

第6章 経験過程の収束

6.1 導入

X, X_1, X_2, \dots, X_n を i.i.d. 確率変数とし, 2 次の積率は有限とする. すると標本平均 $\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j$ に対して, 大数の強法則と中心極限定理を適用すると, $n \rightarrow \infty$ のとき

$$\bar{X}_n \xrightarrow{\text{a.s.}} \mathbb{E}[X] \quad \text{と} \quad \sqrt{n}(\bar{X}_n - \mathbb{E}[X]) \rightsquigarrow \mathbf{N}(0, \sigma^2)$$

となる. ただし, σ^2 は X の分散である. したがって, \mathbb{R} 上の連続関数 g で $\mathbb{E}[g^2(X)] < \infty$ なるものに対して, $n \rightarrow \infty$ のとき

$$\frac{1}{n} \sum_{j=1}^n g(X_j) \xrightarrow{\text{a.s.}} \mathbb{E}[g(X)] \quad \text{と} \quad \frac{1}{\sqrt{n}} \sum_{j=1}^n (g(X_j) - \mathbb{E}[g(X)]) \rightsquigarrow \mathbf{N}(0, \sigma_g^2)$$

が成り立つ. ただし, σ_g^2 は確率変数 $g(X)$ の分散である. 統計的推測論の多くの文脈では, 関数 g 自身がランダムなものであり, 観測に依存することがある. 標本に基づく g の推定量 \hat{g}_n に対する経験過程 $\frac{1}{n} \sum_{j=1}^n \hat{g}_n(X_j)$ を考察する必要がある. すなわち, \mathcal{G} を実数値関数のある族としたとき, $g \in \mathcal{G}$ で添え字付けられた経験過程

$$\frac{1}{n} \sum_{j=1}^n \{g(X_j) - \mathbb{E}[g(X)]\}$$

の確率の一様な評価が必要となってくる.

6.1.1 \sup の可測性についての注意

以後では, \mathcal{G} を関数族としたとき, 次のような経験過程

$$\sup_{g \in \mathcal{G}} g(X) \tag{6.1}$$

の挙動を調べることになる. すると

$$\mathbb{E} \left[\sup_{g \in \mathcal{G}} g(X) \right]$$

の数学的な厳密な意味が問題になる. 実際, 一般には仮定 (6.1) の可測性が保障されなくなる. しかし, この困難を克服するためには 2 つのアプローチが知られている.

一つは

$$E^* \left[\sup_{g \in \mathcal{G}} g(X) \right] := \sup_{\tilde{\mathcal{G}} \subset \mathcal{G}} \left\{ E \left[\sup_{g \in \tilde{\mathcal{G}}} g(X); \tilde{\mathcal{G}} \text{ は有限集合} \right] \right\}$$

と定義することである. 上の流儀では, 有限集合に対しては \sup を取るだけなので, 可測性が保障される.

もう一つは

$$E^* \left[\sup_{g \in \mathcal{G}} g(X) \right] := \inf \left\{ E[U]; U : \Omega \rightarrow \mathbb{R} \text{ は } U \geq \sup_{g \in \mathcal{G}} g(X) \text{ なる確率変数} \right\}$$

と定義するアプローチである. もちろん, \mathcal{G} が有限集合であれば, ふたつの定義はともに通常の期待値に一致することがわかる. さらに, この講義録で扱う経験過程は漸近連続性¹を持ち, 関数族 \mathcal{G} は可分²である. したがって, $\sup_{g \in \mathcal{G}} g(X)$ の可測性が保障されることになることが知られている.

6.2 経験過程の例

6.2.1 教育と雇用

個人からなる母集団からの標本

$$X_1 = (Y_1, Z_1), X_2 = (Y_2, Z_2), \dots, X_n = (Y_n, Z_n)$$

を観測したとする. ただし, $Y_j \in \{0, 1\}$ ($j = 1, 2, \dots, n$) は個人 j が雇用されている場合は $Y_j = 1$, そうでない場合には $Y_j = 0$ である. また, $Z_j \in \mathbb{R}$ は教育年限である. ここでの興味は, 教育年限と雇用との関係を理解することである. そのために以下のような関数

$$g^*(z) = \Pr(Y = 1 | Z = z)$$

を想定する. ただし, Y と Z は generic な確率変数とする. 自然な仮定として, g^* は非減少関数とする. いま, g^* を含む関数族として

$$\Lambda_1 := \{g : \mathbb{R} \rightarrow (0, 1); g \text{ は非減少} \}$$

¹この定義については, 定義 ?? を参照のこと.

²まず, \mathcal{G} に対して, 位相を \sup ノルム $\|g\|_\infty := \sup |g(x)|$ で入れる. このとき, \mathcal{G} は稠密な可算部分集合を持つことである.

を考える. 関数 g^* に対する自然な推定量は最尤推定量である.

$$\hat{g}_n \in \arg \min_{g \in \Lambda_1} \left[\sum_{j=1}^n \left\{ Y_j \log g(Z_j) + (1 - Y_j) \log(1 - g(Z_j)) \right\} \right] \quad (6.2)$$

である.

ここで, \mathbf{Q} を確率変数 Z の母集団分布としたとき, 推定量 (6.2) の精度を評価する尺度として

$$\|\hat{g}_n - g^*\|_{L_2(\mathbf{Q})} = \left(\int_{\mathbb{R}} (\hat{g}_n(z) - g^*(z))^2 d\mathbf{Q}(z) \right)^{1/2}$$

を考えることができる. 後に説明する道具を用いると

$$\|\hat{g}_n - g^*\|_{L_2(\mathbf{Q})} = O_P\left(\frac{1}{n^{1/3}}\right)$$

となることがわかる.

目的関数の仮定として

$$\Lambda_2 := \left\{ g : \mathbb{R} \rightarrow (0, 1); 0 \leq \frac{dg}{dz} \leq M, g \text{ は concave} \right\}$$

を考えることもできる. この文脈では

$$\|\hat{g}_n - g^*\|_{L_2(\mathbf{Q})} = O_P\left(\frac{1}{n^{2/5}}\right)$$

となることが証明される.

最後に, 母数 $\theta \in \mathbb{R}$ で添え付けられた母数モデル

$$\Lambda_3 := \left\{ g : \mathbb{R} \rightarrow (0, 1); g(z) = g^*(\theta z), \theta \in \mathbb{R}, g^*(z) = \frac{e^z}{1 + e^z} \right\}$$

を考える. すると θ^* を真の母数³とし, $\hat{\theta}_n$ をその最尤推定量としたとき,

$$\|\hat{g}_n - g^*\|_{L_2(\mathbf{Q})} \leq C|\hat{\theta}_n - \theta^*| = O_P\left(\frac{1}{\sqrt{n}}\right)$$

となる. ただし, C は generic 定数である.

³すなわち,

$$g^*(z) = \frac{e^{\theta^* z}}{1 + e^{\theta^* z}}$$

が成立する.

6.2.2 Kullback-Leibler 偏差

$\mathbb{X}(\subset \mathbb{R})$ を標本空間とし, m を \mathbb{R} 上の Lebesgue 測度とする. 真の p.d.f. を含む母数空間 Θ の要素 $\theta \in \Theta$ で添え付けられた関数族 (Lebesgue 測度 m に関する p.d.f. の族)

$$\{p_\theta; \theta \in \Theta\}$$

を考える. 真の分布はこの関数族に含まれると仮定し, 真の p.d.f. に対応する母数を θ^* と書くことにする.

いま, p_{θ^*} からの i.i.d. 標本 X_1, X_2, \dots, X_n を観測したとする. このときの推定精度の評価尺度として, **Hellinger** 距離 h と呼ばれるものを考えよう.

$$h(p, q) = \left(\frac{1}{2} \int_{\mathbb{X}} (\sqrt{p} - \sqrt{q})^2 dm \right)^{1/2}, \quad (p, q \text{ は } \mathbb{X} \text{ 上の p.d.f.}).$$

Hellinger 距離は

$$KL(p, q) = \int_{\mathbb{X}} \log \left(\frac{p(x)}{q(x)} \right) p(x) dm(x) \quad (6.3)$$

によって定義される Kullback-Leibler 偏差で制御される.

注意 6.1. (6.3) の右辺の積分は, q が p に対して絶対連続のときは問題なく定義される. 一方, 絶対連続でない場合には, $KL(p, q) = \infty$ と約束する. \square

命題 6.2. 任意の p.d.f. p, q (Lebesgue 測度 m に関する p.d.f.) に対して, 以下が成立する.

- (1) $KL(p, q) \geq 0$.
- (2) $h^2(p, q) \leq \frac{1}{2} KL(p, q)$.

Proof. $v > 0$ に対して, 不等式

$$\log v \leq v - 1; \quad \frac{1}{2} \log v \leq \sqrt{v} - 1$$

が成立することに注意する. (1) の証明:

$$\begin{aligned}
 \text{KL}(\mathbf{p}, \mathbf{q}) &= \int \log\left(\frac{\mathbf{p}(x)}{\mathbf{q}(x)}\right) \mathbf{p}(x) \, d\mathbf{m}(x) \\
 &= - \int \log\left(\frac{\mathbf{q}(x)}{\mathbf{p}(x)}\right) \mathbf{p}(x) \, d\mathbf{m}(x) \\
 &\geq - \int \left\{ \frac{\mathbf{q}(x)}{\mathbf{p}(x)} - 1 \right\} \mathbf{p}(x) \, d\mathbf{m}(x) \\
 &= - \int \mathbf{q}(x) \, d\mathbf{m}(x) + \int \mathbf{p}(x) \, d\mathbf{m}(x) \\
 &= 0
 \end{aligned}$$

から示せた.

(2) の証明:

$$\begin{aligned}
 \frac{1}{2}\text{KL}(\mathbf{p}, \mathbf{q}) &= \int \frac{1}{2} \log\left(\frac{\mathbf{p}(x)}{\mathbf{q}(x)}\right) \mathbf{p}(x) \, d\mathbf{m}(x) \\
 &= - \int \frac{1}{2} \log\left(\frac{\mathbf{q}(x)}{\mathbf{p}(x)}\right) \mathbf{p}(x) \, d\mathbf{m}(x) \\
 &\geq - \int \left\{ \sqrt{\frac{\mathbf{q}(x)}{\mathbf{p}(x)}} - 1 \right\} \mathbf{p}(x) \, d\mathbf{m}(x) \\
 &= \int \mathbf{p}(x) \, d\mathbf{m}(x) - \int \sqrt{\mathbf{p}(x)\mathbf{q}(x)} \, d\mathbf{m}(x) \\
 &= \frac{1}{2} \left\{ \int \mathbf{p}(x) \, d\mathbf{m}(x) + \int \mathbf{q}(x) \, d\mathbf{m}(x) - \int 2\sqrt{\mathbf{p}(x)\mathbf{q}(x)} \, d\mathbf{m}(x) \right\} \\
 &= \frac{1}{2} \int \{ \sqrt{\mathbf{p}(x)} - \sqrt{\mathbf{q}(x)} \}^2 \, d\mathbf{m}(x) \\
 &= h^2(\mathbf{p}, \mathbf{q})
 \end{aligned}$$

から示せた. □

注意 6.3. 上の証明において, 積分領域は $\{x \in \mathbb{R}; \mathbf{p}(x) > 0\} =: A$ となっていることに注意せよ. さらに

$$\int_A \mathbf{p}(x) \, d\mathbf{m}(x) = \int_A \mathbf{p}(x) \, d\mathbf{m}(x) + \underbrace{\int_{A^c} \mathbf{p}(x) \, d\mathbf{m}(x)}_{=0} = \int_{\mathbb{R}} \mathbf{p}(x) \, d\mathbf{m}(x) = 1$$

である.

