2021 年度　数理科学特別講義 C, G
大阪府立大学
8 月 26 日 第 2 コマ

今野良彦

日本女子大学 理学部 数物科学科

2021 年 8 月 26 日

実数値確率変数列の収束についての復習

(1) $\{W_n\}_{n\geq 1}$, W は同じ確率空間 $(\Omega, \mathcal{F}, \mathbb{P})$ 上で定義されている実数値確率変数列とする．

(2) 確率変数列の収束の 3 つのモード（様式）

(2a) 確率収束：任意の $\epsilon > 0$ に対して,

$$\lim_{n\to\infty} \mathbb{P}\left(\{\omega \in \Omega;\ |W_n(\omega) - W(\omega)| > \epsilon\}\right) = \lim_{n\to\infty} \mathbb{P}(|W_n - W| > \epsilon) = 0.$$

この確率収束することを $W_n \xrightarrow{P} W\ (n \to \infty)$ と記す．

(2b) 概収束：$\mathbb{P}\left(\{\omega \in \Omega;\ \lim_{n\to\infty} W_n(\omega) \neq W(\omega)\}\right) = 0$. この収束を $W_n \xrightarrow{a.s.} W\ (n \to \infty)$ と記す．

(2c) 分布 (法則) 収束：$\mathbb{R}$ 上の任意の有界連続函数 h に対して,

$$\lim_{n\to\infty} \mathbb{E}[h(W_n)] = \mathbb{E}[W] \qquad (n \to \infty)$$

$\mathbb{E}[h(W)]$

記法: 任意の $B \in \mathcal{B}(\mathbb{R})$ 対して, $P_n(B) = \mathbb{P}(B \in W_n)$ が成り立つことを $P_n = \mathbb{P} \circ W_n^{-1}$ と記す．

3

$$\mathbb{E}[h(W_n)] = \mathbb{P}(h(W_n)) = \int_\Omega h(W_n(\omega))\, d\mathbb{P}(\omega)$$
$$= \int_{\mathbb{R}} h(x)\, dP_n(x).$$

$$\mathbb{E}[h(W)] = \mathbb{P}(h(W)) = \int_\Omega h(W(\omega))\, d\mathbb{P}(\omega)$$
$$= \int_{\mathbb{R}} h(x)\, dP(x).$$

$\lim_{n\to\infty} \mathbb{E}[h(W_n)] = \mathbb{E}[h(W)]$ ($\forall h$: 有界連続)

$\Longrightarrow$ $W_n \xrightarrow{L} W$ $(n \to \infty)$ と記す

(3) 法則収束の同値条件 (2c) $\Longleftrightarrow$ (3a) $\Longleftrightarrow$ (3b).

(3a) $\mathbb{R}$ 上の任意の有界連続函数 h に対して,

$$\int_{\mathbb{R}} h(x)\, dP_n(x) = \int_{\mathbb{R}} h(x)\, dP(x).$$

ただし, $P_n = \mathbb{P} \circ W_n^{-1}$, $P = \mathbb{P} \circ W^{-1}$. これは確率測度（分布）$P_n$ が確率測度 P に弱収束すると理解できるので, $P_n \Rightarrow P\ (n \to \infty)$ と記す.

(3b) $\lim_{n\to\infty} F_n(x) = F(x)$. ただし, x は F の連続点である. また, $F_n(x) = P_n((-\infty, x])$, $F(x) = P((-\infty, x])$ である.

(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.

