2021 年度　数理科学特別講義 C, G
大阪府立大学
8 月 25 日 第 4 コマ

今野良彦

日本女子大学 理学部 数物科学科

2021 年 8 月 25 日

1. イントロダクション：経験過程とはなにか？
    - 記号の導入
    - 経験過程の古典的な極限定理
    - 現代経験過程理論
    - 授業計画と参考文献
    - 統計モデルと統計推測についての注意

記号の導入

(1) $(\Omega, \mathcal{F}, \mathbb{P})$：確率空間
(2) $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$：測度空間. ただし, $\mathbb{R}$ は実数全体, $\mathcal{B}(\mathbb{R})$ は $\mathbb{R}$ のボレル加法族. すなわち, $\mathbb{R}$ の開集合を含む最小の σ 加法族.
(3) $X : \Omega \to \mathbb{R}$ は確率変数とする. すなわち, 任意の $B \in \mathcal{B}(\mathbb{R})$ に対して,

$$X^{-1}(B) = \{\omega \in \Omega;\ X(\omega) \in B\} \in \mathcal{F}.$$

(4) X の分布函数 F を次で定義：任意の $x \in \mathbb{R}$ に対して,

$$F(x) := \mathbb{P}(X \le x) = \mathbb{P}(\{\omega \in \Omega;\ X(\omega) \le x\}).$$

また, X の分布 P を次で定義：任意の $B \in \mathcal{B}(\mathbb{R})$ に対して

$$P(B) := \mathbb{P}(X \in B) = \mathbb{P}(\{\omega \in \Omega;\ X(\omega) \in B\}).$$

したがって, $B = (-\infty, x]$ とすれば

$$P(B) = F(x).$$

(5) $X_1, X_2, \ldots, X_n$ を X の独立な複製とする. 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).

(6) 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.

(7) $X_1, X_2, \ldots, X_n$ 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

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

で定める.

$\widehat{F}_n(x)$ のグラフを描く.
$n = 3$ とし, 観測 $X_1 = 3,\ X_2 = 1,\ X_3 = 4$ を得たとする.