$\theta^* \in \Theta$ を真の母数とする. X_1, X_2, \dots, X_n を母集団分布 \mathbf{p}_{θ^*} からラ

ンダム標本としたとき, p_{θ^*} の最尤推定量は

$$\begin{aligned} p_{\hat{\theta}_n} &\in \arg \min_{\theta \in \Theta} \left\{ \sum_{j=1}^n \log \left(\frac{p_{\theta^*}(X_j)}{p_{\theta}(X_j)} \right) \right\} \\ &\Leftrightarrow p_{\hat{\theta}_n} \in \arg \max_{\theta \in \Theta} \left\{ \sum_{j=1}^n \log p_{\theta}(X_j) \right\} \end{aligned}$$

で与えられる. これは Kullback-Leibler 偏差の経験過程版と解釈できる. 最尤推定量の定義から

$$\begin{aligned} 0 &\geq \frac{1}{n} \sum_{j=1}^n \log p_{\theta^*}(X_j) - \frac{1}{n} \sum_{j=1}^n \log p_{\hat{\theta}_n}(X_j) = \frac{1}{n} \sum_{j=1}^n \log \left(\frac{p_{\theta^*}(X_j)}{p_{\hat{\theta}_n}(X_j)} \right) \\ &= \frac{1}{n} \sum_{j=1}^n \log \left(\frac{p_{\theta^*}(X_j)}{p_{\hat{\theta}_n}(X_j)} \right) - \text{KL}(p_{\theta^*}, p_{\hat{\theta}_n}) + \text{KL}(p_{\theta^*}, p_{\hat{\theta}_n}) \end{aligned}$$

となる. したがって

$$\begin{aligned} \text{KL}(p_{\theta^*}, p_{\hat{\theta}_n}) &\leq \left| \frac{1}{n} \sum_{j=1}^n \log \left(\frac{p_{\theta^*}(X_j)}{p_{\hat{\theta}_n}(X_j)} \right) - \text{KL}(p_{\theta^*}, p_{\hat{\theta}_n}) \right| \\ &\leq \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{j=1}^n \log \left(\frac{p_{\theta^*}(X_j)}{p_{\theta}(X_j)} \right) - \text{KL}(p_{\theta^*}, p_{\theta}) \right| \end{aligned}$$

を得る. しかし, 任意の固定した $\theta \in \Theta$ に対して

$$\frac{1}{n} \sum_{j=1}^n \log \left(\frac{p_{\theta^*}(X_j)}{p_{\theta}(X_j)} \right) - \text{KL}(p_{\theta^*}, p_{\theta}) = O_P(n^{-1/2})$$

が成り立つこと⁴を知っている. このことから, 中心極限定理の一樣バージョンを導出できれば, $\text{KL}(p_{\theta^*}, p_{\hat{\theta}_n})$ の 0 への収束のスピードが評価できることが期待される.

6.3 計量エントロピー, 被覆数と ϵ 網

この章では, (\mathbb{D}, d) を一般の擬距離空間とする. ただし, 擬距離関数 d は, 「 $d(x, y) = 0 \Rightarrow x = y (\forall x, y \in \mathbb{D})$ 」を除いた距離関数の公理⁵をみたしている. $x \in \mathbb{D}$ と $\epsilon > 0$ に対して, 点 x を中心とした半径 ϵ の開球 $B_d(x, \epsilon)$ を

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

で定める.

⁴中心極限定理から用意に想像できる.

⁵任意の $x, y, z \in \mathbb{D}$ に対して, ① $d(x, y) \geq 0$; ② $d(x, y) = d(y, x)$; ③ $d(x, z) \leq d(x, y) + d(y, z)$, をみたす.

定義 6.4. (\mathbb{D}, d) を擬距離空間とし, $A \subset \mathbb{D}$ とする.

- 擬距離空間 \mathbb{D} における部分集合 A の半径 $\epsilon (> 0)$ の被覆とは, ϵ 球の有限集合族 \mathcal{C} で

$$\bigcup_{C \in \mathcal{C}} C \supset A$$

となるものである.

- A の ϵ 被覆集合族の全体を $\text{Cover}(A, \epsilon)$ と記す.
- 部分集合 A の ϵ 被覆 \mathcal{C} に対して, 被覆 \mathcal{C} で使われた ϵ 開球の中心の集合を $\text{Centers}(\mathcal{C}, \epsilon)$ と記すことにする.

(1) 部分集合 A に対する被覆数 $N(\epsilon, A, d)$ を A を被覆するために必要な ϵ 開球の個数の最小で定義する. すなわち

$$N(\epsilon, A, d) = \min_{\mathcal{C} \in \text{Cover}(A)} \#(\text{Centers}(\mathcal{C}, \epsilon))$$

である. ここで, 有限集合 B に対して, $\#(B)$ は B の元の個数である.

(2) $H(\epsilon, A, d) = \log N(\epsilon, A, d)$ を部分集合 A の ϵ エントロピーということにする.

(3) 部分集合 A は全有界であるとな, 任意の $\epsilon > 0$ に対して $H(\epsilon, A, d) < \infty$ が成立することである.

注意 6.5. 全有界な集合のみに興味があるので, ϵ 網の中心が A に含まれるかいは重要ではない. 実際, A の被覆 $\bigcup_{j=1}^N B_d(x_j, \epsilon) \cap A$ が存在すれば, $x'_j \in A$ をうまく取り, $\bigcup_{j=1}^N B_d(x'_j, 2\epsilon) \supset A$ とできることがわかる. \square

6.3.1 関数族のエントロピー

\mathbb{R} 上の確率測度を \mathbb{Q} とし, $1 \leq p < \infty$ とする. このとき, 実数値関数の集合 \mathcal{G} に対する距離 d を

$$d(f, g) = \|f - g\|_{L_p(\mathbb{Q})} = \left(\int_{\mathbb{R}} |f - g|^p d\mathbb{Q} \right)^{1/p} \quad (f, g \in L_p(\mathbb{R}))$$

で定める. 距離 d に関する関数族 \mathcal{G} の $\epsilon (> 0)$ エントロピーを $H_p(\epsilon, \mathcal{G}, d)$ を $H(\epsilon, \mathcal{G}, \mathbb{Q})$ や $H(\epsilon, \mathcal{G}, \|\cdot\|_{p, \mathbb{Q}})$ とも記すことにする. この場合には, \mathcal{G} は距離空間

$$L_p(\mathbb{Q}) := \{g : \mathbb{R} \rightarrow \mathbb{R}; g \text{ は可測関数で } \int_{\mathbb{R}} |g|^p d\mathbb{Q} < \infty\}$$

に含まれる. さらに, $\|g\|_{\infty} = \sup_{x \in \mathbb{R}} |g(x)|$ に関するエントロピーを $H_{\infty}(\epsilon, \mathcal{G})$ と記すことにする.

$N(\epsilon, A, \text{距離関数})$ と $N(\epsilon, A, \text{ノルム})$ 等の記法を統一する必要あり.

(2025/02/13 記)

定義 6.6. \mathcal{G} を \mathbb{R} 上の実数値関数のある部分集合族とし, \mathbb{Q} を \mathbb{R} 上の確率測度とする. さらに, $\epsilon > 0$ とする. このとき, $N_{p,B}(\epsilon, \mathcal{G}, \mathbb{Q})$ で次の条件をみたす関数の組 $\{(g_j^L, g_j^R)\}_{j=1}^N$ の最小の個数で定義する.

- すべての $j = 1, 2, \dots, N$ に対して

$$\|g_j^L - g_j^R\|_{L_p(\mathbb{Q})} \leq \epsilon.$$

- すべての $g \in \mathcal{G}$ に対して, ある $j \in \{1, 2, \dots, N\}$ が存在して

$$g_j^L(x) \leq g(x) \leq g_j^R(x) \quad (x \in \mathbb{R})$$

が成り立つ.

$H_{p,B}(\epsilon, \mathcal{G}, \mathbb{Q}) = \log N_{p,B}(\epsilon, \mathcal{G}, \mathbb{Q})$ を関数族 \mathcal{G} の括弧付き ϵ エントロピーとよぶ.

次の命題は異なるエントロピー間の関係を述べたものである.

命題 6.7. (1)

$$H_p(\epsilon, \mathcal{G}, \mathbb{Q}) \leq H_{p,B}(\epsilon, \mathcal{G}, \mathbb{Q})$$

(2)

$$H_{p,B}(\epsilon, \mathcal{G}, \mathbb{Q}) \leq H_\infty\left(\frac{\epsilon}{2}, \mathcal{G}\right)$$

(3) $A \subset \mathbb{D}$ とし, d, d' を \mathbb{D} 上の擬距離で

$$d(x, y) \leq d'(x, y) \quad (\forall x, y \in \mathbb{D})$$

が成立するとする. このとき

$$H(\epsilon, A, d) \leq H(\epsilon, A, d')$$

が成り立つ.

Proof. (1) $\epsilon > 0$ と $1 \leq p < \infty$ とし, \mathbb{Q} を \mathbb{R} 上の確率測度とする. 関数族 \mathcal{G} に対して, 以下が成立する. $\log N := H_{p,B}(\epsilon, \mathcal{G}, \mathbb{Q})$ とし, $\{(g_j^R, g_j^L)\}_{j=1}^N$ を関数の組とする. すると $\forall g \in \mathcal{G}$ に対して, $\exists j_0 \in \{1, 2, \dots, N\}$ があって

$$g_{j_0}^L(x) \leq g(x) \leq g_{j_0}^R(x) \quad \text{かつ} \quad \|g_{j_0}^L - g_{j_0}^R\|_{L_p(\mathbb{Q})} \leq \epsilon$$

なので

$$\|g - g_{j_0}^R\|_{L_p(\mathcal{Q})} \leq \epsilon$$

となるので, $\{g_1^R, g_2^R, \dots, g_N^R\}$ は網となる. よって

$$H_p(\epsilon, \mathcal{G}, \mathcal{Q}) \leq H_{p,B}(\epsilon, \mathcal{G}, \mathcal{Q})$$

がわかる.

(2) の証明: $\log N := H_\infty(\epsilon/2, \mathcal{G})$ とし, $\{g_1, g_2, \dots, g_N\}$ を関数の組とする. このとき

$$g_j^L(x) := g_j(x) - \frac{\epsilon}{2}, \quad g_j^R(x) := g_j(x) + \frac{\epsilon}{2}$$

とおく. すると $\forall g \in \mathcal{G}$ に対して, $\exists j_0 \in \{1, 2, \dots, N\}$ があって

$$g_{j_0}^R(x) \leq g(x) \leq g_{j_0}^L(x)$$

となるので

$$H_{p,B}(\epsilon, \mathcal{G}, \mathcal{Q}) \leq H_\infty\left(\frac{\epsilon}{2}, \mathcal{G}\right)$$

がわかる.

(3) の証明: $\log N := H(\epsilon, A, d')$ とし, g_1, g_2, \dots, g_N を関数の組とする. $\forall g \in \mathcal{G}$ に対して, $\exists j_0 \in \{1, 2, \dots, N\}$ があって

$$d(g, g_{j_0}) \leq d'(g, g_{j_0}) \leq \epsilon$$

となるので

$$H(\epsilon, A, d) \leq \log N$$

がわかる. □

6.3.2 ϵ 網

定義 6.8. (\mathbb{D}, d) を擬距離空間とし, $\|x\| = \sqrt{d(x, x)}$ ($x \in \mathbb{D}$) と記す. $\epsilon > 0$ とする. 空でない部分集合 $A \subset \mathbb{D}$ の ϵ 網とは A の有限部分集合 $\{c_1, c_2, \dots, c_N\}$ で

- 任意の $j \neq k$ に対して, $\|c_j - c_k\| \geq \epsilon$,
- 集合 $\{c_1, c_2, \dots, c_N\}$ は包含関係による順序に関して最大である.

命題 6.9. 部分集合 $A \subset \mathbb{D}$ の ϵ 網 $\{c_1, c_2, \dots, c_N\}$ は A の被覆 \mathcal{C} の中心とすることができる.