(7) デルタ法
$p \geq q$ は自然数とし, $\underline{W}_n$, $\underline{W}$ を p 次元確率縦ベクトル列, $\underline{\mu} \in \mathbb{R}^p$ を定数縦ベクトルとする. 函数 $\underline{g}: \mathbb{R}^p \longrightarrow \mathbb{R}^q$ の導関数 $\underline{\dot{g}}$ は $\underline{\mu} \in \mathbb{R}^p$ の近傍である連続であるとする. このとき,

$$\sqrt{n}(\underline{W}_n - \underline{\mu}) \xrightarrow{d} \underline{W} \qquad (n \to \infty)$$

ならば,

$$\sqrt{n}\left(\underline{g}(\underline{W}_n) - \underline{g}(\underline{\mu})\right) \xrightarrow{d} \underline{\dot{g}}(\underline{\mu})\underline{W} \qquad (n \to \infty).$$

特に,

$$\sqrt{n}(\underline{W}_n - \underline{\mu}) \xrightarrow{d} N_p(0, \Sigma) \qquad (n \to \infty)$$

ならば,

$$\sqrt{n}\left(\underline{g}(\underline{W}_n) - \underline{g}(\underline{\mu})\right) \xrightarrow{d} N_q\left(\underline{0}, \underline{\dot{g}}(\underline{\mu})\Sigma\underline{\dot{g}}(\underline{\mu})^\top\right) \qquad (n \to \infty)$$

ただし, $\boldsymbol{\Sigma}$ は $p \times p$ の正値対称行列である. また,
$\underline{w} = (w_1, w_2, \ldots, w_p)^\top$ と $\underline{g} = (g_1, g_2, \ldots, g_q)^\top$ としたとき,

$$\dot{\underline{g}}(\underline{w}) = \begin{pmatrix} \frac{\partial g_1}{\partial w_1} & \frac{\partial g_1}{\partial w_2} & \cdots & \frac{\partial g_1}{\partial w_p} \\ \frac{\partial g_2}{\partial w_1} & \frac{\partial g_2}{\partial w_2} & \cdots & \frac{\partial g_2}{\partial w_p} \\ \vdots & & \ddots & \vdots \\ \frac{\partial g_q}{\partial w_1} & \frac{\partial g_1}{\partial w_2} & \cdot & \frac{\partial g_q}{\partial x_p} \end{pmatrix} \longrightarrow w_p$$

である.

(8) 大数の弱法則
$\{X_n\}_{n\geq 1}$ は確率空間 $(\Omega, \mathcal{F}, \mathbb{P})$ 上で定義された独立同一分布に従う確率変数列で,

$$\mathbb{E}[X_1] = \mu \, (-\infty < \mu < \mu)$$

とする. このとき,

$$\overline{X}_n := \frac{\sum_{n=1}^{\infty} X_n}{n}$$

と定義したとき,

$$\overline{X}_n \xrightarrow{P} \mu \, (n \to \infty)$$

となる.

(9) $\mathbb{E}[X_1^2] < \infty$ の場合の証明. チェビシェフの不等式を思い出す:
$\mathbb{E}[W^2] < \infty$, $\mathbb{E}[W] = \mu$ なる確率変数に対して, 任意の $\epsilon > 0$ に対して,

$$\mathbb{P}(|W - \mu| > \epsilon) \leq \frac{\mathbb{E}[(W-\mu)^2]}{\varepsilon^2}.$$

これより

$$\mathbb{P}(|\overline{X}_n - \mu| > \epsilon) \leq \frac{\mathbb{E}[(\overline{X}_n - \mu)^2]}{\epsilon} = \frac{\mathbb{E}[(X_1 - \mu)^2]}{n\epsilon^2}.$$

よって,

$$\lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mu| > \epsilon) = 0$$

よって, $\overline{X}_n \xrightarrow{P} \mu$ は簡単にわかる.
$\mathbb{E}[|X_1|] < \infty$ のもとで,

$$\overline{X}_n \xrightarrow{a.s.} \mu$$

を示すのは手間がかかる. たとえば, ① $\mathbb{E}[X_1^4] < \infty$ の場合には, 概収束の意味で大数の法則（大数の強法則）を示すことは比較的容易である. さらに, ② バックワード・マルチンゲールの収束理論を使うと比較的簡単に弱法則から強法則の格上げは比較的ルーティンの議論に帰着できる.

