

M-推定量の漸近正規性について考えて前回から議論してきた。設定は以下のものであった。確率変数  $X_1, \dots, X_n$  が確率分布  $P$  に従うとする。次のような関数を考える。

$$\Psi_n(\theta) \triangleq \frac{1}{n} \sum_{i=1}^n \psi_\theta(X_i) = \mathbb{P}_n \psi_\theta$$

$$\Psi(\theta) = P\psi_\theta = \mathbb{E}\psi_\theta(X)$$

いま、推定量の列  $\hat{\theta}_n$  は  $\Psi_n(\theta) = 0$  の解で、 $\Psi(\theta) = 0$  の解  $\theta_0$  に確率収束すると仮定する。

このときに漸近正規性、すなわち、 $\sqrt{n}(\hat{\theta}_n - \theta_0)$  が正規分布に収束する条件として、次のような定理を得た。

**Theorem 5.21.** ユークリッド空間の開部分集合  $\Theta$  でノルム  $\|\cdot\|$  が定義されているとする。以下を満たすベクトル値可測関数  $x \mapsto \psi_\theta(x), \theta \in \Theta$  を定義する。

1 可測関数  $\dot{\psi}, P\dot{\psi}^2 < \infty$  が存在し、

$$\|\psi_{\theta_1}(x) - \psi_{\theta_2}(x)\| \leq \dot{\psi}(x)\|\theta_1 - \theta_2\|.$$

2  $P\|\psi_{\theta_0}\|^2 < \infty$ .

3  $\theta \mapsto P\psi_\theta$  が  $\theta_0$  において微分可能であり、 $\theta_0$  におけるヤコビアン  $V_{\theta_0}^{-1}$  は非特異である。

このとき、 $\mathbb{P}_n \psi_{\hat{\theta}_n} = o_P(n^{-1/2})$  かつ  $\hat{\theta}_n \xrightarrow{P} \theta_0$  ならば、次式が成り立つ。

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -V_{\theta_0}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{\theta_0}(X_i) + o_P(1)$$

上の定理は M-推定量の漸近正規性について、 $\psi_\theta$  のリプシッツ連続性を条件においている。しかし、たとえばメジアン の推定で用いられる。  $\psi_\theta(x) = \text{sgn}(x - \theta)$  はリプシッツ連続でなく上の定理では漸近正規性を示すことはできない。この条件を弱めることを考えていく。

まず M-推定量では、方程式の解を解く問題と、関数の最大化の問題を同一視できることを思い出す。すなわち最大化すべき、規準関数として、 $\theta \mapsto Pm_\theta$  を考える。この関数が  $\theta_0$  において2階連続微分可能であれば、 $\theta_0$  回りのテーラー展開により、

$$Pm_\theta = Pm_{\theta_0} + \frac{1}{2}(\theta - \theta_0)^T V_{\theta_0}(\theta - \theta_0) + o(\|\theta - \theta_0\|^2).$$

ただし、 $V_{\theta_0} = \frac{\partial^2}{\partial \theta_0^2} Pm_\theta$  であり、 $\frac{\partial}{\partial \theta_0} Pm_\theta = 0$  となることを用いた。この展開により、次の定理を示すことができる。

**Theorem 5.23.** ユークリッド空間の開部分集合  $\Theta$  でノルム  $\|\cdot\|$  が定義されているとする。以下を満たす実数値可測関数  $x \mapsto m_\theta(x), \theta \in \Theta$  を定義する。

1  $\theta \mapsto m_\theta(x)$  は  $\theta_0$  で微分可能 ( $P$ -a.s  $x$ ) であり、その微分は  $\dot{m}_{\theta_0}(x)$ 。

2  $\theta_0$  の近傍の全ての  $\theta_1, \theta_2$  と可測関数  $\dot{m}$  ( $P\dot{m}^2 < \infty$ ) に対して、

$$|m_{\theta_1}(x) - m_{\theta_2}(x)| \leq \dot{m}(x)\|\theta_1 - \theta_2\|.$$

3  $\theta \mapsto Pm_\theta$  は  $\theta_0$  回りで非特異な対称ヤコビ行列  $V_{\theta_0}$  を用いて2次のテイラー展開可能。

このとき、 $\mathbb{P}_n m_{\hat{\theta}_n} \geq \sup_\theta \mathbb{P}_n m_\theta - o_p(n^{-1})$  と  $\hat{\theta}_n \xrightarrow{P} \theta_0$  が成り立つとすれば、

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -V_{\theta_0}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{m}_{\theta_0}(X_i) + o_p(1)$$

が成り立つ。特に、

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow N(0, V_{\theta_0}^{-1} P\dot{m}_{\theta_0}\dot{m}_{\theta_0}^T V_{\theta_0}^{-1}).$$

**Example 5.24.** (Median) 標本  $X_1, \dots, X_n$  に対する標本メジアンは  $\theta \mapsto -\sum_{i=1}^n |X_i - \theta|$  を最大にする (6/30 寺田君の資料を参照). サンプルの分布関数  $F$  はそのメジアン  $\theta_0$  で微分可能であり導関数  $f(\theta_0) > 0$  とする. このとき, 標本メジアンは漸近正規的である.