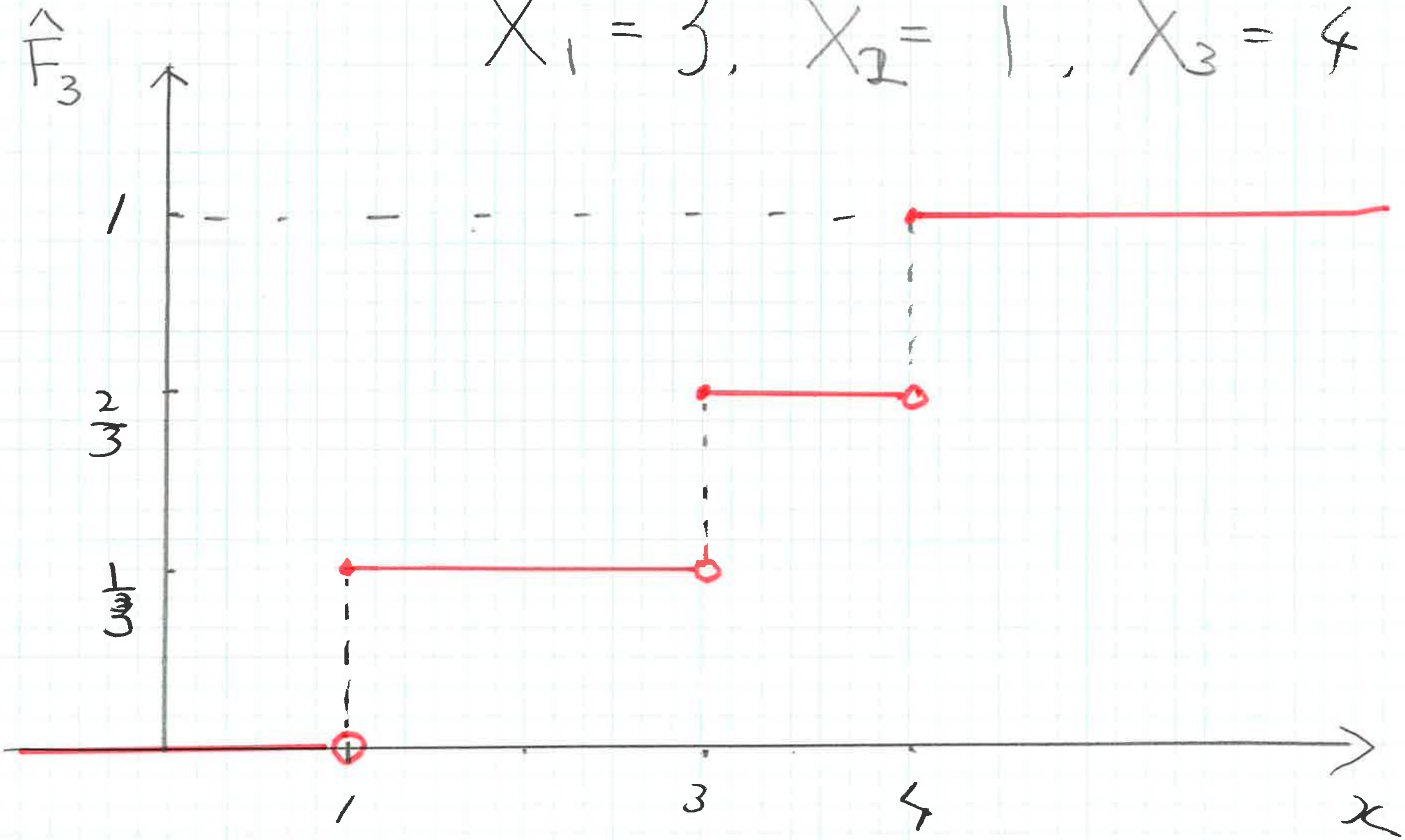
$$\mathbb{1}_{(-\infty, x]}(3) = \begin{cases} 1 & (x \geq 3) \\ 0 & (x < 3) \end{cases}$$

となることに注意して,

$$\widehat{F}_3(x) = \frac{1}{3}\Big(\mathbb{1}_{(-\infty, x]}(3) + \mathbb{1}_{(-\infty, x]}(1) + \mathbb{1}_{(-\infty, x]}(4)\Big)$$

を描いてみよう.

$X_1 = 3,\ X_2 = 1,\ X_3 = 4$



$\hat{F}_3$ のグラフ

No. 6

経験過程の古典的な極限定理

(8) 興味のあることは F_n と F および P_n と P の関係である. すぐにわかることは, 各 x と B に対して,

$$\mathbb{E}[\widehat{F}_n(x)] = F(x), \qquad \mathbb{E}[P_n(B)] = P(B).$$

(9) 大数の強法則の復習：$Z_1, Z_2, \ldots, Z_n$ を確率空間 $(\Omega, \mathscr{F}, \mathbb{P})$ 上の独立同一分布に従う確率変数列とする.
$\mathbb{E}[|Z_1|] < \infty$ $(\mu = \mathbb{E}[Z_1])$ ならば

$$\overline{Z}_n := \frac{Z_1 + Z_2 + \cdots + Z_n}{n} \xrightarrow{a.s.} \mu \qquad (n \to \infty)$$

となる. すなわち,

$$\mathbb{P}\left(\lim_{n\to\infty} \overline{Z}_n = \mu\right) = 1.$$

(10) 各 x と B に対して, $\mathbb{1}_{(-\infty,x]}(X_1), \ldots, \mathbb{1}_{(-\infty,x]}(X_n)$ と $\mathbb{1}_B(X_1), \ldots, \mathbb{1}_B(X_n)$ は独立に同一分布に従う確率変数列で $\mathbb{E}[\mathbb{1}_{(-\infty,x]}(X_1)] = F(x)$ なので, 大数の強法則より

$$\widehat{F}_n(x) \xrightarrow{a.s.} F(x), \qquad P_n(B) \xrightarrow{a.s.} P(B) \qquad (n \to \infty)$$

がわかる.

