2021 年度　数理科学特別講義 C, G
大阪府立大学
8 月 27 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 27 日

内容

(1) Vapnik-Chervonenkis の定理 (定理 11.3)

(2) VC 次元, VC 族

(3) パッキング数

(4) VC 次元による計量エントロピーの評価
(定理 11.5)

(5) VC 族 ⇒ GC 族 (Gilvenko-Cantelli)

(6) 法則(分布)収束

(7) Donsker 族

記号

$(X, \mathcal{T})$：位相空間

$\Longleftrightarrow$ $\mathcal{T}$ は X の部分集合族で以下をみたす：

(1) $\phi, X \in \mathcal{T}$.

(2) $O_1, O_2 \in \mathcal{T} \Rightarrow O_1 \cap O_2 \in \mathcal{T}$.

(3) $\{O_\lambda\}_{\lambda \in \Lambda}$ を $\mathcal{T}$ の元からなる集合族とすると，

$$\bigcup_{\lambda \in \Lambda} O_\lambda \in \mathcal{T}.$$

注意　$\mathcal{T}$ の要素を開集合という。

$(\Omega, \mathcal{F}, \mathbb{P})$：確率空間

$(\mathcal{X}, \mathcal{T})$：位相空間

$\mathcal{B}(\mathcal{X})$：$\mathcal{T}$を含む最小のσ加法族

$$\mathcal{B}(\mathcal{X}) = \sigma(\mathcal{T})$$

$X : \Omega \to \mathcal{X}$は確率要素。すなわち，$\forall B \in \mathcal{B}(\mathcal{X})$に対し

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

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

$P = \mathbb{P} \circ X^{-1}$である。

$(\mathcal{X}, \mathcal{B}(\mathcal{X}), P)$：$X$によって誘導された確率空間.

定義11.2　$\mathcal{D}$を$\mathcal{X}$の部分集合の集まりとする．$n \in \mathbb{N}$と

$\{x_1, x_2, \ldots, x_n\} \in \mathcal{X}$に対し．

$$\Delta^{\mathcal{D}}(x_1, \ldots, x_n) := \#\{D \cap \{x_1, \ldots, x_n\} : D \in \mathcal{D}\} \leq 2^n.$$

ただし，集合Aに対し，#(A)はAの要素の個数

さらに．

$$VC^{\mathcal{D}}(n) := \sup\{\Delta^{\mathcal{D}}(x_1, \ldots, x_n) ; x_1, \ldots, x_n \in \mathcal{X}\}$$

例11.2　$\mathcal{X} = \mathbb{R}$とし．

$$\mathcal{D} = \{(-\infty, r] : r \in \mathbb{R}\}$$

としたとき．$\{x_1, \ldots, x_n\} \in \mathbb{R}$に対し．$\Delta^{\mathcal{D}}(x_1, \ldots, x_n) \leq n+1$.

定理 11.3 $X_1, \ldots, X_n$ を X の独立複製とする。

このとき、次の2つは同値：

(1) $\frac{1}{n} \log \Delta^{\mathcal{D}}(X_1, \ldots, X_n) \xrightarrow{P} 0 \quad (n \to \infty)$.

(2) $\sup_{D \in \mathcal{D}} | P_n(D) - P(D) | \xrightarrow{P} 0 \quad (n \to \infty)$

ただし、

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

$P(D) = \mathbb{P}(X \in D)$.

$X_1, \ldots, X_n$ は $\mathcal{X}$ 値確率要素。

$D \subset \mathcal{X}$. $\mathcal{D}$: $\mathcal{X}$ の部分集合族

(1) ⟹ (2) の証明

$$\mathcal{G} := \{ \mathbb{I}_D : \mathcal{X} \longrightarrow \{0,1\} \; ; \; D \in \mathcal{D} \}$$

とおく。すると

$$|g(x)| \leq 1 \quad (x \in \mathcal{X}). \qquad (a)$$

さらに、$\delta > 0$ に対し

$$N(\delta, \mathcal{G}, \|\cdot\|_{L^\infty(P_n)}) = \Delta^{\mathcal{D}}(X_1, \dots, X_n)$$