*Proof.*  $m_\theta(x) = |x - \theta| - |x|$  とする. 三角不等式

$$\begin{aligned} |m_{\theta_1}(x) - m_{\theta_2}(x)| &= ||x - \theta_1| - |x| - (|x - \theta_2| - |x|)| \\ &= ||x - \theta_1| - |x - \theta_2|| \\ &\leq |x - \theta_1 - (x - \theta_2)| \\ &= |\theta_1 - \theta_2|. \end{aligned}$$

さらに,  $Pm^2 = \mathbb{E}1 = 1 < \infty$  である. よって,  $m(x) = 1$  として, Theorem 5.23 の条件 2 が成り立つ.

次に, 条件 1 をチェックする.  $\theta \mapsto m_\theta(x)$  は  $x = \theta_0$  のときを除いて  $\theta$  微分可能であり, その導関数は  $m_{\theta_0}(x) = -\text{sgn}(x - \theta_0)$  となる. (補足図参照). よって条件 1 も満たされる.

最後に条件 3 をチェックする. まず

$$Pm_\theta = \theta F(0) + \int_{(0, \theta]} (\theta - 2x) dF(x) - \theta(1 - F(\theta)) = 2 \int_0^\theta F(x) dx - \theta.$$

が成り立つことを示す.

まず, 左の等号は  $\theta > 0$  とすれば ( $\theta \leq 0$  のときも全く同様),

$$\begin{aligned} Pm_\theta &= \int_{\mathbb{R}} \{|x - \theta| - |x|\} dF(x) \\ &= \int_{-\infty}^0 \{-(x - \theta) + x\} dF(x) + \int_0^\theta \{-(x - \theta) - x\} dF(x) + \int_\theta^\infty \{x - \theta - x\} dF(x) \\ &= \theta F(0) + \int_0^\theta (\theta - 2x) dF(x) - \theta(1 - F(\theta)) \end{aligned}$$

として求まる. 右側の等号は部分積分により

$$\begin{aligned} 2 \int_0^\theta F(x) dx - \theta &= 2 \left\{ [xF(x)]_0^\theta - \int_0^\theta x dF(x) \right\} - \theta \\ &= \theta F(\theta) + \int_0^\theta (-2x) dF(x) - \theta(1 - F(\theta)) \\ &= \theta F(0) + \int_0^\theta (\theta - 2x) dF(x) - \theta(1 - F(\theta)). \end{aligned}$$

よって示せた.

これより,

$$\begin{aligned} \frac{\partial}{\partial \theta} Pm_\theta &= 2F(\theta) - 1 \\ \frac{\partial^2}{\partial \theta^2} Pm_\theta &= 2f(\theta) \end{aligned}$$

よって,  $Pm_\theta$  を  $\theta_0$  回りでテーラー展開すると

$$Pm_\theta = Pm_{\theta_0} + \frac{1}{2} 2f(\theta_0)(\theta - \theta_0)^2 + o(|\theta - \theta_0|^2).$$

よって,  $V_{\theta_0} = 2f(\theta_0) > 0$  (仮定より). よって条件 3 も示された.

ゆえに Theorem 5.23 より

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow N\left(0, \frac{1}{\{2f(\theta_0)\}^2}\right).$$

□

**Example 5.25** (Misspecified model). サンプル  $X_1, \dots, X_n$  に対してモデル  $\{p_\theta : \theta \in \Theta\}$  を考える. 真のモデルがこのモデルに含まれていないと仮定する (misspecified). このモデルから対数尤度  $\sum \log p_\theta(X_i)$  を最大にする  $\hat{\theta}_n$  を推定量として得たとする.  $\hat{\theta}_n$  の漸近的な振る舞いはどうなるだろうか?

$\hat{\theta}_n$  は誤ったモデルを使用したために, 漸近的にも誤った振る舞いをすると考えるかもしれないが, 実はそうではない. まず,  $\hat{\theta}_n$  は  $\theta \mapsto P \log p_\theta$  (真の分布での期待値) を最大にする  $\theta_0$  に対して, 一致性を持つ.  $p_{\theta_0}$  は真の分布  $P$  からのカルバックライブラーダイバージェンス  $-P \log(p_\theta/p)$  を最小にする  $\{p_\theta : \theta \in \Theta\}$  の元であることが分かる. すなわち真のモデルのモデル空間への射影が  $p_{\theta_0}$  である.