(10) 中心極限定理（CLT）の復習：$\{X_n\}_{n\geq 1}$ を確率空間 $(\Omega, \mathcal{F}, \mathbb{P})$ 上の独立同一分布に従う確率変数列とし,

$$\mathbb{E}[X_1] = \mu, \qquad \mathbb{V}[X_1] = \mathbb{E}[(X_1-\mu)^2] = \sigma^2 \qquad (0 < \sigma < \infty).$$

このとき,

$$\mathbb{P}\left(\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \leq z\right) \longrightarrow \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt \qquad (z \in \mathbb{R}).$$

上式の右辺の被積分関数は標準正規分布の確率密度関数である.

(11) 中心極限定理の証明についてのコメント

1. 特性函数を利用する方法. 杉浦 (2020) を参照.
2. リンデベルグの証明. Knight (1996) または加藤 (2016) を参照.
3. スタインの方法による表明. 小池 (2020) を参照.

(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.

(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.

(12c) $X_1, X_2, \ldots, X_n$ は X の独立複製とし, θ_0 の最尤推定量 $\widehat{\theta}_n$ を

$$\widehat{\theta}_n := \operatorname*{argmax}_{\theta\in\Theta} P_n p_\theta, \qquad P_n p_\theta = \frac{1}{n}\sum_{i=1}^{n} p_\theta(X_i)$$

で定義する. $\widehat{\theta}_n$ は一意的ではないかもしれないが, 存在はすると仮定する. また, $B \in \mathcal{B}(\mathbb{R})$ に対して,

$$P_n(B) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_B(X_i)$$

であった.

(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

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

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

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

θ_0

は可測.

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

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

証明の要点

$\theta \in \Theta$ に対して

$$R(\theta) := P_0(-\log p_\theta) = \int_{\mathbb{R}} (-\log p_\theta(x))\, dP_0(x),$$

$$P_0(B) = \int_B p_0(x)\, d\mu(x) \quad (\forall B \in \mathcal{B}(\mathbb{R}))$$

とおく。さらに

$$R_n(\theta) := P_n(-\log p_\theta) = \frac{1}{n}\sum_{i=1}^{n}(-\log p_\theta(X_i)).$$

このとき

$$\sup_{\theta \in \Theta} |R_n(\theta) - R(\theta)| \xrightarrow{a.s.} 0 \quad (n \to \infty)$$

が証明できたとする。

No. 1

函数

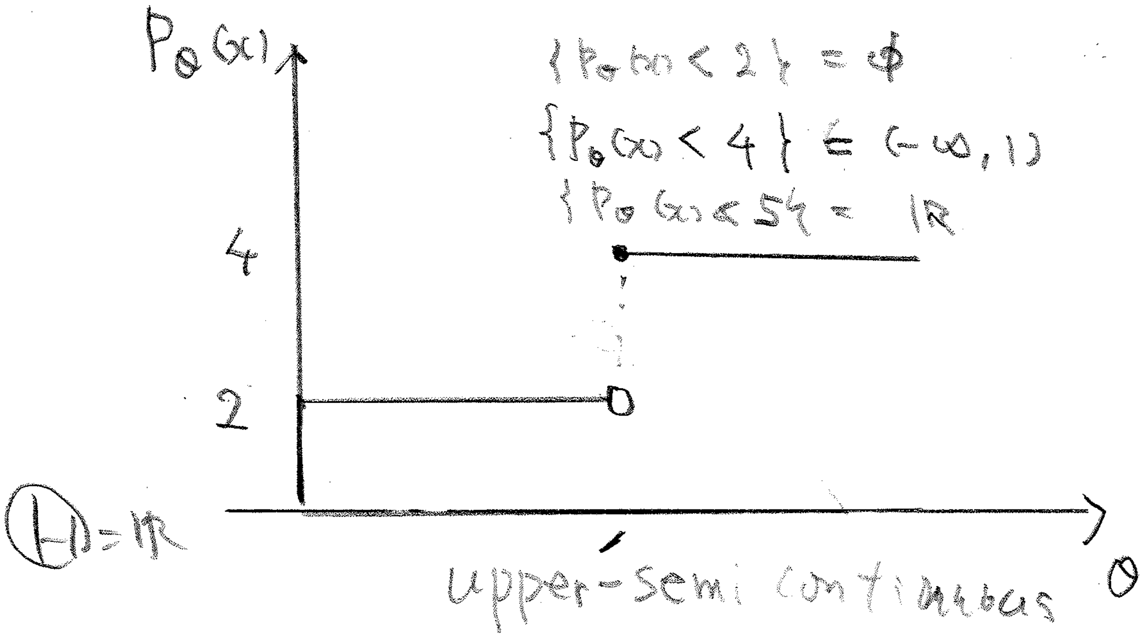
$\Theta \ni \theta \longmapsto p_\theta(x) \in [0, \infty)$