と

$$N(\delta, \mathcal{G}, \|\cdot\|_{L^1(P_n)}) \leq N(\delta, \mathcal{G}, \|\cdot\|_{L^\infty(P_n)})$$

より

$$\frac{1}{n} H(\delta, \mathcal{G}, \|\cdot\|_{L^1(P_n)}) \leq \frac{1}{n} \log \Delta^{\mathcal{D}}(X_1, \dots, X_n). \qquad (b)$$

定理11.2は $\mathbb{R}$ を $\mathcal{X}$ にかえても成立することに注意する.

(a)と(b)より, 定理11.2の条件をみたしているので

$$\sup_{D\in\mathcal{D}} |P_n(D)-P(D)| = \sup_{g\in\mathcal{G}} |(P_n-P)g|$$

$$= \|P_n-P\|_{\mathcal{G}} \xrightarrow{P} 0 \quad (n\to\infty).$$

□

定義 11.3

$\mathcal{D}$ は VC族

$\Longleftrightarrow$

ある定数 $C>0$ と $V>0$ が存在して，$^\forall n \in \mathbb{N}$ に対して，

$$VC^{\mathcal{D}}(n) \leq C n^V$$

が成り立つ．

ただし，

$$VC^{\mathcal{D}}(n) = \sup\{\Delta^{\mathcal{D}}(x_1,\ldots,x_n);\ x_1,\ldots,x_n \in \mathcal{X}\}$$

$$\Delta^{\mathcal{D}}(x_1,\ldots,x_n) = \#\{D\cap\{x_1,\ldots,x_n\};\ D\in\mathcal{D}\} \leq 2^n.$$

定理 11.4　$\mathcal{D}$ が VC 族ならば

$$\sup_{D \in \mathcal{D}} |P_n(D) - P(D)| \xrightarrow{P} 0 \quad (n \to \infty).$$

証明

$$\mathcal{G} = \{ \mathbb{I}_D(x) \ (x \in \mathcal{X}) \ ; \ D \in \mathcal{D} \}$$

とおく。$\delta > 0$ に対して，（V は定義 11.3）

$$\frac{1}{n} H(\delta, \mathcal{G}, \|\cdot\|_{L^1(P_n)}) \le \frac{1}{n} \log \Delta^{\mathcal{D}}(X_1, \dots, X_n)$$

$$\le \log \mathrm{VC}^{\mathcal{D}}(n) \le \frac{1}{n} \log (C n^V)$$

$$= V \frac{\log n}{n} + \frac{C}{n} \to 0 \quad (n \to \infty).$$

定理 11.2 よりわかる。（定理 11.3 でもよい）. □

No.9

定義 11.4　集合族 $\mathcal{D}$ の VC 次元 $VC^{\mathcal{D}}$ とは

$$VC^{\mathcal{D}} := \inf\{n \in \mathbb{N} ; VC^{\mathcal{D}}(n) < 2^n\}.$$

このような n がないとき，$VC^{\mathcal{D}} = \infty$.

