記号の訂正（スライド18）

$$\Psi(\theta, x) = \frac{d}{d\theta} \log p_\theta(x)$$ p×1 ベクトル

$$\dot{\Psi}(\theta, x) = \frac{d^2}{d\theta^2} \log p_\theta(x)$$ p×p 行列

log

(12e) 最尤推定量量の漸近正規性 (Fergason (2017, pp.135-137):

1. $\Theta \in \mathbb{R}^p$ の開集合.
2. p_θ の θ に関する 2 階微分が存在して, 函数 $\Theta \ni \mapsto \frac{d^2}{d\theta^2} p_\theta(x) =: \dot{\Psi}(\theta, x)$ は μ に関してほとんどいたるところで連続. さらに, その 2 階微分は $\int p_\theta \, dP_0$ の積分と順序交換できる.
3. 函数 $G(x) > 0$ が存在して, $P_0 G < \infty$ であり, θ_0 の近傍の θ に対して,

G $\dot{\psi}$

$$\mathrm{tr}\,(\dot{\Psi}(\theta, x)\dot{\Psi}^\top(\theta, x)) \le K(x) \quad (a.e.\,\mu). \quad \left(\dot{\Psi}(\theta, x) := \frac{d^2}{d\theta^2} p_\theta(x)\right).$$

$J(\theta)$

log

4. $p \times p$ の行列 $J = -P_0 \dot{\Psi}(\theta, \cdot)]$ は正定値.
5. $p_\theta = p_{\theta'}\,(a.e.\,\mu) \iff \theta = \theta'$.

トル

このとき, $\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} N_p(0, J^{-1})$.

$\{J(\theta_0)\}^{-1}$

$\psi(\theta, x) = \frac{d}{d\theta} \cancel{p_\theta}\ \frac{d}{d\theta} \log p_\theta$

(4) 確率収束間の関係

(4a) 概収束するならば, 確率収束する. この逆は一般には成立しない. 成立する場合にも確認が大変！

(4b) 確率収束するならば, 分布収束

(5) スラツキーの定理

(5a) $W_n \xrightarrow{L} W$ かつ $Z_n \xrightarrow{P} c\ (n \to \infty)$ のとき,
$(W, Z_n) \xrightarrow{L} (W, c)\ (n \to \infty)$.

(6) 連続写像定理

(6a) $h : \mathbb{R} \longrightarrow \mathbb{R}$ は連続函数とする. $W_n \xrightarrow{*} W\ (n \to \infty)$ ならば,

$$h(W_n) \xrightarrow{*} h(W)\ (n \to \infty).$$

ただし, $* = P$, a.s., L である.

(12) X を確率空間 $(\Omega, \mathcal{F}, \mathbb{P})$ 上の実数値確率変数とし, $P_0(B) := \mathbb{P}(X \in B)$ $(B \in \mathcal{B}(\mathbb{R}))$ とする. μ を $\mathcal{B}(\mathbb{R})$ 上の Lebesgue 測度とする. $\forall B \in \mathcal{B}(\mathbb{R})$ に対して,
$\mu(B) = 0 \implies P_0(B)$ が常になりたつとする. これを $P_0 \ll \mu$ とかく.

$\longrightarrow$ $P_0(B) = 0$

(12a) Θ は母数空間 (この次元では無限次元でもよい) とし, 統計モデル

$$\mathcal{P} := \{p_\theta : \theta \in \Theta,\ p_\theta \ll \mu \text{ は確率密度函数 }\}$$

(12b) $\theta_0 := \operatorname*{argmax}_{\theta \in \Theta} P_0 p_\theta$ は一意的に存在すると仮定する. θ_0 は真のモデル P_0 に対応する真の母数と考えることができる. ただし, argmax の意味は函数

$$\Theta \ni \theta \mapsto P_0 p_\theta = \int_{\mathbb{R}} p_\theta(x)\, dP_0(x) \in [0, 0) \cup \{\infty\}$$

の最大を与える θ の値である.

(12d) 最尤推定量の一致性について (Fergason (2017, pp.128-129):

1 $\Theta \subset \mathbb{R}^p$ はコンパクトと仮定する.

2 x を固定すると, 写像 $\Theta \ni \theta \mapsto p_\theta(x) \in [0, \infty)$ は上半連続.

3 $P_0|G| < \infty$ なる実数値可測函数が存在して, $\forall\theta \in \Theta$ に対して

$$\log p_\theta(x) - \log p_{\theta_0}(x) \leq G(x) \qquad (a.e.\,\mu).$$

4 $\forall\theta \in \Theta$ と小さな $\delta > 0$ 1 に対して, 写像

$$\mathbb{R} \ni x \mapsto \sup_{|\theta'-\theta|<\delta} p_{\theta'}(x) \in [0, \infty)$$

$\Theta_0 \to \Theta$

は可測.

5 $p_\theta = p_{\theta'}\,(a.e.\,\mu) \iff \theta = \theta'$.

