colon [0,1] \longrightarrow \mathbb{R}, \|\theta - \theta^*\|_\infty < \epsilon, \int_0^1 |\theta^{(m)}(x)|^2\,\mathrm{d}x \le 1\right\}$$

はコンパクトであることを示すことができる. θ^* が一意的に存在（たとえば, X の密度が, 零にならない測度に関して絶対連続であればよい）すれば, 最小 2 乗推定量 $\widehat{\theta}_n$ は一様ノルムに関して一致性をもつこと知られている.

10.3 漸近正規性

この章の以下の節では, $\Theta \subset \mathbb{R}^p$ とする.

定義 10.9. 母数 $\boldsymbol{\theta}_0$ の推定量列 $\{\widehat{\boldsymbol{\theta}}_n\}_{n=1}^\infty$ が漸近線形であるとは, 以下のように表現できるときをいう:

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \sqrt{n}\widehat{\mathsf{P}}_n\underline{\boldsymbol{\ell}} + o_P(1).$$

ただし

$$\underline{\boldsymbol{\ell}(x)} = \begin{pmatrix} \ell_1(x) \\ \ell_2(x) \\ \vdots \\ \ell_p(x) \end{pmatrix} : \mathcal{X} \longrightarrow \mathbb{R}^p$$

は

$$\mathsf{P}\underline{\boldsymbol{\ell}} = \begin{pmatrix} \mathsf{P}\ell_1 \\ \mathsf{P}\ell_2 \\ \vdots \\ \mathsf{P}\ell_p \end{pmatrix} = \begin{pmatrix} \int_{\mathcal{X}} \ell_1(x)\,\mathrm{d}\mathsf{P}(x) \\ \int_{\mathcal{X}} \ell_2(x)\,\mathrm{d}\mathsf{P}(x) \\ \vdots \\ \int_{\mathcal{X}} \ell_p(x)\,\mathrm{d}\mathsf{P}(x) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} =: \mathbf{0}$$

と $\mathsf{P}\ell_k^2 = \int_{\mathbb{M}} \ell_k^2(x)\,\mathrm{d}\mathsf{P}(x) < \infty\,(k = 1, 2, \ldots, p)$ をみたす. 関数 $\underline{\boldsymbol{\ell}}$ を影響関数という. $p = 1$ のとき,

$$\sigma^2 := \mathsf{P}\ell^2$$

を漸近分散という.

定義 10.10. $p = 1$ とする. 母数 θ^* の 2 つの推定量列 $\{\widehat{\theta}_{n,1}\}_{n=1}^\infty$ と $\{\widehat{\theta}_{n,2}\}_{n=1}^\infty$ が漸近線形であり, 漸近分散が σ_1^2 と σ_2^2 とする. このとき,

$$e_{1,2} = \frac{\sigma_2^2}{\sigma_1^2}$$

を $\widehat{\theta}_{n,2}$ と比較した $\widehat{\theta}_{n,1}$ の漸近相対効率という.

10.4 M 推定量の漸近正規性のための強い十分条件

以下の条件は, 漸近正規性が成立するための十分条件である. これが, 十分条件であることを証明するのは易しいが, 厳しい条件である. 次章では, よりゆるい十分条件を議論する.

- <u>条件 a</u>: $\epsilon > 0$ が存在して, $|\boldsymbol{\theta} - \boldsymbol{\theta}^*| < \epsilon$ なる θ に対して, 写像

$$\boldsymbol{\theta} \mapsto \gamma_{\boldsymbol{\theta}}(x)$$

はすべての x に対して, 微分可能で, 導関数

$$\boldsymbol{\psi}_{\boldsymbol{\theta}}(x) := \frac{\partial}{\partial \boldsymbol{\theta}} \gamma_{\boldsymbol{\theta}}(x) = \left(\frac{\partial}{\partial \theta_1} \gamma_{\boldsymbol{\theta}}(x), \ldots, \frac{\partial}{\partial \theta_p} \gamma_{\boldsymbol{\theta}}(x) \right)^\top \qquad (x \in \mathcal{X})$$

をもつ. ここで, $\mathbb{R}^p$ の Euclid ノルムを $|\cdot|$ と書いた.