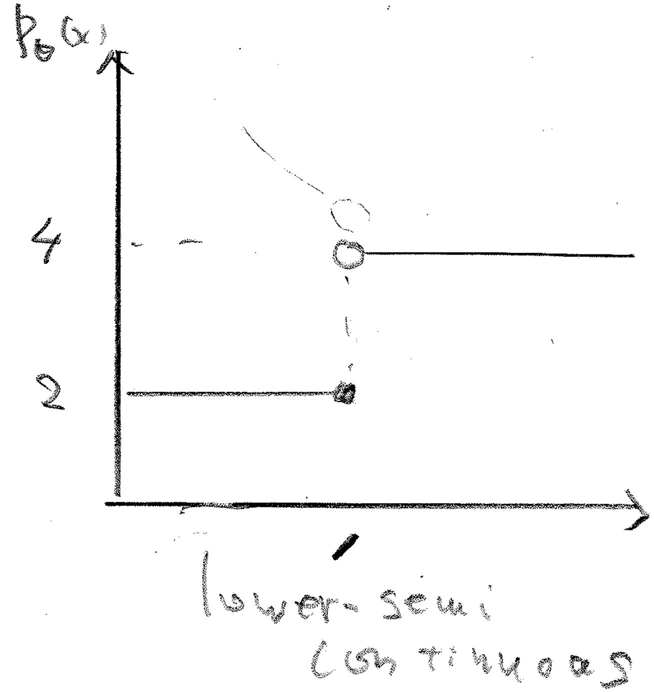
下半連続
$\{p_\theta(x) > r\}$ が開

は上半連続（xを固定する）

$\Longleftrightarrow$

$\{\theta \in \Theta \,;\, p_\theta(x) < r\}$ はすべての $r \in \mathbb{R}$ に対して開集合





函数 $\Theta \ni \theta \longmapsto R(\theta) \in [0, \infty]$ の連続性より，

$$\hat{\theta}_n \xrightarrow{a.s.} \theta_0 \quad (n \to \infty).$$

No.2

(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 $\dot{\psi}$

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

J(θ)

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

トル

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

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

証明の要点

$\theta \in \Theta$ に対して、

$$\dot{l}_n(\theta) := \sum_{i=1}^{n} \Psi(\theta, X_i) \quad : \ p \times 1$$

とおく。$\dot{l}_n(\theta)$ を θ_0 のまわりで展開：

$$\dot{l}_n(\theta) = \dot{l}_n(\theta_0) + \Bigl(\underbrace{\int_0^1 \sum_{i=1}^{n} \dot{\Psi}\{\theta_0 + \lambda(\theta - \theta_0), X_i\}\, d\lambda}_{p \times p \text{ の行列}}\Bigr)(\theta - \theta_0) \qquad (a)$$