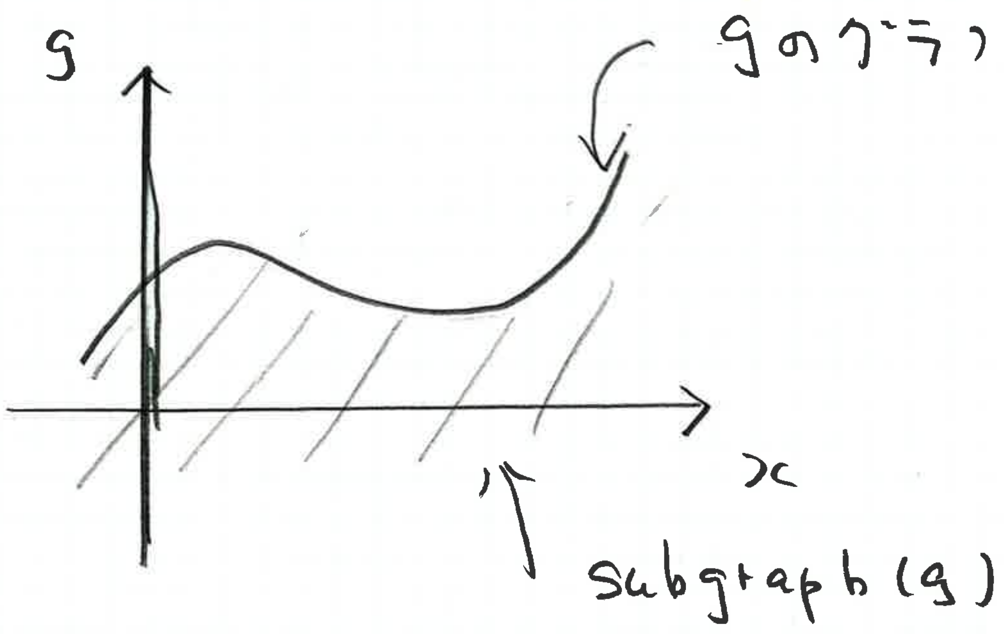
補題 11.3　次は同値

(1)　$\mathcal{D}$ は VC 族.

(2)　$VC^{\mathcal{D}} < \infty$,

定義11.5 函数 $g : \mathcal{X} \longrightarrow \mathbb{R}$ のサブグラフとは　　No.10

$$\mathrm{subgraph}(g) := \{ (x, t) \in \mathcal{X} \times \mathbb{R} \,;\, g(x) \geq t \}$$



$\mathcal{X} = \mathbb{R}$

$\mathcal{G}$ は VC族 $\iff$ $\{ \mathrm{subgraph}(g) \,;\, g \in \mathcal{G} \}$ は VC族.

定義 11.6 S を（準）距離空間 (M, d_M) の空でない部分集合とする．$\forall \delta > 0$ に対して，S の δ パッキング数 $D(\delta, S, \|\cdot\|_{d_M})$ は次をみたす最大の自然数 N である：

S の元 $\{A_1, \ldots, A_N\}$ が存在して，

$$d_M(s_k, s_j) > \delta \quad (\forall k \neq j,\ k, j = 1, \ldots, N).$$

ただし，$\|x\|_{d_M}^2 = d_M(x, x)$．（定理 11.5 の証明で使う）

補題 11.4 $\delta > 0$ とする．

$$N(\delta, S, \|\cdot\|_{d_M}) \leq D(\delta, S, \|\cdot\|_{d_M}) \leq N\left(\frac{\delta}{2}, S, \|\cdot\|_{d_M}\right).$$

Q : $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ 上の確率測度

$\mathcal{G} := \{ g : \mathcal{X} \longrightarrow \mathbb{R} ;$ g は可測函数 $\}$

$N(\delta, \mathcal{G}, \|\cdot\|_{L^1(Q)})$: $\mathcal{G}$ の δ 被覆数 $(\delta > 0)$

$$\|g\|_{L^1(Q)} = \int_{\mathcal{X}} |g(x)| \, dQ(x).$$

$$G(x) = \sup_{g \in \mathcal{G}} |g(x)| \quad (x \in \mathcal{X})$$: 包絡函数

定理 11.5 $G \in L^1(Q)$ とし，関数族 $\mathcal{G}$ は

$$VC^{\mathcal{D}} = V < \infty, \quad \mathcal{D} = \{ \mathrm{subgraph}(g) ; g \in \mathcal{G} \}$$

とする．このとき，V に依存する定数 A が存在して，
$\forall \delta > 0$ に対して

$$N(\delta QG, \mathcal{G}, \|\cdot\|_{L^1(Q)}) \le \max\{ A\delta^{-2V}, e^{\delta/4} \}.$$