- <u>条件 b</u>: $\boldsymbol{\theta} \to \boldsymbol{\theta}^*$ のとき,

$$P(\boldsymbol{\psi}_{\boldsymbol{\theta}} - \boldsymbol{\psi}_{\boldsymbol{\theta}_0}) = V(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o(1)|\boldsymbol{\theta} - \boldsymbol{\theta}_0|.$$

ただし, V は $p \times p$ の正定値行列である.

- <u>条件 c</u>: ある $\epsilon > 0$ が存在して, 族

$$\mathcal{G} =: \{\boldsymbol{\psi}_{\boldsymbol{\theta}} : |\boldsymbol{\theta} - \boldsymbol{\theta}^*| < \epsilon\}$$

を考える. $\sqrt{n}(\widehat{\mathsf{P}}_n - \mathsf{P})$ は $\mathcal{G}$ 上で漸近連続である. さらに,

$$\begin{aligned} \lim_{\boldsymbol{\theta} \to \boldsymbol{\theta}^*} \|\boldsymbol{\psi}_{\boldsymbol{\theta}_0} - \boldsymbol{\psi}_{\boldsymbol{\theta}}\|_{L^2(\mathsf{P})} &= \lim_{\boldsymbol{\theta} \to \boldsymbol{\theta}_0} \sqrt{\mathsf{P}\{|\boldsymbol{\psi}_{\boldsymbol{\theta}_0} - \boldsymbol{\psi}_{\boldsymbol{\theta}}|^2\}} \\ &= \lim_{\boldsymbol{\theta} \to \boldsymbol{\theta}_0} \sqrt{\mathsf{P}\{(\boldsymbol{\psi}_{\boldsymbol{\theta}_0} - \boldsymbol{\psi}_{\boldsymbol{\theta}})^\top (\boldsymbol{\psi}_{\boldsymbol{\theta}_0} - \boldsymbol{\psi}_{\boldsymbol{\theta}})\}} = 0 \end{aligned}$$

である.

補題 10.11. 条件 a, b, c を仮定する. このとき, $\widehat{\boldsymbol{\theta}}_n$ は漸近線形で, 影響関数

$$\underline{\boldsymbol{\ell}}(x) = -V^{-1} \boldsymbol{\psi}_{\boldsymbol{\theta}^*}(x)$$

をもつ. したがって

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*) \rightsquigarrow \mathsf{N}_p(\mathbf{0}, V^{-1} J V^{-1}).$$

ただし,

$$J := \mathsf{P}[\boldsymbol{\psi}_{\boldsymbol{\theta}^*} \boldsymbol{\psi}_{\boldsymbol{\theta}^*}^\top].$$

Proof. $\boldsymbol{\theta}^*$ は Θ の内点であり

$$\Theta \ni \boldsymbol{\theta} \mapsto \mathsf{P}\gamma_{\boldsymbol{\theta}} \in [0,\, \infty)$$

を最小化する. したがって,

$$\mathsf{P}\psi_{\boldsymbol{\theta}^*} = \mathbf{0}.$$

推定量 $\widehat{\boldsymbol{\theta}}_n$ は一致性を持つので, スコア関数の解としてよい:

$$\widehat{\mathsf{P}}_n \psi_{\widehat{\boldsymbol{\theta}}_n} = \mathbf{0}.$$

このスコア関数を書きかえると

$$\begin{aligned}
\mathbf{0} &= \widehat{\mathsf{P}}_n \psi_{\widehat{\boldsymbol{\theta}}_n} = \widehat{\mathsf{P}}_n(\psi_{\widehat{\boldsymbol{\theta}}_n} - \psi_{\boldsymbol{\theta}^*}) + \widehat{\mathsf{P}}_n \psi_{\boldsymbol{\theta}^*} \\
&= (\widehat{\mathsf{P}}_n - \mathsf{P})(\psi_{\widehat{\boldsymbol{\theta}}_n} - \psi_{\boldsymbol{\theta}^*}) + \mathsf{P}\psi_{\widehat{\boldsymbol{\theta}}_n} - \mathsf{P}\psi_{\boldsymbol{\theta}^*} + \widehat{\mathsf{P}}_n \psi_{\boldsymbol{\theta}^*} \\
&= (\widehat{\mathsf{P}}_n - \mathsf{P})(\psi_{\widehat{\boldsymbol{\theta}}_n} - \psi_{\boldsymbol{\theta}^*}) + \mathsf{P}\psi_{\widehat{\boldsymbol{\theta}}_n} + \widehat{\mathsf{P}}_n \psi_{\boldsymbol{\theta}^*}. \qquad (\because \mathsf{P}\psi_{\boldsymbol{\theta}^*} = \mathbf{0})
\end{aligned}$$