ここで、最尤推定量 $\hat{\theta}_n$ は

$$\dot{l}_n(\hat{\theta}_n) = 0$$

をみたすので、(a) に $\theta = \hat{\theta}_n$ を代入すれば

No. 1

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

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

となる。さらに、

$$P_0\Psi(\theta_0,\cdot) = 0, \quad \underset{\substack{\| \\ -P_0\dot{\Psi}(\theta_0,\cdot)}}{} P_0[\Psi(\theta_0,\cdot)\Psi^T(\theta_0,\cdot)] = J(\theta_0)$$

に注意する。CLT より

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

$$\xrightarrow{d} \cancel{N_p(0, \{J(\theta_0)\}^{-1})}$$

$$N_p(\theta_0, J(\theta_0)).$$

No. 2

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

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

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.

$\|B\|_F^2 = \mathrm{tr}\{BB^T\}$ B: $p\times p$の行列

No. 4

$X_1, X_2, \ldots, X_n$ を X の独立な複製とする．ここで，$P = \mathbb{P} \circ X^{-1}$ とした．したがって，X_i $(i = 1, \ldots, n)$ は独立同一に分布 P に従う (F に従うといってもよい)．

(13) 集合 B に対して，指示函数を

$$\mathbb{1}_B(x) := \begin{cases} 1 & (x \in B), \\ 0 & (x \notin B) \end{cases}$$

で定義する．

(14) $X_1, X_2, \ldots, X_n$ に基づく経験分布函数を

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

経験確率測度を

$$P_n(B) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_B(X_i) \qquad (B \in \mathcal{B}(\mathbb{R})).$$

CLT より，各 x に対して

$$\mathbb{P}\left(\frac{\sqrt{n}(F_n(x)-F(x))}{\sqrt{F(x)(1-F(x))}}\leq z\right)\longrightarrow\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}\,dt \qquad (z\in\mathbb{R}).$$

また，各 $B\in\mathcal{B}(\mathbb{R})$ に対して，

$$\mathbb{P}\left(\frac{\sqrt{n}(P_n(B)-P(B))}{\sqrt{P(B)(1-P(B))}}\leq z\right)\longrightarrow\int_{B}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}\,dt \qquad (z\in\mathbb{R}).$$

点 x を実数上や B を σ 加法族上を動かしたら，どうなるだろうか？
$\sqrt{n}\{F_n(x)-F(x)\}$ を右連続・左極限を持つ函数とみなして，函数型の中心極限定理のようなものを得ることができる．しかし，$P_n(B)$ については，一般にはできない．

そこで, 拡張できると期待される $\sqrt{n}\{F_n(x) - F(x)\}$ を書き換えてみよう.
$x \in \mathbb{R}$ に対して, 函数

$$\mathbb{R} \ni t \mapsto \mathbb{1}_{(-\infty, x]}(t) \in [0, 1]$$

を用いると

$$F(x) = P\mathbb{1}_{(-\infty, x]}(\cdot) \quad F_n(x) = P_n\mathbb{1}_{(-\infty, x]}(\cdot)$$

とかける. したがって,

$$\sqrt{n}\{F_n(x) - F(x)\} = \sqrt{n}(P_n - P)\mathbb{1}_{(-\infty, x]}(\cdot)$$

$\sqrt{n}\{(P_n - P)\mathbb{1}_{(-\infty,x]}(\cdot)$ 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,

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

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

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

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

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

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

$\mathcal{G}_1$ と $\mathcal{G}_3$ は

$$\sup_g |(P_n - P)g|$$

をコントロールできるが, $\mathcal{G}_2$ は

$$\sup_g |(P_n - P)g|$$

をうまくコントロールできない.

この違いは函数族の豊かさによって決まる. あまりおおきな函数族では

$$\sup_g |(P_n - P)g|$$

をうまくコントロールできない.

では, どの函数族がそれほどひろいものではないかをどうやって判定するだろうか？

さらに, このような経験過程を考えると可測性の問題が発生する.

上の議論では, それを気にせずに形式的に議論をすすめていることに注意せよ.