ただし，

$$QG = \int_{\mathcal{X}} G(x)\, dQ(x).$$

証明の要点　一般性を失わず：$G=1$.　No. 14

$\mathcal{D}=\{\mathrm{subgraph}(g)\subset \mathcal{X}\times\mathbb{R}:\ g\in\mathcal{G}\}$

とし，$\delta>0$ に対して，

$N=D(\delta,\ \mathcal{G},\ \|\cdot\|_{L^1(Q)})$

とおく．

$$n\geq\frac{4\log N}{\delta}\qquad (a)$$

のとき，$\mathcal{D}$ はVC族なので，定数 C が存在して，

$D(\delta,\ \mathcal{G},\ \|\cdot\|_{L^1(Q)})\leq C n^{V}$

がわかる．

しかし、

$$N \geq \exp\left\{\frac{\delta}{4}\right\} \quad \Leftarrow \quad (a)$$

のとき、

$$N \leq C\left(\frac{16V}{\delta}\right)^{V}\sqrt{N} \Longrightarrow \sqrt{N} \leq C(16V)^{V}\delta^{-V}$$

$$\Longrightarrow N \leq \underbrace{C^{2}(16V)^{2V}}_{=:A}\delta^{-2V}$$

よって、(a)が成立しないとき、$N < \exp\left\{\frac{\delta}{4}\right\}$

とあわせれば、

$$N(\delta, \mathcal{G}, L^{1}(Q)) \leq D(\delta, \mathcal{G}, L^{1}(Q)) = N$$

$$\leq \max\left\{A\delta^{-2V},\ e^{\delta/4}\right\} =: \text{bound}$$

$QG = 1$ から、 $N(\delta QG, \mathcal{G}, L^{1}(Q)) \leq \text{bound}$. □

系11.1 $\mathcal{G}$を函数族とし，その包絡函数Gが$PG<\infty$をみたし，函数族$\mathcal{G}$がVC族ならば：

$$\| \mathbb{P}_n - P \|_{\mathcal{G}} = \sup_{g \in \mathcal{G}} |(\mathbb{P}_n - P) g| \xrightarrow{P} 0 \quad (n \to \infty)$$

証明 定理11.5と定理11.2を使う。
$(\mathbb{P}_n = Q)$

まず，

$$|\mathbb{P}_n G - PG| \xrightarrow{P} 0$$

である。$\forall \varepsilon > 0$に対して，nを十分大きく取れば

$$\mathbb{P}(\mathbb{P}_n G \geq 2 \cdot PG) \leq \varepsilon \qquad (a)$$

となる。

$2PG \geqq P_nG$ 上では、定理11.5より、

$$\frac{1}{n}\log N(\delta PG, \mathcal{G}, \|\cdot\|_{L^1(P_n)})$$

$$\leq \frac{1}{n}\log N(\tfrac{\delta}{2} P_nG, \mathcal{G}, \|\cdot\|_{L^1(P_n)})$$

$$\leq \frac{1}{n}\log\left(\max\left\{A\left(\tfrac{\delta}{2}\right)^{-\delta}, e^{\delta/8}\right\}\right)$$

$$\longrightarrow 0 \quad (n \to \infty)$$

(a)と合せれば

$$\frac{1}{n}\log N(\delta PG, \mathcal{G}, \|\cdot\|_{L^1(P_n)}) \xrightarrow{P} 0 \quad (n \to \infty)$$

定理11.2より

$$\|P_n - P\|_{\mathcal{G}} \xrightarrow{P} 0 \quad (n \to \infty).$$

□

確率変数列の分布収束の復習

$\{X_n\}_{n\geq 1}, X$: $(\Omega, \mathcal{F}, \mathbb{P})$ 上の確率変数列

以下は同値：

(1) $\lim_{n\to\infty} \mathbb{P}(X_n \leq x) = \mathbb{P}(X \leq x)$

ただし，点 $x \in \mathbb{R}$ は $x \longmapsto \mathbb{P}(X \leq x)$ の連続点

(2) 任意の有界連続関数 $h : \mathbb{R} \to \mathbb{R}$ に対して，

$$\lim_{n\to\infty} \mathbb{E}[h(X_n)] = \mathbb{E}[h(X)].$$

CLTの復習

$Z_1, \ldots, Z_n$: $(\Omega, \mathcal{F}, \mathbb{P})$ 上で定義された i.i.d. 確率変数あり

で，$\mathbb{E}[Z_i] = \mu$，$\mathbb{V}[Z_i] = \sigma^2$ $(0 < \sigma < \infty)$

$(i = 1, \ldots, n)$.