いま, 条件 b と $\{\psi_{\boldsymbol{\theta}} : |\boldsymbol{\theta} - \boldsymbol{\theta}^*| \le \epsilon\}$ の漸近連続性より

$$\begin{aligned}
\mathbf{0} &= (\widehat{\mathsf{P}}_n - \mathsf{P})(\psi_{\widehat{\boldsymbol{\theta}}_n} - \psi_{\boldsymbol{\theta}^*}) + P(\psi_{\widehat{\boldsymbol{\theta}}_n} - \psi_{\boldsymbol{\theta}^*}) + \widehat{\mathsf{P}}_n \psi_{\widehat{\boldsymbol{\theta}}_n} \\
&= o_P(n^{-1/2}) + V(\psi_{\widehat{\boldsymbol{\theta}}_n} - \psi_{\boldsymbol{\theta}^*}) + o(|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*|) + \widehat{\mathsf{P}}_n \psi_{\boldsymbol{\theta}^*}
\end{aligned}$$

となる. よって,

$$\begin{aligned}
o_P(1) &= V\big(\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*)\big) + o\big(\sqrt{n}|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*|\big) + \sqrt{n}\widehat{\mathsf{P}}_n \psi_{\boldsymbol{\theta}^*} \\
&= o_P(1) + V\big(\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*)\big) + o\big(\sqrt{n}|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*|\big) + O_p(1)
\end{aligned}$$

となるので,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*) = O_P(1).$$

したがって,

$$\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^* = -V\big\{\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*)\big\} + o_P(n^{-1/2})$$

がわかる. 中心極限定理と $\mathsf{P}\boldsymbol{\psi}_{\boldsymbol{\theta}^*} = \mathbf{0}$ より

$$\sqrt{n}\widehat{\mathsf{P}}_n \psi_{\boldsymbol{\theta}^*} \rightsquigarrow \mathsf{N}_p\Bigg(\mathbf{0},\, \mathsf{P}\big(\psi_{\boldsymbol{\theta}^*} \psi_{\boldsymbol{\theta}^*}^\top\big)\Bigg)$$

なので

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*) \rightsquigarrow \mathsf{N}_p\big(\mathbf{0},\, V^{-1} J V^{-1}\big).$$

□

例 10.12 (Huber 型推定量の漸近分布)**.** $\mathbb{M} = \mathbb{R}$, $\Theta = \mathbb{R}$ とし, $0 < k < \infty$ はある固定された定数（統計学者が選択するもの）とする. Huber 型損失関数

$$
\begin{aligned}
\gamma_\theta &= \gamma(x-\theta), \\
\gamma(x) &= x^2 \mathbb{1}\{|x| \le k\} + (2k|x| - k^2)\mathbb{1}\{|x| > k\}, \qquad (x \in \mathbb{R})
\end{aligned}
$$

に対応する推定量を Huber 型推定量とよぶ.

これが, 条件 a, b, c をみたすことを確認する.

条件 a:

$$
\psi_\theta(x) = \begin{cases} 2k & (x-\theta < -k), \\ -2(x-\theta) & (|x-\theta| \le k), \\ -2k & (x-\theta > k). \end{cases}
$$

条件 b:

$$
\frac{\mathrm{d}}{\mathrm{d}\theta} \int \psi_\theta(x)\,\mathrm{d}\mathsf{F}(x) = 2\{\mathsf{F}(k+\theta) - \mathsf{F}(-k+\theta)\}.
$$

ただし, $\mathsf{F}(t) = \mathsf{Pr}(X \le t)\,(t \in \mathbb{R})$ は X の分布関数である. なぜならば, $|\psi_\theta(x)| \le k$ なので, 定理 **??** を適用して

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\theta} \int \psi_\theta(x)\,dF(x) &= \int \frac{\mathrm{d}}{\mathrm{d}\theta}\psi_\theta(x)\,\mathrm{d}\mathsf{F}(x) = 2\int_{-k+\theta}^{k+\theta} \mathrm{d}\mathsf{F}(x) \\
&= 2\{\mathsf{F}(k+\theta) - \mathsf{F}(-k+\theta)\}
\end{aligned}
$$

よりわかる. したがって,

$$
V = 2\{\mathsf{F}(k+\theta_0) - \mathsf{F}(-k+\theta_0)\}.
$$

条件 c: 明らかに $\{\psi_\theta;\, \theta \in \mathbb{R}\}$ は VC グラフ族で, 包絡関数 $\Psi = \sup_{x\in\mathbb{R}} |\psi_\theta(x)|$ は $\Psi \le 2k$ となり, 有界.