Proof. $1 \leq j \leq N$ とする. c_j を中心とする半径 ϵ の球を $B_j := \{x \in \mathbb{D}; \|x - c_j\| < \epsilon\}$ とする. すると $\bigcup_{j=1}^N B_j \supset A$ となる. このことを背理法で示す. そのために, ある $c_0 \in A$ が存在して, $c_0 \notin \bigcup_{j=1}^N B_j$ と仮定する. すると $\{c_1, c_2, \dots, c_N, c_0\}$ も部分集合 A の ϵ 網となる. $\#\{c_1, c_2, \dots, c_N, c_0\} = N + 1$ となるので, $\{c_1, c_2, \dots, c_N\}$ が部分集合 A の ϵ であることと矛盾する. \square

補題 6.10. d を \mathbb{R}^d の Euclid の距離とし, 0_d を \mathbb{R}^d の原点とし, $R > 0$ とする. $A = B_d(0_d, R) \subset \mathbb{R}^d$ に対して,

$$N(\epsilon, A, d) \leq \left(\frac{2R + \epsilon}{\epsilon} \right)^d$$

が成立する.

Proof. $\{c_j\}_{j=1}^N$ を部分集合 A の ϵ 網とする. 命題 6.9 から

$$N(\epsilon, A, d) \leq N \quad \text{かつ} \quad \bigcup_{j=1}^N B_d(c_j, \epsilon) \supset A$$

となる. $j \neq k$ ($j, k \in \{1, 2, \dots, N\}$) に対して, $\|c_j - c_k\| \geq \epsilon$ なので

$$B_d\left(c_j, \frac{\epsilon}{2}\right) \cap B_d\left(c_k, \frac{\epsilon}{2}\right) = \emptyset \quad (6.4)$$

となる. また

$$\bigcup_{j=1}^N B_d\left(c_j, \frac{\epsilon}{2}\right) \subset B_d\left(0_d, R + \frac{\epsilon}{2}\right) \quad (6.5)$$

であることがわかる. すると (6.4) に注意して (6.5) の両辺の体積を比較すると

$$N \text{vol}(B_d(0_1, 1)) \left(\frac{\epsilon}{2}\right)^d \leq \text{vol}(B_d(0_d, 1)) \left(R + \frac{\epsilon}{2}\right)^d \quad (6.6)$$

を得る. ただし, $\text{vol}(B_d(0_d, 1))$ は \mathbb{R}^d の単位球の体積で $\frac{(2\pi)^{d/2}}{d\Gamma(d/2)}$ で与えられる.

(6.6) の両辺を $\left(\frac{\epsilon}{2}\right)^d$ で割って整理すると

$$N \leq \left(\frac{2R + \epsilon}{\epsilon} \right)^d$$

を得る. \square

問 6.1. (6.4) と (6.5) を確認せよ.

例 6.11. □

例 6.12.

$$\mathcal{G} := \{g : [0, 1] \rightarrow [0, 1] \text{ s.t. } \sup_{x \in [0, 1]} |\dot{g}(x)| \leq 1\}$$

とする. このとき, ある $A > 0$ が存在して

$$H_\infty(\epsilon, \mathcal{G}) \leq \frac{A}{\epsilon} \quad (\epsilon > 0) \quad (6.7)$$

となる. (6.7) を証明するために, $N\epsilon > 1$ となるように $N \in \mathbb{N}$ を取り, $a_k = k\epsilon$ ($k = 0, 1, \dots, N-1$), $a_N = 1$ とおく. $B_k = (a_{k-1}, k)$ ($k = 1, 2, \dots, N$) とおき関数 $\tilde{g} : [0, 1] \rightarrow [0, 1]$ を

$$\tilde{g}(x) = \sum_{k=1}^N \epsilon \left\lfloor \frac{g(a_k)}{\epsilon} \right\rfloor$$

で定める. ただし, $[a]$ は $a \in \mathbb{R}$ の整数部分である. $\sup_{x \in [0, 1]} |\dot{g}(x)| \leq 1$ から

$$\sup_{x \in [0, 1]} |g - \tilde{g}| \leq 2\epsilon \quad (6.8)$$

となる. また, \tilde{g} の構成から $\{k\epsilon\}_{k=0}^{N-1} \cup \{1\}$ に値をとる. さらに

$$\begin{aligned} & |\tilde{g}(a_k) - \tilde{g}(a_{k-1})| \\ & \leq |\tilde{g}(a_k) - g(a_k)| + |g(a_k) - g(a_{k-1})| + |\tilde{g}(a_{k-1}) - g(a_{k-1})| \\ & \leq |\tilde{g}(a_k) - g(a_k)| + \left| \int_{a_{k-1}}^{a_k} \dot{g}(t) \, dm(t) \right| + |\tilde{g}(a_{k-1}) - g(a_{k-1})| \\ & \leq |\tilde{g}(a_k) - g(a_k)| + \int_{a_{k-1}}^{a_k} |\dot{g}(t)| \, dm(t) + |\tilde{g}(a_{k-1}) - g(a_{k-1})| \\ & \leq |\tilde{g}(a_k) - g(a_k)| + \int_{a_{k-1}}^{a_k} dm(t) + |\tilde{g}(a_{k-1}) - g(a_{k-1})| \\ & \leq |\tilde{g}(a_k) - g(a_k)| + (a_k - a_{k-1}) + |\tilde{g}(a_{k-1}) - g(a_{k-1})| \\ & \leq 3\epsilon \end{aligned} \quad (6.9)$$

となる. $\tilde{g}(a_0)$ の取りうる点の個数は $\lfloor \epsilon^{-1} \rfloor + 1$ である. (6.9) から $\tilde{g}(a_1)$ の取りうる値の点は $\{\tilde{g}(a_0) + k\epsilon\}_{k=-3}^3$ の 7 点である. 以上から, \tilde{g} の取りうる点の数は最大

$$(\lfloor \epsilon^{-1} \rfloor + 1) 7^{\lfloor \epsilon^{-1} \rfloor}$$

である. したがって, 上の式と (6.8) から, ある $A' (> 0)$ が存在して

$$H_\infty(2\epsilon, \mathcal{G}) \leq \log(\lfloor \epsilon^{-1} \rfloor + 1) 7^{\lfloor \epsilon^{-1} \rfloor} \leq \frac{1}{\epsilon} \log 7 + \log\left(\frac{1}{\epsilon} + 1\right) \leq \frac{A'}{\epsilon}$$

となることがわかる. よって

$$H_\infty(\epsilon, \mathcal{G}) \leq \frac{2A'}{\epsilon} =: \frac{A}{\epsilon}$$

が示せた. □

例 6.13. □

例 6.14. $L > 0$ を定数とる. 区間 $\mathcal{I} \subset \mathbb{R}$ 上の実数値関数 $g: \mathcal{I} \rightarrow \mathbb{R}$ が Lipschitz 係数 $L (> 0)$ の Lipschitz 関数とする. 任意の $x, x' \in \mathcal{I}$ に対し,

$$|g(x) - g(x')| \leq L|x - x'|$$

をみたすことである.

一般性を失うことなく, $\mathcal{I} = (0, 1]$ としよ. さらに,

$$\mathcal{G} := \{g: (0, 1] \rightarrow [0, 1]; g \text{ Lipschitz 係数 } 1 \text{ の Lipschitz 関数}\}$$

とする.

$\epsilon > 0$ を固定し, 閉区間 $[0, 1]$ を $N \leq 1 + \frac{1}{\epsilon}$ 個の区間 $(a_{j-1}, a_j]$ ($j = 1, 2, \dots, N$) で分割する. ただし

$$a_j - a_{j-1} \leq \epsilon \quad (j = 1, 2, \dots, N)$$

である. 任意の $g \in \mathcal{G}$ と $x \in (a_{j-1}, a_j]$ ($j = 1, 2, \dots, N$) に対して

$$\frac{\tilde{g}(x)}{\epsilon} = \lfloor \frac{g(a_j)}{\epsilon} \rfloor$$

とする. ただし, $r \in \mathbb{R}$ に対して, $\lfloor r \rfloor$ を r を越えない最大の整数とする.

このとき, $x \in (a_{j-1}, a_j]$ ($j = 1, 2, \dots, N$) に対して

$$\tilde{g}(a_j) - \epsilon < \lfloor \frac{g(a_j)}{\epsilon} \rfloor \epsilon < \tilde{g}(a_j)$$

なので

$$g(a_j) - g(x) - \epsilon < \lfloor \frac{g(a_j)}{\epsilon} \rfloor \epsilon - g(x) - \epsilon < \tilde{g}(x) - g(x) < g(a_j) - g(x)$$

となる. よって, $x \in (a_{j-1}, a_j]$ ($j = 1, 2, \dots, N$) に対して

$$|g(x) - \tilde{g}(x)| \leq |g(x) - g(a_j)| + \epsilon \leq 2\epsilon$$

を得る.

$$\lfloor \frac{g(a_1)}{\epsilon} \rfloor, \lfloor \frac{g(a_2)}{\epsilon} \rfloor, \dots, \lfloor \frac{g(a_N)}{\epsilon} \rfloor$$

に対して, 最大 $\left(1 + \frac{1}{\epsilon}\right)$ の選択があり, $j \in \{1, 2, \dots, N-1\}$ を固定すると

$$\begin{aligned} \left| \lfloor \frac{g(a_{j+1})}{\epsilon} \rfloor - \lfloor \frac{g(a_j)}{\epsilon} \rfloor \right| &\leq \left| \lfloor \frac{g(a_{j+1})}{\epsilon} \rfloor - \frac{g(a_{j+1})}{\epsilon} \right| \\ &\quad + \left| \frac{g(a_j)}{\epsilon} - \lfloor \frac{g(a_j)}{\epsilon} \rfloor \right| + \left| \frac{g(a_{j+1})}{\epsilon} - \lfloor \frac{g(a_{j+1})}{\epsilon} \rfloor \right| \\ &\quad + \frac{|g(a_j) - g(a_{j+1})|}{\epsilon} \\ &\leq 2 + \frac{1}{\epsilon} |a_{j+1} - a_j| \leq 3 \end{aligned}$$

となる. 一度, $\lfloor \frac{g(a_j)}{\epsilon} \rfloor$ を選ぶと, $\lfloor \frac{g(a_{j+1})}{\epsilon} \rfloor$ の選び方は最大 7 通りある. したがって, $g \in \mathcal{G}^\epsilon$ が動くとき, 関数 \tilde{g} の個数の最大は

$$\left(1 + \frac{1}{\epsilon}\right) \underbrace{\times 7 \times 7 \times \dots \times 7}_{N-1} \leq \left(1 + \frac{1}{\epsilon}\right) 7^{1/\epsilon}.$$

よって, $0 < \epsilon < 1$ に対して,

$$H_\infty(\epsilon, \mathcal{G}) \leq \log \left(1 + \frac{1}{\epsilon}\right) + \frac{1}{\epsilon} \log 7.$$

□

6.4 括弧付きエントロピーによる Glivenko-Cantelli 族の定理

定義 6.15. P を \mathbb{R} 上の確率測度とし, $X, X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} P$ とする. 関数族 \mathcal{G} は P -Glivenko-Cantelli 族であるとは

$$\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{j=1}^n g(X_j) - \mathbb{E}[g(X)] \right| \xrightarrow{\text{a.s.}} 0$$

をみたすことである. P -Glivenko-Cantelli 族を簡単に P -GC 族ともいうことにする.

定理 6.16. P を \mathbb{R} 上の確率測度とする. \mathcal{G} を関数族とする. 任意の $\epsilon > 0$ に対して, $H_B(\epsilon, \mathcal{G}, \|\cdot\|_{L_1(P)}) < \infty$ が成立するとする. このとき, 関数族 \mathcal{G} は P -GC 族である.