さらに我々は,  $\sqrt{n}(\hat{\theta}_n - \theta_0)$  が平均 0, 共分散行列  $V_{\theta_0}^{-1} P \dot{\ell}_{\theta_0} \dot{\ell}_{\theta_0}^T V_{\theta_0}^{-1}$  の正規分布に分布収束することを期待する. ただし,  $\ell_\theta = \log p_\theta$  であり,  $V_{\theta_0}$  は写像  $\theta \mapsto P \log p_\theta$  のヤコビ行列である.

$m_\theta = \log p_\theta$  とおけば, Theorem 5.23 によりこれを示すことができる.

以上のような考察から, モデルが真の分布からそれほど遠くなければ,  $p_{\hat{\theta}_n}$  は真の分布の妥当な推定になりうる.

**Example 5.26** (Exponential frailty model). 生存時間のペアのサンプル  $(X_1, Y_1), \dots, (X_n, Y_n)$  を考える. 例えば  $X_i$  は”父親”の生存時間で  $Y_i$  は”息子”の生存時間などである. 観測されない値  $z_i$  が与えられた時に,  $X_i$  と  $Y_i$  は独立で, それぞれ指数分布  $Exp(z_i), Exp(\theta z_i)$  に従うと仮定する. パラメータ  $\theta$  を推定する問題を考える.

まず,  $z_1, \dots, z_n$  はある未知の分布に従う確率変数  $Z_1, \dots, Z_n$  の実現値であるとする.

一つ目のアプローチは  $\theta$  が既知のときの,  $z_i$  が与えられた時の  $X_i + \theta Y_i$  の十分性に基づいた方法である.

$Z_i = z$  が与えられた時, 統計量  $X_i + \theta Y_i$  はガンマ分布  $Gamma(2, z)$  に従う.

$\because X_i \sim Exp(z) = Gamma(1, z), \theta Y_i \sim \theta Exp(\theta z) = \theta Gamma(1, \theta z) = Gamma(1, z)$ .  $z$  が与えられたとき,  $X_i, Y_i$  は独立と仮定していたので, ガンマ分布の再生性から  $X_i + \theta Y_i = Gamma(2, z)$

さて,  $X_i + \theta Y_i$  は  $z_i$  に対する十分統計量より, フィッシャーの因子分解定理から,  $(X, Y)$  の同時分布  $p_\theta(x, y|z)$  は次のように分解できる.

$$p_\theta(x, y|z) = h_\theta(x, y) g_\theta(x + \theta y|z)$$

ただし

$$\begin{aligned} p_\theta(x, y|z) &= z e^{-zx} \theta z e^{-\theta zy} \\ &= \frac{\theta}{x + \theta y} z^2 (x + \theta y) e^{-z(x + \theta y)} \end{aligned}$$

より,

$$h_\theta(x, y) = \frac{\theta}{x + \theta y}, \quad g_\theta(x + \theta y|z) = z^2 s e^{-zs}, \quad \text{where } s = x + \theta y.$$

$g_\theta(s|z)$  は  $Gamma(2, z)$  の密度関数である.

$X_i + \theta Y_i$  は  $z_i$  に依存するが,  $z_i$  は観測できないので,  $g_\theta(x + \theta y|z)$  を尤度から無視して,  $h_\theta(x, y)$  だけを使うということを考えるかもしれないが, この”条件付き尤度”  $h_\theta(x, y)$  は尤度とは異なる振る舞いをするのでそうはいかない. なぜなら,  $h_\theta$  に対応する,  $\theta$  を求めるための条件付き尤度方程式

$$\frac{\dot{h}_\theta}{h_\theta}(x, y) = \frac{\partial}{\partial \theta} \log h_\theta(x, y)$$

は真の分布で期待値をとったとき,  $\theta > 0$  をどのようにとっても 0 とならないからである. 実際

$$\frac{\dot{h}_\theta}{h_\theta}(x, y) = \frac{x}{\theta(x + \theta y)} > 0$$

である.

このバイアスは十分統計量によって条件付けることで修正できる.

$$\psi_{\theta}(X, Y) = 2\theta \frac{\dot{h}_{\theta}}{h_{\theta}}(X, Y) - 2\theta E_{\theta} \left( \frac{\dot{h}_{\theta}}{h_{\theta}}(X, Y) | X + \theta Y \right) = \frac{X - \theta Y}{X + \theta Y}.$$

とする.

$\therefore$  2つ目の等号が成り立つことを示す.

$$2\theta \frac{\dot{h}_{\theta}}{h_{\theta}}(X, Y) = \frac{2X}{X + \theta Y}.$$

よって

$$2\theta E_{\theta} \left( \frac{\dot{h}_{\theta}}{h_{\theta}}(X, Y) | X + \theta Y \right) = 1$$

を示せばよい. が, 示せない...

$\hat{\theta}_n$  を  $\mathbb{P}_n \psi_{\theta} = 0$  の解として得る. このようにして求められた  $\hat{\theta}_n$  の漸近的性質は次回述べる.