いま，

$$\bar{Z}_n = \frac{1}{n}(Z_1 + \cdots + Z_n)$$

$$X_n = \frac{\sqrt{n}(\bar{Z}_n - \mu)}{\sigma}$$

$$X \sim N(0, 1)$$

任意の有界連続関数 $h : \mathbb{R} \to \mathbb{R}$ に対して

$$\lim_{n \to \infty} \mathbb{E}[h(X_n)] = \mathbb{E}[h(X)].$$

$\mathcal{K}$ 上の norm を $\|\cdot\|_K$ を

$$\|\cdot\|_K = \sup_{x\in\mathbb{R}} |f(x)|$$

で定める.

そうすると, $\|\cdot\|_K$ によって, $\mathcal{K}$ の位相 $\mathcal{T}_{\|\cdot\|_K}$ が定まる. $\mathcal{K}$ 上の Borel 集合族は

$$\mathcal{B}(\mathcal{K}) = \sigma[\mathcal{T}_{\|\cdot\|_K}]$$

である.

いま,

$$\mathcal{H} = \{f : \mathcal{K} \longrightarrow \mathbb{R} \ ; \ f \text{は連続函数}\}$$

すなわち, $\mathbb{R}$ の任意の開集合 $O \subset \mathbb{R}$ に対して,

$$f^{-1}(O) \in \mathcal{T}_{\|\cdot\|_K}.$$

ドンスカー族

No.21

記号

$\mathcal{G} := \{ g : \mathbb{R} \to \mathbb{R} ;\ g$は可測関数$\}$

$X : \Omega \to \mathbb{R}$: $(\Omega, \mathcal{F}, \mathbb{P})$上の確率変数

$X_1, X_2, \ldots, X_n$: Xの独立複製.

$P = \mathbb{P} \circ X^{-1}$.

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

$$Pg = \mathbb{E}[g(X)]$$

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

$\mathcal{G}$で添え字付けられた経験過程

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

ここで、$g(x)$ の分散を

$$\sigma^2(g) := Pg^2 - (Pg)^2$$

$g_1, g_2 \in \mathcal{G}$ としたとき、$g_1(x)$ と $g_2(x)$ の共分散を

$$\sigma(g_1, g_2) = P(g_1 g_2) - (Pg_1)(Pg_2)$$

と書く。

<u>CLT</u>　各 g に対し、$\sigma^2(g) < \infty$ のとき、

$$\nu_n(g) \xrightarrow{L} N(0, \sigma^2(g)) \qquad (n \to \infty).$$

有限個の $g_1, \dots, g_p \in \mathcal{G}$ に対して.

$$\begin{pmatrix} \nu_n(g_1) \\ \vdots \\ \nu_n(g_p) \end{pmatrix} \xrightarrow{\mathcal{L}} N_p(0, \Sigma_{g_1, \dots, g_p})$$

$$\Sigma_{g_1, \dots, g_p} = (\sigma(g_i, g_j))_{i,j=1,\dots,p}.$$

<u>定義 13.3</u> ν は関数族 $\mathcal{G}$ で添え字付けられた確率過程とする. 各 $p \in \mathbb{N}$ と各 $\{g_1, \dots, g_p\} \in \mathcal{G}$ に対して.

$$(\nu(g_1), \dots, \nu(g_p))^T \sim N_p(0, \Sigma_{g_1, \dots, g_r}).$$

のとき ν は $\mathcal{G}$ で添え字付けられた P ブラウン橋という.