このとき, $\widehat{\theta}_n \xrightarrow{a.s.} \theta_0\ (n \to \infty)$.

$$\frac{1}{\sqrt{n}}\dot{\ell}_n(\theta_0) = B_n\sqrt{n}\,(\hat{\theta}_n - \theta_0)$$

$$B_n := -\int_0^1 \frac{1}{n}\sum_{i=1}^{n}\dot{\Psi}\{\theta_0 + s(\hat{\theta}_n - \theta_0), X_i\}\,ds$$

となる。さらに、

$$-P_0\dot{\Psi}(\theta_0,\cdot)$$

$$\shortparallel$$

$$P_0\Psi(\theta_0,\cdot) = 0, \quad P_0\left[\Psi(\theta_0,\cdot)\Psi^T(\theta_0,\cdot)\right] = J(\theta_0)$$

に注意する。CLT より

$$\frac{1}{\sqrt{n}}\dot{\ell}_n(\theta_0) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\Psi(\theta_0, X_i)\right)$$

$$\xrightarrow{d} \cancel{N_p(0,\{J(\theta_0)\})}$$

$$N_p(\cancel{\theta_0}, J(\theta_0)).$$

No. 2

0 ゼロベクトル

さらに、

$$B_n \xrightarrow{a.s.} J(\theta_0) \quad (n\to\infty) \qquad (b)$$

がわかれば、スラツキーの定理と連続写像定理より

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N_p(0, J^{-1}(\theta_0)).$$

注意 (b)を示すために、$\delta>0$に対して、

$$\sup_{\theta\in B_\delta(\theta_0)} \| P_n\dot{\Psi}(\theta,\cdot) - P_0\dot{\Psi}(\theta,\cdot)\|_F \xrightarrow{a.s.} 0 \quad (n\to\infty)$$

が証明できると楽である。ただし、 No.3

$B_\delta(\theta_0)=\{\theta\in\Theta;\ |\theta-\theta_0|<\delta\}$, $|\cdot|$: Euclid norm.

$\sqrt{n}\{(P_n - P)\mathbb{1}_{(-\infty,x]}(\cdot)$ の $\mathbb{1}_{(-\infty,x]}(\cdot)$ のところを $g : \mathbb{R} \longrightarrow \mathbb{R}$ なる有界な可測函数族 $\mathcal{G}$ の要素で置きかえれば,

$$\sqrt{n}(P_n - P)g, \qquad (g \in \mathcal{G})$$

とみなせる.

Gilvenko-Cantelli の定理

$\mathcal{G}_1 = \{g(t) = \mathbb{1}_{(-\infty,x]}(t)\ (x \in \mathbb{R})\}$ のとき,

$$\sup_{g\in G} |(P_n - P)g| \xrightarrow{a.s.} 0 \qquad (n \to \infty).$$

$G \longrightarrow \mathcal{G}_1$

だが, $\mathcal{G}_2 = \{g(t) = \mathbb{1}_B(t)\ (B \in \mathcal{B}(\mathbb{R}))\}$ のとき,

$$\sup_{g\in G} |(P_n - P)g| \qquad (n \to \infty)$$

$G \longrightarrow \mathcal{G}_2$

はうまくコントロールできないことが知られている. $\mathcal{G}_0$ 以外の意味のある函数族はあるのか?

X は確率空間 $(\Omega, \mathcal{F}, \mathbb{P})$ 上で定義された実数値確率変数で, $P = \mathbb{P} \circ X^{-1}$ とする. (Θ, d_T) はある距離空間の部分集合 (無限次元でもよい) とする. 統計モデルを $\mathcal{P} := \{p_\theta; , \theta \in \Theta, p_\theta$ は $\mathbb{R}$ 上の σ 有限な測度 μ に関する密度函数$\}$ さらに,

$$\theta_0 := \mathrm{argmax}_{\theta\in\Theta} P_0 p_\theta, \quad \widehat{\theta}_n := \mathrm{argmax}_{\theta\in\Theta} P_n p_\theta$$

で定める. いま

$$\mathcal{G}_3 := \left\{ \log\left(\frac{p + p_0}{2p_0}\right) \mathbb{1}_{\{x\in\mathbb{R};\, p_0>0\}};\ p \in \mathcal{P} \right\}$$

で定める. すると, ゆるい条件のもと

$$\sup_{g\in\mathcal{G}_3} |(P_n - P)g| \xrightarrow{P} 0 \qquad (n \to \infty)$$

$$p_0 = \frac{dP_0}{d\mu}$$

がわかる. さらに,

$$d_H(p_{\widehat{\theta}_n}, p_0) \xrightarrow{P} 0 \qquad (n \to \infty).$$

ただし, $d_H^2(p, p') := \frac{1}{2}\int_{\mathbb{R}} \left(\sqrt{p} - \sqrt{p'}\right)^2 d\mu(x)\ (p, p' \in \mathcal{P})$