(11) Glivenko–Cantelli の定理：$n \to \infty$ のとき,

$$\sup_{-\infty<x<\infty} |\widehat{F}_n(x) - F(x)| = \sup_{B\in\mathcal{I}} |P_n(B) - P(B)| \xrightarrow{a.s.} 0.$$

ただし, $\mathcal{I} = \{(-\infty, x];\ x \in \mathbb{R}\}$.

(12) ところが

$$\sup_{B\in\mathcal{B}(\mathbb{R})} |P_n(B) - P(B)| \qquad (n \to \infty)$$

は, 一般には 0 に収束しない. $\mathcal{B}(\mathbb{R})$ は集合族として大きすぎる. 集合族 $\mathcal{I}$ はコントロールできる程度の大きさ.

(13) 中心極限定理（CLT）の復習：$Z_1, Z_2, \ldots, Z_n$ を確率空間 $(\Omega, \mathscr{F}, \mathbb{P})$ 上の独立同一分布に従う確率変数列とし,

$$\mathbb{E}[Z_1] = \mu, \qquad \mathbb{V}[Z_1] = \mathbb{E}[(Z_1 - \mu)^2] = \sigma^2 \qquad (0 < \sigma < \infty).$$

このとき,

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

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.

(14) CLT より, 各 x に対して

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

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

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

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？
$\sqrt{n}\{\widehat{F}_n(x) - F(x)\}$ 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.

現代経験過程理論

(15) $\mathbb{R}$ 上の過程 $\sqrt{n}\{\widehat{F}_n(x) - F(x)\}$ を拡張する.
$\mathcal{G} := \{g : \mathbb{R} \to \mathbb{R}$ は適当な可測函数$\}$ とし,

$$Pg := \int_{\mathbb{R}} g(x)\, dP(x), \qquad P_n g := \frac{1}{n}\sum_{i=1}^{n} g(X_i).$$

経験過程

$$\sqrt{n}\{P_n g - Pg\}$$

を $\mathcal{G}$ 上の過程とみなす.

$g(t) = \mathbb{1}_{(-\infty,x]}(t)$ とすれば, $\sqrt{n}\{\widehat{F}_n(x) - F(x)\}$ が得られる.

また, $g(t) = \mathbb{1}_B(t)$ とすれば, $\sqrt{n}\{P_n(B) - P(B)\}$ が得られる.

うまくいく, いかないか (函数型の大数の強法則や中心極限定理をえることができる) は函数族 $\mathcal{G}$ の大きさ（豊かさ）に依存すること（計量エントロピー・VC 族）がしられている.

(16) 現代経験過程理論により, 母数モデル（統計モデルを添え字づける母数空間が有限次元）より複雑なセミ・パラメトリックモデル（母数空間が有限次元と無限次元の空間の直積でかけ, 推測のターゲットは有限次元母数で, 無限次元母数は擾乱母数として扱うことができるモデル）およびノンパラメトリックモデル（無限次元母数空間で推測のターゲットが無限次元母数：言葉が変であるが, 統計学の慣習に従う）において, 現代経験過程理論は強力である！また, 統計的機械学習理論の必須の道具である.

授業計画

- 8 月 25 日 4 コマ：導入
- 8 月 25 日 5 コマ：確率論の復習
- 8 月 26 日 2 コマ：古典的な大標本理論のまとめ
- 8 月 26 日 3 コマ：指数型確率不等式
- 8 月 26 日 4 コマ：計量エントロピーと Glivenko-Cantelli の定理（一様型大数法則）
- 8 月 27 日 2 コマ：VC 族, Donsker の定理：一様中心極限定理
- 8 月 27 日 3 コマ：M 推定量の漸近分布について
- 8 月 27 日 4 コマ：まとめとレポートの説明

参考文献

(a) 確率論に関わる文献

(a1) 清水泰隆：『統計学への確率論, その先へ』, 内田老鶴圃 (2019).
(a2) 服部哲弥：『確率変数の収束と大数の完全法則』, 共立出版 (2019).
(a3) 原啓介：『測度・確率・ルベーグ積分』, 講談社 (2017).
(a4) 吉田伸夫：『ルベーグ積分入門』, 日本評論社 (2021).