Proof. $\epsilon > 0$ を固定する. 定理の仮定から $N := N_{1,B}(\epsilon, \mathcal{G}, \mathbf{P}) < \infty$ とおく. すると N 個の関数の組 $\{g_j^L, g_j^R\}_{j=1}^N$ が存在して

$$\max_{j=1,2,\dots,N} \|g_j^L - g_j^R\|_{L_1(\mathbf{P})} \leq \epsilon$$

となる. さらに, 任意の $g \in \mathcal{G}$ に対して, ある $\exists j_0 \in \{1, 2, \dots, N\}$ が存在して

$$g_{j_0}^L \leq g \leq g_{j_0}^R \Rightarrow g_{j_0}^L - g_{j_0}^R \leq g - g_{j_0}^R \leq 0 \Rightarrow |g_{j_0}^R - g| \leq |g_{j_0}^R - g_{j_0}^L|$$

となる. したがって

$$\begin{aligned} \int g \, d(\widehat{\mathbf{P}}_n - \mathbf{P}) &= \int g \, d\widehat{\mathbf{P}}_n - \int g \, d\mathbf{P} \\ &\leq \int g_{j_0}^R \, d\widehat{\mathbf{P}}_n - \int g \, d\mathbf{P} \\ &= \int g_{j_0}^R \, d(\widehat{\mathbf{P}}_n - \mathbf{P}) + \int (g_{j_0}^R - g) \, d\mathbf{P} \\ &\leq \int g_{j_0}^R \, d(\widehat{\mathbf{P}}_n - \mathbf{P}) + \int |g_{j_0}^R - g| \, d\mathbf{P} \\ &\leq \int g_{j_0}^R \, d(\widehat{\mathbf{P}}_n - \mathbf{P}) + \int |g_{j_0}^R - g_{j_0}^L| \, d\mathbf{P} \\ &\leq \int g_{j_0}^R \, d(\widehat{\mathbf{P}}_n - \mathbf{P}) + \epsilon \end{aligned} \quad (6.10)$$

となる. 同様に

$$\begin{aligned} \int g \, d(\widehat{\mathbf{P}}_n - \mathbf{P}) &= \int g \, d\widehat{\mathbf{P}}_n - \int g \, d\mathbf{P} \\ &\geq \int g_{j_0}^L \, d\widehat{\mathbf{P}}_n - \int g \, d\mathbf{P} \\ &= \int g_{j_0}^L \, d(\widehat{\mathbf{P}}_n - \mathbf{P}) + \int (g_{j_0}^L - g) \, d\mathbf{P} \\ &\geq \int g_{j_0}^L \, d(\widehat{\mathbf{P}}_n - \mathbf{P}) - \int |g_{j_0}^L - g| \, d\mathbf{P} \\ &\geq \int g_{j_0}^L \, d(\widehat{\mathbf{P}}_n - \mathbf{P}) - \int |g_{j_0}^L - g_{j_0}^R| \, d\mathbf{P} \\ &\geq \int g_{j_0}^L \, d(\widehat{\mathbf{P}}_n - \mathbf{P}) - \epsilon \end{aligned} \quad (6.11)$$

となる. (6.10) と (6.11) を合わせると

$$\begin{aligned} - \max_{j=1,2,\dots,N} \left| \int g_{j_0}^L \, d(\widehat{\mathbf{P}}_n - \mathbf{P}) \right| - \epsilon &\leq \int g \, d(\widehat{\mathbf{P}}_n - \mathbf{P}) \\ &\leq \max_{j=1,2,\dots,N} \left| \int g_{j_0}^R \, d(\widehat{\mathbf{P}}_n - \mathbf{P}) \right| + \epsilon \end{aligned} \quad (6.12)$$

を得る. $\{g_j^L, g_j^R\}$ の個数は有限個なので, 古典的な大数の強法則から

$$\begin{aligned} \max_{j=1,2,\dots,N} \left| \int g_j^L d(\widehat{P}_n - P) \right| &\xrightarrow{\text{a.s.}} 0 \quad (n \rightarrow \infty) \\ \max_{j=1,2,\dots,N} \left| \int g_j^R d(\widehat{P}_n - P) \right| &\xrightarrow{\text{a.s.}} 0 \quad (n \rightarrow \infty) \end{aligned}$$

となる. よって, 十分大きな n に対して

$$\Pr \left(\max_{j=1,2,\dots,N} \left| \int g_j^L d(\widehat{P}_n - P) \right| \leq \epsilon \right) = 1 \quad (6.13)$$

と

$$\Pr \left(\max_{j=1,2,\dots,N} \left| \int g_j^R d(\widehat{P}_n - P) \right| \leq \epsilon \right) = 1 \quad (6.14)$$

となる. (6.12) – (6.14) を合わせると十分大きな n に対して

$$\Pr \left(\sup_{g \in \mathcal{G}} \left| \int g d(\widehat{P}_n - P) \right| \leq 2\epsilon \right) = 1$$

が成立することがわかる. すなわち

$$\lim_{n \rightarrow \infty} \Pr \left(\sup_{g \in \mathcal{G}} \left| \int g d(\widehat{P}_n - P) \right| \leq 2\epsilon \right) = 1$$

がわかる. よって, 定理は証明された. \square

定義 6.17. 実数値関数のある族 \mathcal{G} に対し, 関数

$$G(x) := \sup_{g \in \mathcal{G}} |g(x)| \quad (x \in \mathbb{R})$$

を関数族 \mathcal{G} の封筒関数 (envelope) という.

補題 6.18. P を \mathbb{R} 上の確率測度, \mathcal{G} を関数族, G をその封筒関数とする. このとき, 任意の $\epsilon > 0$ に対して, $H_{1,B}(\epsilon, \mathcal{G}, P) < \infty$ ならば, $G \in L_1(P)$ である.

Proof. $\forall \epsilon > 0$ に対して, $H_{1,B}(\epsilon, \mathcal{G}, P) < \infty$ なので, 命題 6.7(1) から, $H_1(\epsilon, \mathcal{G}, P) < \infty$ である. よって, $(\mathcal{G}, \|\cdot\|_{L_1(P)})$ は全有界となる. すなわち, $N := H_{1,B}(\epsilon, \mathcal{G}, P)$ と書いた⁶とき, ある $\{g_1, g_2, \dots, g_N\} \subset \mathcal{G}$ が存在して,

$$\mathcal{G} \subset \bigcup_{j=1}^N B(g_j, \epsilon)$$

とできる. ここで, $L_1(P)$ は完備であることに注意すると, \mathcal{G} の閉包 $\text{cl}(\mathcal{G})$ も完備となる. よって, $\text{cl}(\mathcal{G})$ はコンパクトであること⁷がわかる. さ

⁶あとの都合で, $N := H_1(\epsilon, \mathcal{G}, P)$ とせずに, 上のように置いた.

⁷距離空間 \mathbb{X} の部分集合 A について, 以下は同値である.

らに, 写像 $\mathcal{G} \ni g \mapsto \|g\|_{L_1(\mathbb{P})}$ は連続なので, この写像による \mathcal{G} の像 $\{\|g\|_{L_1(\mathbb{P})}; g \in \mathcal{G}\} \subset \mathbb{R}$ もコンパクトとなる. よって, ある $R > 0$ が存在して

$$\{\|g\|_{L_1(\mathbb{P})}; g \in \mathcal{G}\} \subset [-R, R] \Rightarrow \sup_{g \in \mathcal{G}} \|g\|_{L_1(\mathbb{P})} \leq R$$

となる. いま, $\epsilon > 0$ を固定して, $\{g_j^L, g_j^R\}_{j=1}^N \subset \mathcal{G}$ をうまくとると, $\forall g \in \mathcal{G}$ に対して, ある $j \in \{1, 2, \dots, N\}$ が存在して

$$g_j^L \leq g \leq g_j^R \Rightarrow |g| \leq |g_j^L| + |g_j^R - g_j^L| \quad \text{かつ} \quad \|g_j^R - g_j^L\|_{L_1(\mathbb{P})} \leq \epsilon$$

とできる. このことから

$$\begin{aligned} \int G(x) d\mathbb{P}(x) &= \int \sup_{g \in \mathcal{G}} |g(x)| d\mathbb{P}(x) \\ &\leq \max_{j \in \{1, 2, \dots, N\}} \int \{|g_j^L(x)| + |g_j^R(x) - g_j^L(x)|\} d\mathbb{P}(x) \\ &\leq \sum_{j=1}^N \left\{ \underbrace{\int |g_j^L(x)| d\mathbb{P}(x)}_{\leq R} + \underbrace{\int |g_j^R(x) - g_j^L(x)| d\mathbb{P}(x)}_{\leq \epsilon} \right\} d\mathbb{P} \\ &\leq N(R + \epsilon) < \infty \end{aligned}$$

がわかる. よって, 主張は証明された. \square

命題 6.19. \mathbb{P} を \mathbb{R} 上の確率測度, \mathcal{G} を関数族, G をその封筒関数とする. 関数族 \mathcal{G} は \mathbb{P} -GC 族で $L_1(\mathbb{P})$ 有界ならば, $G \in L_1(\mathbb{P})$ である.

Proof. 標本 X_1, X_2, \dots, X_n に基づく経験測度を $\hat{\mathbb{P}}_n$ と書き, 標本 X_1, X_2, \dots, X_{n-1} に基づく経験測度を $\hat{\mathbb{P}}_{n-1}$ と書くことにする. 関数族 \mathcal{G} は \mathbb{P} -GC 族なので

$$\frac{1}{n} \sup_{g \in \mathcal{G}} |\hat{\mathbb{P}}_n g - \mathbb{P}g| \xrightarrow{\text{a.s.}} 0$$

となる.

$$\begin{aligned} \frac{1}{n} |g(X_n) - \mathbb{P}g| &= \frac{1}{n} |n\hat{\mathbb{P}}_n - (n-1)\hat{\mathbb{P}}_{n-1} - \mathbb{P}g| \\ &\leq |\hat{\mathbb{P}}_n - \mathbb{P}g| + \frac{n-1}{n} |\hat{\mathbb{P}}_{n-1} - \mathbb{P}g| \end{aligned}$$

(1) A はコンパクト.

(2) A の任意の点列は A の中に収束する部分列を持つ.

(3) A は \mathbb{X} の部分集合として完備かつ全有界.

[12, p.48] を参照のこと.

と書き直すことができる. したがって

$$\sup_{g \in \mathcal{G}} \frac{1}{n} |g(X_n) - \mathbb{P}g| \leq \sup_{g \in \mathcal{G}} |\hat{\mathbb{P}}_n - \mathbb{P}g| + \frac{n-1}{n} \sup_{g \in \mathcal{G}} |\hat{\mathbb{P}}_{n-1} - \mathbb{P}g|$$

$$\xrightarrow{\text{a.s.}} 0$$

がわかる. よって

$$\Pr\left(\sup_{g \in \mathcal{G}} |g(X_n) - \mathbb{P}g| \geq n, \text{i.o.}\right) = 0 \quad (6.15)$$

となる. したがって, 補題 1.44 の対偶から

$$\sum_{n=1}^{\infty} \Pr(\sup_{g \in \mathcal{G}} |g(X_n) - \mathbb{P}g| \geq n) < \infty$$

がわかる. よって

$$\begin{aligned} \mathbb{E}\left[\sup_{g \in \mathcal{G}} |g(X_n) - \mathbb{P}g|\right] &= \int_0^{\infty} \Pr(\sup_{g \in \mathcal{G}} |g(X_n) - \mathbb{P}g| > t) dt \quad (\because \text{命題 1.31}) \\ &\leq \sum_{n=1}^{\infty} \Pr(\sup_{g \in \mathcal{G}} |g(X_n) - \mathbb{P}g| \geq n) < \infty \end{aligned} \quad (6.16)$$

となる. 任意の $g \in \mathcal{G}$ に対して, $\|g\|_{L_1(\mathbb{P})} < \infty$ であることと (6.16) から

$$\mathbb{E}[G] \leq \mathbb{E}\left[\sup_{g \in \mathcal{G}} |g(X_n) - \mathbb{P}g|\right] + \sup_{g \in \mathcal{G}} |\mathbb{P}g| < \infty$$

が示せた. □

6.5 対称化トリックと Dudley のエントロピー積分による Glivenko-Cantelli 族の定理

定理 6.20. \mathbb{P} を \mathbb{R} 上の確率測度, \mathcal{G} を関数族, G をその封筒関数とする. $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathbb{P}$ に対して, $\hat{\mathbb{P}}_n$ をこれらに基づく経験測度とする. すなわち

$$\hat{\mathbb{P}}_n(B) = \frac{\#\{j \in \{1, 2, \dots, n\}; X_j \in B\}}{n} \quad (B \in \mathcal{B}(\mathbb{R}))$$

である. $G \in L_1(\mathbb{P})$ かつ, $\forall \delta > 0$ に対して

$$\frac{1}{n} H(\delta, \mathcal{G}, \|\cdot\|_{L_1(\hat{\mathbb{P}}_n)}) \xrightarrow{\mathbb{P}} 0 \quad (n \rightarrow \infty) \quad (6.17)$$

が成り立つとき, \mathcal{G} は \mathbb{P} -Glivenko-Cantelli 族である.