以上より, 漸近正規性は補題 10.11 より従う.

例 10.13. 確率変数 X の分布関数 F は中央値 θ^* で微分可能とし, 中央値の漸近理論を述べる. $X_1, X_2, \ldots, X_n$ は X の独立複製とし,

$$
\widehat{\mathsf{F}}_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{(-\infty,\,x]}(X_i) \qquad (x \in \mathbb{R})
$$

とする. 標本中央値は Huber 型推定量の極限の場合（$k \to \infty$）とみなすことができる. しかし,

$$
\gamma_\theta(x) = |x-\theta|
$$

は $x = \theta$ で微分できないので, 条件 a をみたさない. そのため, 少し工夫が必要となる.

簡単のために, 標本数は奇数とする. 標本中央値 $\widehat{\theta}_n$ は, スコア方程式

$$\widehat{\mathsf{F}}_n(\widehat{\theta}_n) - \frac{1}{2} = 0$$

をみたす. このとき,

$$\begin{aligned} 0 &= \widehat{\mathsf{F}}_n(\widehat{\theta}_n) - \mathsf{F}(\theta^*) \\ &= [\widehat{\mathsf{F}}_n(\widehat{\theta}_n) - \mathsf{F}(\widehat{\theta}_n)] + [\mathsf{F}(\widehat{\theta}_n) - \mathsf{F}(\theta^*)] \\ &= \frac{1}{\sqrt{n}} W_n(\widehat{\theta}_n) + [\mathsf{F}(\widehat{\theta}_n) - \mathsf{F}(\theta^*)] \end{aligned}$$

をえる. ただし,

$$W_n(x) = \sqrt{n}(\widehat{\mathsf{F}}_n(x) - \mathsf{F}(x))$$

は経験過程である. 分布関数 F は θ^* で連続であり, $\mathsf{E}[|X-\theta|] < \infty$ のとき, 例 **??** より,

$$\widehat{\theta}_n \xrightarrow{\mathrm{P}} \theta^* \qquad (n \to \infty)$$

となる. 半閉区間の集合族 $\{(-\infty, r]; r \in \mathbb{R}\}$ は VC 族なので, 経験過程 W_n は漸近連続性をもつので,

$$W_n(\widehat{\theta}_n) = W_n(\theta^*) + o_P(1)$$

となる. したがって,

$$\begin{aligned} 0 &= W_n(\widehat{\theta}_n) + \sqrt{n}(\mathsf{F}(\widehat{\theta}_n) - \mathsf{F}(\theta^*)) \\ &= W_n(\theta_\theta) + \sqrt{n}\{\mathsf{p}(\theta^*) + o(1)\}(\widehat{\theta}_n - \theta^*) + o_P(1) \end{aligned}$$

である. ただし,

$$\mathsf{p}(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta}\mathsf{F}(\theta).$$

これより

$$\sqrt{n}(\widehat{\theta}_n - \theta^*) = -\frac{W_n(\theta^*)}{\mathsf{p}(\theta^*)} + o_P(1).$$

よって

$$\sqrt{n}(\widehat{\theta}_n - \theta^*) \rightsquigarrow \mathsf{N}(0, \sigma^2)$$

となる. ただし, 漸近分散は

$$\sigma^2 = \frac{1}{\{\mathsf{p}(\theta^*)\}^2}\mathbb{V}[W_n(\theta^*)] = \frac{1}{4\{\mathsf{p}(\theta^*)\}^2}$$

となる. なぜならば

$$\begin{aligned} \mathsf{Var}[W_n(\theta^*)] &= \mathsf{Var}\left[\sqrt{n}\frac{1}{n}\sum_{i=1}^{n}(\mathbb{1}_{(-\infty,\theta^*]}(X_i) - \mathsf{F}(\theta^*))\right] \\ &= \mathsf{Var}[\mathbb{1}_{(-\infty,\theta^*]}(X_1)] = \mathsf{F}(\theta^*)\{1 - \mathsf{F}(\theta^*)\} \\ &= \mathsf{E}[\{\mathbb{1}_{(-\infty,\theta^*]}(X_1)\}^2] - \{\mathsf{E}[\mathbb{1}_{(-\infty,\theta^*]}(X_1)]\}^2 \\ &= \mathsf{F}(\theta^*) - \{\mathsf{F}(\theta^*)\}^2 = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4} \end{aligned}$$

よりわかる.