(b) 経験過程に関わる文献

(b1) Búhlmann, Peter and van der Geer, Sara: Statistics for high-dimensional data. Springer–Verlag(2011).
(b2) Pollard, David: Convergence of stochastic processes. Springer(1984).
(b3) van de Geer, Sara: Empirical processes in M-estimation. Cambridge Univ. Press(2000).
(b4) van de Geer, Sara: Empirical Process Theory. (accessed 2021/04/09) (2020). これが一番大事な文献.
(b5) van der Vaart, Aad.W.: Asymptotic Statistics, Cammbridge Univ. Press(1998).
(b6) van der Vaart, Aad W. and Wellner, Jon A.: Weak convergence and empirical processes. Springer(1996).

(c) 経験過程にかかわる日本語文献

(c1) 久保木久孝・鈴木武：『セミパラメトリック推測と経験過程』，朝倉出版 (2015).

(c2) 渡辺澄夫：『代数幾何と学習理論』，森北出版 (2006).

その他, 成書（日本語はあまりない）や web 上の lecture notes がごろごろある. 数理統計学や統計的学習理論のなかで一番使われている道具といっても過言ではない.

統計モデル

1. $(\Omega, \mathcal{F}, \mathbb{P})$ を確率空間.
2. $X : \Omega \longrightarrow \mathbb{R}$ は (generic) 実数値確率変数.
3. $P_0(B) := \mathbb{P}(X \in B)$ $(\forall B \in \mathcal{B}(\mathbb{R}))$: 真のモデル.
4. $\mathcal{P}$: 統計モデル（測度空間 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上の確率測度の集まり）. すなわち,

$$\mathcal{P} = \{P : \mathcal{B}(\mathbb{R}) \longrightarrow [0, 1] \text{ は確率測度 }\}.$$

5. $X_1, X_2, \ldots, X_n$ を P_0 からランダム標本 (独立同一に P_0 に従う実数値確率変数)
6. P_n は推定されたモデル. $B \in \mathcal{B}(\mathbb{R})$ に対して,

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

統計モデルの母数化

1. Θ: 母数空間, すなわち, 統計モデル $\mathcal{P}$ の要素（確率測度）を添え字付ける集合. Θ の元 θ を母数という.
2. 写像
$$\Theta \ni \theta \longmapsto P_\theta \in \mathcal{P}$$
は識別可能性であるとする. すなわち, $\forall \theta, \theta' \in \Theta$ に対して,
$$P_\theta(B) = P_{\theta'}(B) \; (\forall B \in \mathcal{B}(\mathbb{R}))$$
のとき, $\theta = \theta'$ とする. ただし, P_θ は確率測度とする.
3. 母数から統計モデルへの写像がよい性質をみたせば, 統計モデル $\mathcal{P}$ を調べることは Θ を調べることとだいたい考えてよい.

パラメトリック・セミパラメトリック・ノンパラメトリックモデル

(1) パラメトリックモデル＝母数空間 Θ が有限次元の統計モデル.
例：

$$\mathcal{P} = \left\{ \frac{dP}{dx} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2 \right); \right.$$
$$\left. (\mu, \sigma) \in (-\infty, \infty) \times (0, \infty) =: \Theta \right\}$$

(2) ノンパラメトリックモデル (言葉の矛盾！)= 母数空間 Θ が無限次元の統計モデル = 無限次元統計モデル. 分布形を仮定しないモデル.
(3) セミパラメトリックモデル: $\Theta = \Xi \times \Lambda$. ただし,

- Ξ は有限次元母数空間で推測の対象
- Λ は無限次元母数空間で直接の推測対象ではない. 局外母数空間.

例：Cox 比例ハザードモデル.