Proof. この定理の証明の準備のための主張を述べた後に、証明を与える。ここでは、証明の道筋 ① と ② を解説する。

① 素朴な証明

Step 1. 対称化トリックによって、評価したい事象の確率を Rademacher 列を用いた表現で上限を与える。すなわち、任意の $g \in \mathcal{G}$ と $\delta > 0$ に対して

$$\Pr\left(\left|\int g d(\widehat{P}_n - P)\right| > \frac{\delta}{2}\right) \leq \frac{1}{2} \quad (6.18)$$

が成立すると

$$\Pr\left(\sup_{g \in \mathcal{G}} \left| \int g d(\widehat{P}_n - P) \right| > \delta\right) \leq 4\Pr\left(\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{j=1}^n \epsilon_j g(X_j) \right| > \frac{\delta}{4}\right) \quad (6.19)$$

が成立する。ただし、 X_j ($j = 1, 2, \dots, n$) とは独立な $\{\epsilon_j\}_{j=1}^n$ は Rademacher 列である。なお、(6.18) は Chebyshev の不等式を用いることで簡単に確認できることを注意しておく。

Step 2. \mathcal{G} を有限集合⁸としたとき、Hoeffding の不等式を X_1, X_2, \dots, X_n を条件付きにして、 $\frac{1}{n} \sum_{j=1}^n \epsilon_j g(X_j)$ に対して適用すると、 $t > 0$ に対して

$$\Pr\left(\max_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{j=1}^n g(X_j) \right| > K \sqrt{\frac{2(\log N + t)}{n}}\right) \leq e^{-t} \quad (6.20)$$

と評価である。さらに、Chebyshev の不等式を用いて (6.18) が成立することが確認できる。このときから (6.19) と (6.20) を用いて

$$\Pr\left(\max_{g \in \mathcal{G}} \left| \int g d(\widehat{P}_n - P) \right| > \frac{K}{4} \sqrt{\frac{2(\log N + t)}{n}}\right) \leq 8e^{-t} \quad (6.21)$$

を示す。ただし、 $N = \#(\mathcal{G})$ である。

Step 3. 定理 6.20 のエントロピー条件 (6.17) をみたく \mathcal{G} に対して

$$G(x) := \sup_{g \in \mathcal{G}} |g(x)| \leq K \quad (6.22)$$

を仮定して、 \mathcal{G} が P-GC 族であることを証明する。

Step 4. \mathcal{G} の封筒関数 G と $K > 0$ に対して、 $\mathcal{G}_K := \{g \mathbf{1}\{G \leq K\}\}$ と定める。すると Step 3 の議論から、 \mathcal{G}_K は P-GC 族であることがわかる。任意の $\delta > 0$ に対して、 $G \in L_1(P)$ だから、 K_0 をうまく取ると

$$\int_{G > K_0} G dP < \delta$$

⁸関数族 \mathcal{G} のとき、任意の $\delta > 0$ に対して、有限であるときは δ 網が存在する。この δ 網の有限個の中心に対して、ここで得られた結果を用いる。

とでき, $\{G\mathbb{1}\{G \geq K_0\}\} \cup \mathcal{G}_{K_0}$ も P-GC 族であることを利用して, (6.22) がなくとも \mathcal{G} は P-GC 族であることを最後に示す.

② chaining argument と Dudley のエントロピー積分を用いる証明

Step 1. 劣 Gauss を仮定して, 最大不等式を求める.

Step 2. chaining argument を用いて, 任意の 2 の関数 $g_1, g_2 \in \mathcal{G}$ に対して, 劣 Gauss 性が成立するときに, 2 つの差の sup が Dudley 積分で上から評価できることを証明 (定理 6.27) する.

Step 3. X_1, X_2, \dots, X_n を条件付けして, $\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{j=1}^n \epsilon_j g(X_j) \right|$ の期待値を Dudley のエントロピー積分を用いて上から評価する.

Step 4. Step 3 の評価を用いて, $\forall \delta > 0$ に対して

$$\frac{1}{n} H(\delta, \mathcal{G}, \|\cdot\|_{L_2(\hat{P}_n)}) \xrightarrow{P} 0 \quad (n \rightarrow \infty) \quad (6.23)$$

ならば, \mathcal{G} は P-GC 族であることを示す.

Step 5. $K > 0$ と封筒関数に対して, $\mathcal{G}_K := \{g\mathbb{1}\{G \leq K; g \in \mathcal{G}\}$ とおく. 関数族 \mathcal{G} は, 定理 6.20 の条件 (6.17) と条件 (6.22) をみたすとき, $\forall \delta > 0$ に対して

$$\frac{1}{n} H(\delta, \mathcal{G}_K, \|\cdot\|_{L_2(\hat{P}_n)}) \xrightarrow{P} 0 \quad (n \rightarrow \infty)$$

であることを確認する. すると Step 4 の議論から \mathcal{G}_K は P-GC 族であることがわかる.

Step 6. ① の Step 4 と同じ議論により, 定理の証明を完成させる. \square

注意 6.21. 以下の証明では, 定理の仮定のもと

$$\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{j=1}^n g(X_j) - \int g dP \right| \xrightarrow{P} 0 \quad (n \rightarrow \infty) \quad (6.24)$$

を証明する. すると (6.24) と逆マルチンゲールの収束定理の議論から

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

が成立することが知られている. \square

6.6 定理 6.20 の証明のための準備

6.6.1 対称化トリック

補題 6.22. X_1, X_2, \dots, X_n を独立な確率過程, 各過程 $X_i = \{X_{i,s}\}_{s \in \mathcal{T}}$ は中心化されている⁹ とする. すなわち, $E[X_{i,s}] = 0$ である. ただし,

⁹ $X_{j,s} = g(X_j) - P g(s = g, \mathcal{T} = \mathcal{G})$ と対応させる.

\mathcal{T} は添え字集合である. $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ は Rademacher 確率変数列で, X_1, X_2, \dots, X_n とは独立とする. このとき

$$\frac{1}{2} \mathbb{E} \left[\sup_{s \in \mathcal{T}} \left| \sum_{j=1}^n \epsilon_j X_{j,s} \right| \right] \leq \mathbb{E} \left[\sup_{s \in \mathcal{T}} \left| \sum_{j=1}^n X_{j,s} \right| \right] \leq 2 \mathbb{E} \left[\sup_{s \in \mathcal{T}} \left| \sum_{j=1}^n \epsilon_j X_{j,s} \right| \right] \quad (6.25)$$

と

$$\mathbb{E} \left[\sup_{s \in \mathcal{T}} \sum_{j=1}^n X_{j,s} \right] \leq 2 \mathbb{E} \left[\sup_{s \in \mathcal{T}} \sum_{j=1}^n \epsilon_j X_{j,s} \right] \quad (6.26)$$

が成立する.

Proof. ①(6.25) の 2 番目の不等号の証明: X'_j を X_j の独立複製とする. 各確率過程 X_j は中心化されていることに注意すると

$$\begin{aligned} \mathbb{E} \left[\sup_{s \in \mathcal{T}} \left| \sum_{j=1}^n X_{j,s} \right| \right] &= \mathbb{E} \left[\sup_{s \in \mathcal{T}} \left| \sum_{j=1}^n \{X_{j,s} - \mathbb{E}[X'_{j,s}]\} \right| \right] \\ &= \mathbb{E} \left[\sup_{s \in \mathcal{T}} \left| \mathbb{E} \left[\sum_{j=1}^n \{X_{j,s} - X'_{j,s}\} \mid X'_j \right] \right| \right] \\ &\leq \mathbb{E} \left[\sup_{s \in \mathcal{T}} \left| \sum_{j=1}^n \{X_{j,s} - X'_{j,s}\} \right| \right] \\ &\quad (\because \text{Jensen の不等式と towerng propaty}) \\ &= \mathbb{E} \left[\sup_{s \in \mathcal{T}} \left| \sum_{j=1}^n \epsilon_j \{X_{j,s} - X'_{j,s}\} \right| \right] \\ &\quad (\because X_{j,s} - X'_{j,s} \text{ の分布の対称性}) \\ &\leq 2 \mathbb{E} \left[\sup_{s \in \mathcal{T}} \left| \sum_{j=1}^n \epsilon_j X_{j,s} \right| \right] \end{aligned}$$

からわかる¹⁰.

¹⁰ 上の変形の 1 番目の不等号は関数 $x \mapsto |x|$ に対して, Jensen の不等式を適用すると

$$\begin{aligned} \mathbb{E} \left[\sup_{s \in \mathcal{T}} \left| \mathbb{E} \left[\sum_{j=1}^n \{X_{j,s} - X'_{j,s}\} \mid X'_j \right] \right| \right] &\leq \mathbb{E} \left[\sup_{s \in \mathcal{T}} \mathbb{E} \left[\left| \sum_{j=1}^n \{X_{j,s} - X'_{j,s}\} \right| \mid X'_j \right] \right] \\ &\leq \mathbb{E} \left[\mathbb{E} \left[\sup_{s \in \mathcal{T}} \left| \sum_{j=1}^n \{X_{j,s} - X'_{j,s}\} \right| \mid X'_j \right] \right] \\ &= \mathbb{E} \left[\sup_{s \in \mathcal{T}} \left| \sum_{j=1}^n \{X_{j,s} - X'_{j,s}\} \right| \right] \end{aligned}$$

に注意すればよい.

② (6.25) の 1 番目の不等号の証明: ① と同じように示せばよい.

③ (6.26) の証明: ① と同様. \square

補題 6.23. \mathcal{G} を関数族, \mathbf{P} を \mathbb{R} 上の確率測度, $X_1, X_2, \dots, X_n \stackrel{i.i.d.}{\sim} \mathbf{P}$ とする. さらに, 任意の $g \in \mathcal{G}$ とある $\delta > 0$ に対して

$$\Pr\left(\left|\int g d(\hat{\mathbf{P}}_n - \mathbf{P})\right| > \frac{\delta}{2}\right) \leq \frac{1}{2} \quad (6.18)$$

が成立する¹¹ とする. ただし, $\hat{\mathbf{P}}_n$ は X_1, X_2, \dots, X_n に基づく経験確率測度である. このとき

$$\Pr\left(\sup_{g \in \mathcal{G}} \left|\int g d(\hat{\mathbf{P}}_n - \mathbf{P})\right| > \delta\right) \leq 2 \Pr\left(\sup_{g \in \mathcal{G}} \left|\int g d(\hat{\mathbf{P}}_n - \hat{\mathbf{P}}'_n)\right| > \frac{\delta}{2}\right)$$

が成り立つ. ただし, $\hat{\mathbf{P}}'_n$ は, (X_1, X_2, \dots, X_n) の独立複製 $(X'_1, X'_2, \dots, X'_n)$ に基づく経験確率測度である.