10.5 M 推定量の漸近正規性のためのより弱い十分条件

損失関数 $\gamma_{\boldsymbol{\theta}}$ の微分可能性についての条件をゆるめる.

- 条件 aa: 関数

$$\boldsymbol{\psi}_{\boldsymbol{\theta}^*,i} : \mathbb{M} \longrightarrow \mathbb{R} \qquad (i = 1, 2, \ldots, p),$$
$$\boldsymbol{\psi}_{\boldsymbol{\theta}^*} = (\boldsymbol{\psi}_{\boldsymbol{\theta}^*,1}, \boldsymbol{\psi}_{\boldsymbol{\theta}^*,2}, \ldots, \boldsymbol{\psi}_{\boldsymbol{\theta}^*,p})^\top$$

  は $\mathsf{P}\boldsymbol{\psi}^2_{\theta^*,i} < \infty$ で

$$\lim_{\boldsymbol{\theta}\to\boldsymbol{\theta}^*} \frac{\|\gamma_{\boldsymbol{\theta}} - \gamma_{\boldsymbol{\theta}^*} - (\boldsymbol{\theta}-\boldsymbol{\theta}^*)^\top \boldsymbol{\psi}_{\boldsymbol{\theta}^*}\|_{L^2(\mathsf{P})}}{|\boldsymbol{\theta}-\boldsymbol{\theta}^*|} \longrightarrow 0$$

  が成立する.

- 条件 bb: $\boldsymbol{\theta} \to \boldsymbol{\theta}^*$ のとき,

$$\mathsf{P}(\gamma_{\boldsymbol{\theta}} - \gamma_{\boldsymbol{\theta}^*}) = \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}^*)^\top V(\boldsymbol{\theta}-\boldsymbol{\theta}^*) + o(1)|\boldsymbol{\theta}-\boldsymbol{\theta}^*|^2.$$

  ただし, V は $p \times p$ の正値対称行列である.

- 条件 cc:

$$g_{\boldsymbol{\theta}}(x) = \begin{cases} \dfrac{\gamma_{\boldsymbol{\theta}}(x) - \gamma_{\boldsymbol{\theta}^*}(x) - (\boldsymbol{\theta}-\boldsymbol{\theta}^*)^\top \boldsymbol{\psi}_{\boldsymbol{\theta}^*}(x)}{|\boldsymbol{\theta}-\boldsymbol{\theta}^*|} & (\boldsymbol{\theta} \neq \boldsymbol{\theta}^*) \\ 0 & (\boldsymbol{\theta} = \boldsymbol{\theta}^*) \end{cases}$$

  と定義する. ある $\epsilon > 0$ が存在して, 関数族

$$\mathcal{G} := \{g_{\boldsymbol{\theta}};\, 0 < |\boldsymbol{\theta}-\boldsymbol{\theta}^*| < \epsilon\}$$

  を考える. $\sqrt{n}(\widehat{\mathsf{P}}_n - \mathsf{P})$ は $\mathcal{G}$ 上で漸近連続とする.

補題 10.14. 条件 aa, bb, cc が成立すると仮定する. このとき, $\widehat{\boldsymbol{\theta}}_n$ は影響関数

$$\boldsymbol{\ell}(x) = -V^{-1}\boldsymbol{\psi}_{\boldsymbol{\theta}^*}$$

をもつ. したがって,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*) \rightsquigarrow \mathsf{N}_p(\mathbf{0}, V^{-1}JV^{-1})$$

が成立する. ただし, $J = \mathsf{P}(\boldsymbol{\psi}_{\boldsymbol{\theta}^*}\boldsymbol{\psi}_{\boldsymbol{\theta}^*}^\top)$ である.

