統計数学特別講義 (2025/06/21)

本日の講義内容

(1) 記号の復習

(2) Glivenko-Cantelli 族の定義

(3) 計量エントロピーの復習

(4) 十分条件の比較

(5) 括弧付きエントロピーによる GC 族の定理の証明

記号

$\mathcal{G}$: $\mathbb{R}$ 上の実数値関数の族

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

$(\Omega, \mathcal{A}, \Pr)$: 背景の確率空間

$X : \Omega \to \mathbb{R}$: 確率変数

$$Q(B) = \Pr(X \in B) \quad ({}^\forall B \in \mathcal{B}(\mathbb{R})).$$

$g \in \mathcal{G}$ に対して

$$\|g\|_\infty = \sup_{x \in \mathbb{R}} |g(x)|$$

$$\|g\|_p = \left(\int |g|^p \, dQ \right)^{1/p}$$

計量エントロピー・被覆数・ε網

- $(\mathbb{D}, d)$：擬距離空間

① $d(x,y) \geq 0$, ② $d(x,y) = d(y,x)$ ③ $d(x,z) \leq d(x,y) + d(y,z)$

$(\forall x, y, z \in \mathbb{D})$

- $x \in \mathbb{D}$ と $\varepsilon > 0$ に対して

$$B_d(x, \varepsilon) = \{ y \in \mathbb{D} ; d(x,y) < \varepsilon \}$$

- ε被覆．$\mathcal{C}$：半径 ε の球の集合族

$\mathcal{C}$ は $A \subset \mathbb{D}$ の ε被覆 $\Leftrightarrow$ $\bigcup_{C \in \mathcal{C}} C \supset A$.

$\mathrm{Cover}(A, \varepsilon)$: Aの ε 被覆の全体.

$\mathrm{Centers}(\mathcal{C}, \varepsilon)$: $\mathcal{C}$ の球の中心の集合

たとえば

$$\mathcal{C} = \bigcup_{j=1}^{N} B_d(x_j, \varepsilon) \text{ としたとき}$$

$$\mathrm{Centers}(\mathcal{C}, \varepsilon) = \{x_1, \dots, x_N\}.$$

被覆数

$$N(\varepsilon, A, d) = \min_{\mathcal{C} \in \mathrm{Cover}(A, \varepsilon)} \#(\mathrm{Centers}(\mathcal{C}, \varepsilon)).$$

$$H(\varepsilon, A, d) := \log N(\varepsilon, A, d)$$

$\uparrow$

Aの ε エントロピー

5

関数族 $\mathcal{G}$ に対して

$$d(f, g) = \left(\int |f-g|^p dQ\right)^{1/p} \quad (f, g \in \mathcal{G})$$

とおく.

$\mathcal{G}$ の ε エントロピー $H(\varepsilon, \mathcal{G}, d)$ を $H(\varepsilon, \mathcal{G}, \|\cdot\|_{L_p(Q)})$

または $H_p(\varepsilon, \mathcal{G}, Q)$ とおく.

$\|g\|_\infty = \sup_x |g(x)|$ に関するエントロピーを $H_\infty(\varepsilon, \mathcal{G})$

とおく

- $N_{p,B}(\varepsilon, \mathcal{G}, Q)$ を次の条件をみたす関数 $\{(g_j^L, g_j^R)\}_{j=1}^{N}$ の最小の個数
  - $\| g_j^L - g_j^R \|_{L_p(Q)} \le \varepsilon$.
  - $\forall g \in \mathcal{G}$ に対して、$\exists j_0 \in \{1, \dots, N\}$ s.t.

$$g_j^L(x) \le g(x) \le g_j^R(x) \qquad (x \in \mathbb{R}).$$

$$H_{p,B}(\varepsilon, \mathcal{G}, Q) := \log N_{p,B}(\varepsilon, \mathcal{G}, Q)$$

- 包絡関数

$$G(x) := \sup_{g \in \mathcal{G}} | g(x) |.$$

ε 網

$\{C_1, \dots, C_N\}$ は A の ε 網

$\iff$ ① $\|C_j - C_k\| \geq \varepsilon \ (j \neq k)$

② $\{C_1, \dots, C_N\}$ は包含関係による順序で最大

- $X_1, X_2, \dots, X_n \overset{\text{i.i.d}}{\sim} P$ とする. ただし. P は $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

上の確率測度

関数族 $\mathcal{G}$ は P-GC族

$$\iff \sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{j=1}^{n} g(X_j) - \int g \, dP \right| \xrightarrow{a.s.} 0 \ (n \to \infty).$$

定理6.16

$\forall \varepsilon > 0$ に対して

$$H_B(\varepsilon, \mathcal{G}, L_1(P)) < \infty \Rightarrow \mathcal{G} \text{ は } P\text{-GC 族}$$

注意

$H_{1,B}(\varepsilon, \mathcal{G}, P) \Rightarrow G \in L_1(P)$.

定理6.20

(本日の内容)

$$\hat{P}_n(B) = \frac{\#\{X_j \in B\}}{n}$$

- $G \in L_1(P)$

ランダムな量

- $\frac{1}{n} H(\varepsilon, \mathcal{G}, L_1(\hat{P}_n)) \xrightarrow{P} 0 \quad (n \to \infty)$

$\Rightarrow \mathcal{G}$ は P-GC 族