Proof. $\mathbf{X} = (X_1, X_2, \dots, X_n)$ とおき, $g \in \mathcal{G}$ に対して, ランダムな部分集合 $A_g \subset \mathbb{R}^n$ を

$$A_g := \left\{ \mathbf{X}; \left|\int g d(\hat{\mathbf{P}}_n - \mathbf{P})\right| > \delta \right\}$$

で定める. さらに

$$A := \bigcup_{g \in \mathcal{G}} A_g$$

とする. 部分集合 A の定義から

$$\mathbf{X} \in A \Leftrightarrow \exists g_{\mathbf{X}} =: g^* \in \mathcal{G} \text{ s.t. } \mathbf{X} \in A_{g^*}$$

である. g^* は \mathbf{X} に依存するので, \mathcal{G} に値を取るランダム関数である. $\hat{\mathbf{P}}_n$

¹¹(6.18) から

$$\frac{1}{2} \leq 1 - \Pr\left(\left|\int g d(\hat{\mathbf{P}}_n - \mathbf{P})\right| > \frac{\delta}{2}\right) = \Pr\left(\left|\int g d(\hat{\mathbf{P}}_n - \mathbf{P})\right| \leq \frac{\delta}{2}\right)$$

となることに注意する.

と \hat{P}'_n の独立性から

$$\begin{aligned}
& \Pr\left(\mathbf{X} \in A_{g^*} \text{ かつ } \left| \int g^* d(\hat{P}'_n - P) \right| \leq \frac{\delta}{2}\right) \\
&= \mathbb{E}\left[\mathbb{1}\left\{\left\{\mathbf{X} \in A_{g^*}\right\} \cap \left\{\left| \int g^* d(\hat{P}'_n - P) \right| \leq \frac{\delta}{2}\right\}\right\}\right] \\
&= \mathbb{E}_{\mathbf{X}}\left[\mathbb{1}\left\{\left\{\mathbf{X} \in A_{g^*}\right\} \mathbb{E}\left[\mathbb{1}\left\{\left| \int g^* d(\hat{P}'_n - P) \right| \leq \frac{\delta}{2}\right\} \middle| \mathbf{X}\right]\right\}\right] \\
&= \mathbb{E}_{\mathbf{X}}\left[\mathbb{1}\left\{\left\{\mathbf{X} \in A_{g^*}\right\} \underbrace{\Pr\left(\left| \int g^* d(\hat{P}'_n - P) \right| \geq \frac{\delta}{2} \middle| \mathbf{X}\right)}_{\geq 1/2 \quad \because (6.18)}\right\}\right] \\
&\geq \frac{1}{2} \Pr(A_{g^*}) \\
&= \frac{1}{2} \Pr\left(\left| \int g^* d(\hat{P}_n - P) \right| > \delta\right) \tag{6.27}
\end{aligned}$$

となる. この不等式を用いると

$$\begin{aligned}
& \Pr\left(\sup_{g \in \mathcal{G}} \left| \int g d(\hat{P}_n - P) \right| > \delta\right) = \Pr\left(\mathbf{X} \in \bigcup_{g \in \mathcal{G}} A_g\right) \\
&\leq \Pr(\mathbf{X} \in A_{g^*}) \\
&= \Pr\left(\left| \int g^* d(\hat{P}_n - P) \right| > \delta\right) \\
&\leq 2 \Pr\left(\mathbf{X} \in A_{g^*} \text{ かつ } \left| \int g^* d(\hat{P}'_n - P) \right| \leq \frac{\delta}{2}\right) \quad (\because (6.27)) \\
&= 2 \Pr\left(\left\{\mathbf{X} \in A_{g^*}\right\} \cap \left\{\left| \int g^* d(\hat{P}'_n - P) \right| \leq \frac{\delta}{2}\right\}\right) \\
&= 2 \Pr\left(\left\{\left| \int g^* d(\hat{P}_n - P) \right| > \delta\right\} \cap \left\{\left| \int g^* d(\hat{P}'_n - P) \right| \leq \frac{\delta}{2}\right\}\right) \\
&\leq 2 \Pr\left(\left| \int g^* d(\hat{P}_n - \hat{P}'_n) \right| > \frac{\delta}{2}\right) \\
&\leq 2 \Pr\left(\sup_{g \in \mathcal{G}} \left| \int g d(\hat{P}_n - \hat{P}'_n) \right| > \frac{\delta}{2}\right)
\end{aligned}$$

がわかる. 最初の不等号は

$$\mathbf{X} \in \bigcup_{g \in \mathcal{G}} A_g \Rightarrow \mathbf{X} \in A_{g^*}$$

なので

$$\left\{\mathbf{X} \in \bigcup_{g \in \mathcal{G}} A_g\right\} \subset \{\mathbf{X} \in A_{g^*}\}$$

からわかる. また, 最後の不等号は

$$\left| \int g^* d(\widehat{\mathbb{P}}_n - \mathbb{P}) \right| > \delta \text{ かつ } \left| \int g^* d(\widehat{\mathbb{P}}'_n - \mathbb{P}) \right| \leq \frac{\delta}{2}$$

ならば

$$\left| \int g^* d(\widehat{\mathbb{P}}_n - \widehat{\mathbb{P}}'_n) \right| > \frac{\delta}{2}$$

であること¹² からわかる. □

系 6.24. \mathcal{G} を関数族, \mathbb{P} を \mathbb{R} 上の確率測度, $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathbb{P}$ とする. $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ を Rademacher 確率変数列で $X_j (j = 1, 2, \dots, n)$ は独立とする. このとき, ある $\delta > 0$ とすべての $g \in \mathcal{G}$ に対して

$$\Pr\left(\left|\int g d(\widehat{\mathbb{P}}_n - \mathbb{P})\right| > \frac{\delta}{2}\right) \leq \frac{1}{2} \quad (6.28)$$

とする. このとき

$$\Pr\left(\sup_{g \in \mathcal{G}} \left|\int g d(\widehat{\mathbb{P}}_n - \mathbb{P})\right| > \delta\right) \leq 4\Pr\left(\sup_{g \in \mathcal{G}} \left|\frac{1}{n} \sum_{j=1}^n \epsilon_j g(X_j)\right| > \frac{\delta}{4}\right)$$

が成り立つ.

¹²三角不等式と条件から

$$\begin{aligned} \delta &< \left|\int g^* d(\widehat{\mathbb{P}}_n - \mathbb{P})\right| \leq \left|\int g^* d(\widehat{\mathbb{P}}_n - \widehat{\mathbb{P}}'_n)\right| + \left|\int g^* d(\widehat{\mathbb{P}}'_n - \mathbb{P})\right| \\ &\leq \left|\int g^* d(\widehat{\mathbb{P}}_n - \widehat{\mathbb{P}}'_n)\right| + \frac{\delta}{2} \end{aligned}$$

から

$$\left|\int g^* d(\widehat{\mathbb{P}}_n - \widehat{\mathbb{P}}'_n)\right| \leq \frac{\delta}{2}$$

がわかる.

Proof. 補題 6.23 から

$$\begin{aligned}
& \Pr\left(\sup_{g \in \mathcal{G}} \left| \int g d(\widehat{\mathbb{P}}_n - \mathbb{P}) \right| > \delta\right) \\
& \leq 2\Pr\left(\sup_{g \in \mathcal{G}} \left| \int g d(\widehat{\mathbb{P}}_n - \widehat{\mathbb{P}}'_n) \right| > \frac{\delta}{2}\right) \\
& = 2\Pr\left(\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{j=1}^n \epsilon_j \{g(X_j) - g(X'_j)\} \right| > \frac{\delta}{2}\right) \\
& \leq 2\left\{ \Pr\left(\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{j=1}^n \epsilon_j g(X_j) \right| > \frac{\delta}{4}\right) + \Pr\left(\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{j=1}^n \epsilon_j g(X'_j) \right| > \frac{\delta}{4}\right) \right\} \\
& = 4\Pr\left(\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{j=1}^n \epsilon_j g(X_j) \right| > \frac{\delta}{4}\right)
\end{aligned}$$

からわかる。最後の不等式は以下のような議論からわかる。 U と V を実数値確率変数とする。このとき

$$|U| \leq \frac{\delta}{4} \text{ かつ } |V| \leq \frac{\delta}{4}$$

ならば

$$|U - V| \leq |U| + |V| \leq \frac{\delta}{2}$$

となる。したがって

$$|U| \leq \frac{\delta}{4} \text{ かつ } |V| \leq \frac{\delta}{4} \text{ ならば } |U - V| \leq \frac{\delta}{2}$$

となる。これの対偶をとれば

$$|U - V| > \frac{\delta}{2} \text{ ならば } |U| > \frac{\delta}{4} \text{ または } |V| > \frac{\delta}{4}.$$

なので

$$\Pr\left(|U - V| > \frac{\delta}{2}\right) \leq \Pr\left(|U| > \frac{\delta}{4}\right) + \Pr\left(|V| > \frac{\delta}{4}\right)$$

よりわかる。 □

補題 6.25. 関数族 \mathcal{G} は有限集合とし、 $\#\mathcal{G} = N > 1$ とする。ある定数 $K > 0$ が存在して

$$\max_{g \in \mathcal{G}} \|g\|_{\infty} \leq K \tag{6.29}$$

と仮定する. このとき, $t > 0$ に対して

$$\Pr\left(\max_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{j=1}^n g(X_j) \right| > K \sqrt{\frac{2(\log N + t)}{n}}\right) \leq e^{-t} \quad (6.20)$$

が成り立つ. さらに, $\log N + t \geq 1$ なる $t > 0$ に対して,

$$\Pr\left(\max_{g \in \mathcal{G}} \left| \int g d(\widehat{\mathbb{P}}_n - \mathbb{P}) \right| > \frac{K}{4} \sqrt{\frac{2(\log N + t)}{n}}\right) \leq 8e^{-t} \quad (6.21)$$

が成り立つ.

Proof. (6.20) の証明: $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ を X_1, X_2, \dots, X_n とは独立な Rademacher 列とする. $g \in \mathcal{G}$ と $j = 1, 2, \dots, n$ に対して

$$\begin{aligned} \mathbb{E}[\epsilon_j g(X_j)] &= \mathbb{E}[\mathbb{E}[\epsilon_j g(X_j) | X_j]] \\ &= \mathbb{E}\left[\Pr(\epsilon_j = 1) \times 1 \times g(X_j) + \Pr(\epsilon_j = -1) \times (-1) \times g(X_j)\right] \\ &= \mathbb{E}\left[\frac{1}{2}g(X_j) - \frac{1}{2}g(X_j)\right] \\ &= 0 \end{aligned}$$

となる. また, (6.29) と $\epsilon_j \in \{-1, 1\}$ ($j = 1, 2, \dots, N$) から

$$|\epsilon_j g(x)| \leq K \quad (x \in \mathbb{R})$$

である. Hoeffding の不等式 (命題 5.14) より, $s > 0$ に対して

$$\begin{aligned} &\Pr\left(\left|\frac{1}{n} \sum_{j=1}^n \epsilon_j g(X_j)\right| > s \mid X_1, X_2, \dots, X_n\right) \\ &= \Pr\left(\left|\frac{1}{n} \sum_{j=1}^n \epsilon_j g(X_j) - \mathbb{E}\left[\frac{1}{n} \sum_{j=1}^n \epsilon_j g(X_j)\right]\right| > s \mid X_1, X_2, \dots, X_n\right) \\ &\leq 2 \exp\left\{-\frac{2n^2 s^2}{4nK^2}\right\} = 2 \exp\left\{-\frac{ns^2}{2K^2}\right\} \end{aligned}$$

を得る. さらに, towering property から

$$\Pr\left(\left|\frac{1}{n} \sum_{j=1}^n \epsilon_j g(X_j)\right| > s\right) \leq 2 \exp\left\{-\frac{ns^2}{2K^2}\right\}$$

を得る. 次に, 上の不等式と union bound を用いると

$$\begin{aligned} \Pr\left(\max_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{j=1}^n \epsilon_j g(X_j) \right| > s\right) &= \Pr\left(\bigcup_{g \in \mathcal{G}} \left\{ \left| \frac{1}{n} \sum_{j=1}^n \epsilon_j g(X_j) \right| > s \right\}\right) \\ &\leq \sum_{g \in \mathcal{G}} \Pr\left(\left| \frac{1}{n} \sum_{j=1}^n \epsilon_j g(X_j) \right| > s\right) \\ &= 2N \exp\left\{-\frac{ns^2}{2K^2}\right\} \\ &= 2 \exp\left\{-\frac{ns^2}{2K^2} + \log N\right\} \end{aligned}$$

を得る. ここで

$$t = \frac{ns^2}{2K^2} - \log N \Leftrightarrow s = K \sqrt{\frac{2(t + \log N)}{n}}$$

となる.

$$\Pr\left(\left| \frac{1}{n} \sum_{j=1}^n \epsilon_j g(X_j) \right| > K \sqrt{\frac{2(t + \log N)}{n}}\right) \leq 2e^{-t} \quad (t > 0)$$

を得る. よって, (6.20) は証明できた.