Proof. $\epsilon > 0$ に対して, $\{g_{\boldsymbol{\theta}} : |\boldsymbol{\theta} - \boldsymbol{\theta}^*| \le \epsilon\}$ は漸近連続なので, $\boldsymbol{\theta} \to \boldsymbol{\theta}^*$ のとき,

$$\widehat{\mathsf{P}}_n(\gamma_{\boldsymbol{\theta}} - \gamma_{\boldsymbol{\theta}^*}) = \frac{1}{2}\Big| V^{1/2}(\boldsymbol{\theta} - \boldsymbol{\theta}^*) + o\big(|\boldsymbol{\theta} - \boldsymbol{\theta}^*|\big) + o_P(n^{-1/2})\Big|^2. \quad (10.5)$$

を得る. これは,

$$\begin{aligned}
\widehat{\mathsf{P}}_n(\gamma_{\boldsymbol{\theta}} - \gamma_{\boldsymbol{\theta}^*}) &= (\widehat{\mathsf{P}}_n - \mathsf{P})(\gamma_{\boldsymbol{\theta}} - \gamma_{\boldsymbol{\theta}^*}) + \mathsf{P}(\gamma_{\boldsymbol{\theta}} - \gamma_{\boldsymbol{\theta}^*}) \\
&= (\widehat{\mathsf{P}}_n - \mathsf{P})\big\{g_{\boldsymbol{\theta}}|\boldsymbol{\theta} - \boldsymbol{\theta}^*| + (\boldsymbol{\theta} - \boldsymbol{\theta}^*)^\top \boldsymbol{\psi}_{\boldsymbol{\theta}^*}\big\} + \mathsf{P}(\gamma_{\boldsymbol{\theta}} - \gamma_{\boldsymbol{\theta}^*}) \\
&= (\widehat{\mathsf{P}}_n - \mathsf{P})g_{\boldsymbol{\theta}}|\boldsymbol{\theta} - \boldsymbol{\theta}^*| + (\boldsymbol{\theta} - \boldsymbol{\theta}^*)^\top \widehat{\mathsf{P}}_n \boldsymbol{\psi}_{\boldsymbol{\theta}^*} - (\boldsymbol{\theta} - \boldsymbol{\theta}^*)^\top \mathsf{P} \boldsymbol{\psi}_{\boldsymbol{\theta}^*} \\
&\qquad + \mathsf{P}(\gamma_{\boldsymbol{\theta}} - \gamma_{\boldsymbol{\theta}^*}) \\
&= o_P(n^{-1/2}|\boldsymbol{\theta} - \boldsymbol{\theta}^*| + (\boldsymbol{\theta} - \boldsymbol{\theta}^*)^\top \widehat{\mathsf{P}}_n \boldsymbol{\psi}_{\boldsymbol{\theta}^*} \\
&\qquad + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)^\top V(\boldsymbol{\theta} - \boldsymbol{\theta}^*) + o\big(|\boldsymbol{\theta} - \boldsymbol{\theta}^*|^2\big) \\
&= \frac{1}{2}\Big| V^{1/2}(\boldsymbol{\theta} - \boldsymbol{\theta}^*) + o\big(|\boldsymbol{\theta} - \boldsymbol{\theta}^*|\big) + o_P(n^{-1/2})\Big|^2
\end{aligned}$$

からわかる.

$$\widehat{\mathsf{P}}_n(\gamma_{\widehat{\boldsymbol{\theta}}_n} - \gamma_{\boldsymbol{\theta}^*}) \le 0$$

なので, $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_n$ に対して, (10.2) を用いれば,

$$|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*| = O_p(n^{-1/2}).$$

さらに, $\tilde{\boldsymbol{\theta}}_n := \boldsymbol{\theta}^* - V^{-1/2}\widehat{\mathsf{P}}_n \boldsymbol{\psi}_{\boldsymbol{\theta}^*}$ に対して, (10.2) を用いれば,