$(\Omega, \mathcal{F}, \mathbb{P})$：確率空間

$\mathcal{G} = \{g : \mathbb{R} \to \mathbb{R} : g$ は有界$\}$

$X : \Omega \to \mathbb{R}$　確率変数

$X_1, \dots, X_n$：X の独立複製

$P = \mathbb{P} \circ X^{-1}$　　$Pg = \mathbb{E}[g(X)]$　$(g \in \mathcal{G})$

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

ν：$\mathcal{G}$ で添字付けられている P ブラウン橋

$\nu_n(g) = \sqrt{n}(P_n - P)$

定義13.4

$\mathcal{G}$はドンスカー族

$\Longleftrightarrow$

任意の $h \in \mathcal{H}$ に対して

$$\lim_{n \to \infty} \mathbb{E}[h(\nu_n)] = \mathbb{E}[h(\nu)].$$

ただし, $\ell^\infty(\mathcal{G}) = \{ k : \mathcal{G} \to \mathbb{R} \,;\, \|k\|_{\mathcal{G}} := \sup_{g \in \mathcal{G}} |k(g)| < \infty \}$

$\mathcal{H} = \{ h : \ell^\infty(\mathcal{G}) \ni k \mapsto h(k) \in \mathbb{R} \,;\, h$は有界連続$\}$,

hは有界 $\Longleftrightarrow$ $\sup_{k \in \ell^\infty(\mathcal{G})} |h(k)| < \infty$

hは連続 $\Longleftrightarrow$ 任意の開集合 $O \subset \mathbb{R}$ に対し,
$h^{-1}(O)$ は $(\ell^\infty(\mathcal{G}), \|\cdot\|_{\mathcal{G}})$ の開集合

定理 14.4 $\mathbb{R}$上の実数値函数の族 $\mathcal{G}$ はVC族で、 No.25

$\mathcal{G}$ の包絡函数 G は $L^2(P)$ に含まれるとする。このとき

$\mathcal{G}$ で添字付けられた経験過程

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

はドンスカー族である。

古典的経験過程

$(\Omega, \mathcal{F}, \mathbb{P})$：確率空間

$X : \Omega \to \mathbb{R}$：確率変数

$X_1, \dots, X_n$：Xの独立複製

$$P = \mathbb{P} \circ X^{-1}, \quad F(x) = \mathbb{P}(X \leq x) = P((-\infty, x]) \quad (\forall x \in \mathbb{R})$$

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(-\infty, x]}(X_i)$$

$$\mathcal{G} = \{ \mathbb{R} \ni t \longmapsto \mathbb{1}_{(-\infty, x]}(t) \in \{0, 1\} \,;\, x \in \mathbb{R} \}.$$

すると、

$$\mathbb{P}_n(g_x) = \hat{F}_n(x) \quad (\forall g_x = \mathbb{1}_{(-\infty, x]})$$

$\mathcal{G}$はVC族であり、包絡関数$G = 1$は$L^2(P)$に含まれるので、$\mathcal{G}$はドンスカー族

よって

$$\mathbb{E}[h(\nu_n)] \longrightarrow \mathbb{E}[h(\nu)]$$

ただし、h は $\mathcal{G}$ 上の任意の有界連続函数.

$g_{x_i} = \mathbb{1}_{(-\infty, x_i]}$, $x_i \in \mathbb{R}$ $(i=1,2,\dots,p)$, $x_1 \leq \cdots \leq x_p$

$$\begin{bmatrix} \nu(g_{x_1}) \\ \vdots \\ \nu(g_{x_p}) \end{bmatrix} \sim N_p\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \Sigma_{g_{x_1},\dots,g_{x_p}} \right)$$

$$\Sigma_{g_{x_1},\dots,g_{x_p}} = \left(F(x_i \wedge x_j)\{1 - F(x_i \vee x_j)\} \right)_{i,j=1,\dots,p}$$

$a \wedge b = \min(a,b)$, $a \vee b = \max(a,b)$