(6.21) の証明: (6.20) と系 6.24 を用いて証明する. そのために条件 (6.28) を確認する. まず, (6.29) から

$$\text{Var}[g(X_1)] \leq \mathbb{E}\{[g(X_1)]^2\} \leq K^2$$

と評価できることに注意する. このとき, 各 $g \in \mathcal{G}$ と $K^2/(n\delta^2) \leq 1/2$ なる $\delta > 0$ に対して, Chebyshev の不等式 (系 1.33) を用いると

$$\begin{aligned} \Pr\left(\left| \int g d(\hat{\mathbb{P}}_n - \mathbb{P}) \right| > \delta\right) &= \Pr\left(\left| \frac{1}{n} \sum_{j=1}^n \left\{ g(X_j) - \mathbb{E}[g(X_j)] \right\} \right| > \delta\right) \\ &\leq \frac{\text{Var}[g(X_1)]}{n\delta^2} \leq \frac{K^2}{n\delta^2} \leq \frac{1}{2} \end{aligned}$$

となる. よって, (6.28) が成立するので, $\log N + t \geq 1$ のとき, 確率の対称化定理 (系 6.24) より

$$\begin{aligned} \Pr\left(\max_{g \in \mathcal{G}} \left| \int g d(\hat{\mathbb{P}}_n - \mathbb{P}) \right| > K \sqrt{\frac{2(t + \log N)}{n}}\right) \\ \leq 4 \Pr\left(\max_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{j=1}^n \epsilon_j g(X_j) \right| > \frac{K}{4} \sqrt{\frac{2(t + \log N)}{n}}\right) \leq 8e^{-t} \end{aligned}$$

を得る. □

6.6.2 Dudley のエントロピー積分

補題 6.26. X_1, X_2, \dots, X_N は劣 Gauss 確率変数列 $\text{sugG}(\nu)$ とする. すなわち, $\forall \lambda > 0$ に対して, $\mathbb{E}[e^{\lambda X_j}] \leq e^{\lambda^2 \nu / 2}$ ($j = 1, 2, \dots, N$) が成立する. このとき

$$\mathbb{E} \left[\max_{j=1,2,\dots,N} X_j \right] \leq \sqrt{2\nu \log N}$$

が成立する.

Proof. Jensen の不等式 (定理 1.36) を用いると

$$\begin{aligned} \exp \left(\lambda \mathbb{E} \left[\max_{j=1,2,\dots,N} X_j \right] \right) &\leq \mathbb{E} \left[\exp \left(\lambda \max_{j=1,2,\dots,N} X_j \right) \right] \\ &= \mathbb{E} \left[\max_{j=1,2,\dots,N} \exp(\lambda X_j) \right] \\ &\leq \sum_{j=1}^N \underbrace{\mathbb{E}[\exp(\lambda X_j)]}_{\leq e^{\lambda^2 \nu / 2}} \\ &\leq N \exp \left(\frac{\lambda^2 \nu}{2} \right) \end{aligned}$$

を得る. 両辺の対数を取ると

$$\mathbb{E} \left[\max_{j=1,2,\dots,N} X_j \right] \leq \frac{\log N}{\lambda} + \frac{\lambda \nu}{2}$$

を得る. この式は任意の $\lambda > 0$ に対して成立するので, $\lambda = \sqrt{2 \log(N)/n}$ を代入すると

$$\mathbb{E} \left[\max_{j=1,2,\dots,N} X_j \right] \leq \sqrt{2\nu \log(N)}$$

がわかる. □

定理 6.27. (\mathbb{T}, d) を距離空間とし, $\{X_t\}_{t \in \mathbb{T}}$ を \mathbb{T} を添え字集合とする確率過程で, 任意の $t_1, t_2 \in \mathbb{T}$ と $\lambda > 0$ に対して

$$\log \left(\mathbb{E} \left[\exp \{ \lambda (X_{t_1} - X_{t_2}) \} \right] \right) \leq \frac{\lambda^2 d^2(t_1, t_2)}{2} \quad (6.30)$$

をみたすとする. このとき, すべての $t_0 \in \mathbb{T}$ に対して

$$\mathbb{E} \left[\sup_{t \in \mathbb{T}} |X_t - X_{t_0}| \right] \leq 12 \int_0^{\delta/2} \sqrt{H(\epsilon, \mathbb{T}, d)} \, d\epsilon$$

が成立する. ただし, $\delta = \sup_{t \in \mathbb{T}} d(t, t_0)$ である. とくに, $D = \sup_{t_1, t_2 \in \mathbb{T}} d(t_1, t_2) < \infty$ のとき

$$\mathbb{E} \left[\sup_{t_1, t_2 \in \mathbb{T}} |X_{t_1} - X_{t_2}| \right] \leq 24 \int_0^{\delta/2} \sqrt{H(\epsilon, \mathbb{T}, d)} d\epsilon$$

が成立する.

Proof. 任意の $\epsilon > 0$ に対して, $H(\epsilon, \mathbb{T}, d) < \infty$ と仮定する. そうでない場合には, 不等式は自明となる.

① \mathbb{T} は高々可算集合の場合: 任意の $j \in \mathbb{N}$ に対して, $\delta_j := \delta 2^{-j}$ と記す. すると, 任意の $j \in \mathbb{N}$ に対して, $N_j := N(\delta_j, \mathbb{T}, d)$ は有限となる. このことより, 部分集合 $\mathbb{T}_j := \{t_1, t_2, \dots, t_{N_j}\} \subset \mathbb{T}$ が存在して

$$\bigcup_{k=1}^{N_j} B_d(x_k, \delta_j), \quad B_d(t_k, \delta_j) := \{t \in \mathbb{T}; d(t, t_k) < \delta_j\}$$

とできる. 各 $j \in \mathbb{N}$ に対して, \mathbb{T} から \mathbb{T}_j への写像 Π_j を $t \in \mathbb{T}$ に対して, ある $t_j (\exists j \in \{1, 2, \dots, N_j\})$ で $d(t, t_j) < \delta_j$ なるものに対応させる写像とする. x に対して, $\#\{j \in \{1, 2, \dots, N_j\}; d(t_j, t) < \delta_j\} \geq 2$ の場合には, そのうちのどれかを対応させればよい. ここで, $\mathbb{T}_0 = \{t_0\}$ とし, $\Pi_0(t) = t_0$ と定義する.

Step 1: つぎに

$$X_t = X_{t_0} + \sum_{j=1}^{\infty} \{X_{\Pi_{j+1}(t)} - X_{\Pi_j(t)}\}$$

と書く. \mathbb{T} は有限集合なので, ある大きな $j_0 \in \mathbb{N}$ が存在して, $X_{\Pi_{j_0}(t)} = X_t$ となっている. すなわち, 上の表現の有限和である.

Step 2: つぎに

$$\mathbb{E} \left[\sup_{t \in \mathbb{T}} |X_t - X_{t_0}| \right] \leq \sum_{j=0}^{\infty} \mathbb{E} \left[\sup_{t \in \mathbb{T}} |X_{\Pi_{j+1}(t)} - X_{\Pi_j(t)}| \right] \quad (6.31)$$

である. また

$$\begin{aligned} \#\{(\Pi_j(t), \Pi_{j+1}(t)); t \in \mathbb{T}\} &\leq N(\delta_j, \mathbb{T}, d) \times N(\delta_{j+1}, \mathbb{T}, d) \\ &\leq \{N(\delta_{j+1}, \mathbb{T}, d)\}^2 \\ &= \exp\left(\log(\{N(\delta_{j+1}, \mathbb{T}, d)\}^2)\right) \\ &= \exp\left(2H(\delta_{j+1}, \mathbb{T}, d)\right) \end{aligned} \quad (6.32)$$

となることがわかる. 三角不等式から

$$d(\Pi_j(t), \Pi_{j+1}(t)) \leq d(\Pi_j(t), t) + d(t, \Pi_{j+1}(t)) \leq \delta_j + \delta_{j+1} \leq 3\delta_{j+1} \quad (6.33)$$

となる. (6.30) と (6.33) から

$$\begin{aligned} \log\left(\mathbb{E}\left[\exp(\lambda(X_{\Pi_{j+1}(t)} - X_{\Pi_j(t)}))\right]\right) &\leq \frac{\lambda^2 d^2(\Pi_{j+1}(t), \Pi_j(t))}{2} \\ &\leq \frac{9\lambda^2 \delta_{j+1}^2}{2} \end{aligned}$$

であるので

$$X_{\Pi_{j+1}(t)} - X_{\Pi_j(t)} \sim \text{subG}(9\delta_{j+1}^2)$$

となることがわかる. よって, 補題 6.26 と (6.32) から

$$\begin{aligned} \mathbb{E}\left[\sup_{t \in \mathbb{T}} |X_{\Pi_{j+1}(t)} - X_{\Pi_j(t)}|\right] &\leq \sqrt{2 \times 9\delta_{j+1}^2 \log(\#\{(\Pi_j(t), \Pi_{j+1}(t)); t \in \mathbb{T}\})} \\ &\leq \sqrt{18\delta_{j+1}^2 \times 2H(\delta_{j+1}, \mathbb{T}, d)} \\ &= 6\delta_{j+1} \sqrt{H(\delta_{j+1}, \mathbb{T}, d)} \end{aligned} \quad (6.34)$$

を得る. (6.31) の右辺に (6.34) を代入すると

$$\begin{aligned} \mathbb{E}\left[\sup_{t \in \mathbb{T}} |X_t - X_{t_0}|\right] &\leq \sum_{j=0}^{\infty} 6\delta_{j+1} \sqrt{H(\delta_{j+1}, \mathbb{T}, d)} \\ &= \sum_{j=1}^{\infty} 6\delta_j \sqrt{H(\delta_j, \mathbb{T}, d)} \\ &= 12 \sum_{j=1}^{\infty} \underbrace{(\delta_j - \delta_{j+1})}_{\delta_j - \delta_{j+1} = \delta_j/2} \sqrt{H(\delta_j, \mathbb{T}, d)} \\ &= 12 \int_0^{\delta/2} \sqrt{H(\epsilon, \mathbb{T}, d)} d\epsilon \end{aligned}$$

を得る.

② \mathbb{T} は非可算集合の場合: $H(\epsilon, \mathbb{T}, d) < \infty$ なので, \mathbb{T} は全有界である.