$$\begin{aligned}
\widehat{\mathsf{P}}_n(\gamma_{\tilde{\boldsymbol{\theta}}_n} - \gamma_{\boldsymbol{\theta}^*}) &= o_P(n^{-1/2})|V^{-1/2}P_n \boldsymbol{\psi}_{\boldsymbol{\theta}^*}| - (\widehat{\mathsf{P}}_n \boldsymbol{\psi}_0)^\top V^{-1}(\widehat{\mathsf{P}}_n \boldsymbol{\psi}_{\boldsymbol{\theta}^*}) \\
&\qquad + \frac{1}{2}(\widehat{\mathsf{P}}_n \boldsymbol{\psi}_0)^\top V^{-1}(P_n \boldsymbol{\psi}_0) + o(|V^{-1/2}\widehat{\mathsf{P}}_n \boldsymbol{\psi}_{\boldsymbol{\theta}^*}|) \\
&= -\frac{1}{2}\big|V^{-1/2}P_n \boldsymbol{\psi}_{\boldsymbol{\theta}^*}\big|^2 + o_P(n^{-1})
\end{aligned}$$

ここで

$$\widehat{\mathsf{P}}_n(\gamma_{\widehat{\boldsymbol{\theta}}_n} - \gamma_{\boldsymbol{\theta}^*}) \le \widehat{\mathsf{P}}_n(\gamma_{\tilde{\boldsymbol{\theta}}_n} - \gamma_{\boldsymbol{\theta}^*})$$

に注意すれば,

$$\begin{aligned}
&O_P(n^{-1/2})|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*| + (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*)^\top \widehat{\mathsf{P}}_n \boldsymbol{\psi}_{\boldsymbol{\theta}^*} + \frac{1}{2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*)^\top V(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*)^\top \\
&\qquad + o(|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*|^2) \\
&\qquad \le -\frac{1}{2}\big|V^{-1/2}\widehat{\mathsf{P}}_n \boldsymbol{\psi}_{\boldsymbol{\theta}^*}\big|^2 + o_P(n^{-1/2})
\end{aligned}$$

より

$$\Big| V^{1/2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*)\Big|^2 = o_P(n^{-1}). \quad (10.6)$$

これは

$$
\begin{aligned}
&o_P(n^{1/2})O_P(n^{-1/2}) + (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*)^\top \widehat{\mathsf{P}}_n \boldsymbol{\psi}_{\boldsymbol{\theta}^*} \\
&\qquad + \frac{1}{2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*)^\top V(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*) + o_P(n^{-1}) \le -\frac{1}{2}\left|V^{-1/2}P_n\boldsymbol{\psi}_{\boldsymbol{\theta}^*}\right|^2 + o_P(n^{-1}) \\
&\Leftrightarrow o_P(n^{-1}) + (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*)^\top \widehat{\mathsf{P}}_n \boldsymbol{\psi}_{\boldsymbol{\theta}^*} + \frac{1}{2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*)^\top V(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*) \\
&\qquad\qquad + o_P(n^{-1}) \le -\frac{1}{2}\left|V^{-1/2}P_n\boldsymbol{\psi}_{\boldsymbol{\theta}^*}\right|^2 + o_P(n^{-1}) \\
&\Leftrightarrow o_P(n^{-1}) + (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*)^\top \widehat{\mathsf{P}}_n \boldsymbol{\psi}_{\boldsymbol{\theta}^*} + \frac{1}{2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*)^\top V(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*) \\
&\qquad\qquad + o_P(n^{-1}) \le -\frac{1}{2}\left|V^{-1/2}P_n\boldsymbol{\psi}_{\boldsymbol{\theta}^*}\right|^2 + o_P(n^{-1}) \\
&\Leftrightarrow (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*)^\top \widehat{\mathsf{P}}_n \boldsymbol{\psi}_{\boldsymbol{\theta}^*} + \frac{1}{2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*)^\top V(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*) \\
&\qquad\qquad + o_P(n^{-1}) \le -\frac{1}{2}\left|V^{-1/2}\widehat{\mathsf{P}}_n\boldsymbol{\psi}_{\boldsymbol{\theta}^*}\right|^2 + o_P(n^{-1}) \\
&\Leftrightarrow \left|V^{1/2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*)\right|^2 = o_P(n^{-1})
\end{aligned}
$$

よりわかる. したがって, (10.6) より

$$
V^{1/2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*) = -V^{-1/2}\widehat{\mathsf{P}}_n\boldsymbol{\psi}_{\boldsymbol{\theta}^*} + o_P(n^{-1/2})
$$

なので

$$
\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^* = -V^{-1}\widehat{\mathsf{P}}_n\boldsymbol{\psi}_{\boldsymbol{\theta}^*} + o_P(n^{-1/2})
$$

より補題は示された. □