□

補題 6.28. X_1, X_2, \dots, X_n を i.i.d. 確率変数列とし, $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ を Rademacher 確率変数列とし, $\{X_j\}_{j=1}^n$ と $\{\epsilon_j\}_{j=1}^n$ は独立とする. この

とき

$$\begin{aligned} & \mathbb{E} \left[\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{j=1}^n \epsilon_j g(X_j) \right| \middle| X_1, X_2, \dots, X_n \right] \\ & \leq \sqrt{2\pi} \frac{\sup_{g \in \mathcal{G}} \|g\|_{L_2(\hat{\mathbb{P}}_n)}}{\sqrt{n}} + 12 \int_0^{D_n} \frac{H(\epsilon, \mathcal{G}, \|\cdot\|_{L_2(\hat{\mathbb{P}}_n)})}{n} d\epsilon \end{aligned}$$

が成立する. ただし

$$\begin{aligned} D_n & := \sup_{g \in \mathcal{G}} \|g\|_{L_\infty(\hat{\mathbb{P}}_n)}; & \|g\|_{L_\infty(\hat{\mathbb{P}}_n)} & := \max_{j=1,2,\dots,n} |g(X_j)| \\ \|g\|_{L_2(\hat{\mathbb{P}}_n)} & := \frac{1}{n} \sum_{j=1}^n g^2(X_j) \end{aligned}$$

である.

Proof. 定理 6.27 を用いるために, 条件 (6.30) を確認する. すなわち, 過程 $\{\sum_{j=1}^n \epsilon_j g(X_j)\}_{g \in \mathcal{G}}$ の増分が劣 Gauss であることを示せばよい. そのために, 任意の $g_1, g_2 \in \mathcal{G}$ を取る. $j = 1, 2, \dots, n$ に対して

$$Y_j := \frac{\epsilon_j}{n} \{g_1(X_j) - g_2(X_j)\}$$

とおく. このとき, $\mathbf{X} := X_1, X_2, \dots, X_n$ を与えたとき

$$\mathbb{E}[Y_j | \mathbf{X}] = 0, \quad Y_j \in \left[\underbrace{-\frac{|g_1(X_j) - g_2(X_j)|}{n}}_a, \underbrace{\frac{|g_1(X_j) - g_2(X_j)|}{n}}_{=b} \right]$$

となるので, 補題 5.13 から, 任意の $\lambda > 0$ に対して

$$\mathbb{E} \left[\exp(\lambda Y_j) \middle| \mathbf{X} \right] \leq \exp \left(\frac{\lambda^2 \{g_1(X_j) - g_2(X_j)\}^2}{2n^2} \right)$$

となる. Y_1, Y_2, \dots, Y_n は独立な確率変数列なので

$$\begin{aligned} \mathbb{E} \left[\exp \left(\sum_{j=1}^n \frac{\lambda Y_j}{n} \right) \middle| \mathbf{X} \right] & = \mathbb{E} \left[\prod_{j=1}^n \exp \left(\frac{\lambda Y_j}{n} \right) \middle| \mathbf{X} \right] \\ & = \prod_{j=1}^n \mathbb{E} \left[\exp \left(\frac{\lambda Y_j}{n} \right) \middle| \mathbf{X} \right] \\ & \leq \prod_{j=1}^n \exp \left(\frac{\lambda^2 \{g_1(X_j) - g_2(X_j)\}^2}{2n^2} \right) \\ & = \exp \left(\sum_{j=1}^n \frac{\lambda^2 \{g_1(X_j) - g_2(X_j)\}^2}{2n^2} \right) \\ & = \exp \left(\frac{\lambda^2 \{ \|g_1 - g_2\|_{L_2(\hat{\mathbb{P}}_n)} / \sqrt{n} \}^2}{2} \right) \end{aligned}$$

となる. よって, $d(g_1, g_2) = \|g_1 - g_2\|_{L_2(\hat{\mathbb{P}}_n)}/\sqrt{n}$ として, (6.30) が成立することが示せた. Dudley のエントロピー上限 (補題 6.28) を適用するために, $g_0 \in \mathcal{G}$ を取る. すると

$$\begin{aligned}
& \mathbb{E} \left[\sup_{g \in \mathcal{G}} \left| n^{-1} \sum_{j=1}^n \epsilon_j \{g(X_j) - g_0(X_j)\} \right| \middle| \mathbf{X} \right] \\
& \leq 12 \int_0^{\delta_n/2} \sqrt{H(\epsilon, \mathcal{G}, n^{-1/2} \|\cdot\|_{L_2(\hat{\mathbb{P}}_n)})} d\epsilon \\
& = 12 \int_0^{\delta_n/2} \sqrt{H(\sqrt{n}\epsilon, \mathcal{G}, \|\cdot\|_{L_2(\hat{\mathbb{P}}_n)})} d\epsilon \\
& = 12 \int_0^{\sqrt{n}\delta_n/2} \sqrt{\frac{H(\epsilon', \mathcal{G}, \|\cdot\|_{L_2(\hat{\mathbb{P}}_n)})}{n}} d\epsilon' \\
& \quad (\epsilon' = \sqrt{n}\epsilon \text{ と変換}) \\
& = 12 \int_0^{D_n} \sqrt{\frac{H(\epsilon, \mathcal{G}, \|\cdot\|_{L_2(\hat{\mathbb{P}}_n)})}{n}} d\epsilon \tag{6.35}
\end{aligned}$$

を得る. ただし, $\delta_n = n^{-1/2} \sup_{g \in \mathcal{G}} \|g - g_0\|_{L_2(\hat{\mathbb{P}})}$ と $D_n := \sup_{g \in \mathcal{G}} \|g - g_0\|_{L_2(\hat{\mathbb{P}})}$ である. しかし, $\sum_{j=1}^n \epsilon_j g_0(X_j)$ に対して, Hoeffding の不等式を用いると

$$\begin{aligned}
& \mathbb{E} \left[\left| n^{-1} \sum_{j=1}^n \epsilon_j g_0(X_j) \right| \middle| \mathbf{X} \right] \\
& = \int_0^\infty \Pr \left(\left| n^{-1} \sum_{j=1}^n \epsilon_j g_0(X_j) \right| > t \middle| \mathbf{X} \right) dt \\
& \quad (\because \text{命題 1.31}) \\
& \leq \int_0^\infty 2 \exp \left(-\frac{2n^2 t^2}{4 \sum_{j=1}^n \{g_0(X_j)\}^2} \right) dt \\
& \quad (\because \text{命題 5.14}) \\
& = \int_{-\infty}^\infty \exp \left(-\frac{n^2 t^2}{2 \sum_{j=1}^n \{g_0(X_j)\}^2} \right) dt \\
& = \int_{-\infty}^\infty \exp \left(-\frac{t^2}{2 \{ \|g_0\|_{L_2(\hat{\mathbb{P}})} / n \}^2} \right) dt \\
& = \sqrt{2\pi} \frac{\|g_0\|_{L_2(\hat{\mathbb{P}}_n)}}{4\sqrt{n}} \\
& \leq \sqrt{2\pi} \frac{\sup_{g \in \mathcal{G}} \|g\|_{L_2(\hat{\mathbb{P}})}}{\sqrt{n}} \tag{6.36}
\end{aligned}$$

を得る. (6.35) と (6.36) から

$$\begin{aligned}
& \mathbb{E} \left[\sup_{g \in \mathcal{G}} \left| n^{-1} \sum_{j=1}^n \epsilon_j g(X_j) \right| \middle| \mathbf{X} \right] \\
& \leq \mathbb{E} \left[\sup_{g \in \mathcal{G}} \left\{ \left| \frac{1}{n} \sum_{j=1}^n \epsilon_j g_0(X_j) \right| + \left| \frac{1}{n} \sum_{j=1}^n \epsilon_j \{g(X_j) - g_0(X_j)\} \right| \right\} \middle| \mathbf{X} \right] \\
& \leq \mathbb{E} \left[\left| n^{-1} \sum_{j=1}^n \epsilon_j g_0(X_j) \right| \middle| \mathbf{X} \right] + \mathbb{E} \left[\sup_{g \in \mathcal{G}} \left| n^{-1} \sum_{j=1}^n \epsilon_j \{g(X_j) - g_0(X_j)\} \right| \middle| \mathbf{X} \right] \\
& \leq \sqrt{2\pi} \frac{\sup_{g \in \mathcal{G}} \|g\|_{L_2(\hat{\mathbb{P}})}}{\sqrt{n}} + 12 \int_0^{D_n} \sqrt{\frac{H(\epsilon, \mathcal{G}, \|\cdot\|_{L_2(\hat{\mathbb{P}}_n)})}{n}} d\epsilon
\end{aligned}$$

を得る. よって, 補題は示された. \square

補題 6.29. ある定数 $K > 0$ に対して, $\sup_{g \in \mathcal{G}} \|g\|_\infty < K$ と仮定する. 任意の $\epsilon > 0$ に対して

$$\frac{1}{n} H(\epsilon, \mathcal{G}, \|\cdot\|_{L_2(\hat{\mathbb{P}}_n)}) \xrightarrow{\mathbb{P}} 0 \quad (n \rightarrow \infty) \quad (6.23)$$

ならば, \mathcal{G} は P-GC 族である.

Proof.

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

を示すために

$$\mathbb{E} \left[\sup_{g \in \mathcal{G}} |\hat{\mathbb{P}}_n g - \mathbb{P}g| \right] \xrightarrow{n \rightarrow \infty} 0$$

を示せばよい. 補題 6.22 の対称化トリックと補題 6.28 から

$$\begin{aligned}
\mathbb{E} \left[\sup_{g \in \mathcal{G}} |\hat{\mathbb{P}}_n g - \mathbb{P}g| \right] & \leq \mathbb{E} \left[\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{j=1}^n \epsilon_j g(X_j) \right| \right] \\
& = \mathbb{E} \left[\mathbb{E} \left[\sup_{g \in \mathcal{G}} \left| n^{-1} \sum_{j=1}^n \epsilon_j g(X_j) \right| \middle| \mathbf{X} \right] \right] \\
& \leq 2\sqrt{2\pi} \frac{K}{\sqrt{n}} + 24 \int_0^K \sqrt{\frac{H(\epsilon, \mathcal{G}, \|\cdot\|_{L_2(\hat{\mathbb{P}}_n)})}{n}} d\epsilon
\end{aligned}$$

となる. 命題 6.7(1)(2) から

$$H(\epsilon, \mathcal{G}, \|\cdot\|_{L_2(\hat{\mathbb{P}}_n)}) \leq H_B(\epsilon, \mathcal{G}, \|\cdot\|_{L_2(\hat{\mathbb{P}}_n)}) \leq H\left(\frac{\epsilon}{2}, \mathcal{G}\right) \leq \left(\frac{2K}{\epsilon}\right)^n$$

となる. すると

$$\begin{aligned} \left| \int_0^K \left\{ \sqrt{\frac{H(\epsilon, \mathcal{G}, \|\cdot\|_{L_2(\hat{P}_n)})}{n}} \right\}^2 d\epsilon \right| &\leq \left| \int_0^K \frac{\log\left(\frac{2K}{\epsilon}\right)^n}{n} d\epsilon \right| \\ &= \left| \int_0^K \log(2K/\epsilon) d\epsilon \right| \\ &< \infty \end{aligned}$$

なので, 命題 2.6(1) から $\left\{ \sqrt{\frac{H(\epsilon, \mathcal{G}, \|\cdot\|_{L_2(\hat{P}_n)})}{n}} \right\}_{n=1}^{\infty}$ は一様可積分であることがわかる. したがって, (6.23) と定理 2.8(4) から

$$\int_0^K \sqrt{\frac{H(\epsilon, \mathcal{G}, \|\cdot\|_{L_2(\hat{P}_n)})}{n}} d\epsilon \xrightarrow{n \rightarrow \infty} 0$$

がわかる. 以上から

$$\mathbb{E} \left[\sup_{g \in \mathcal{G}} |\hat{P}_n g - P g| \right] \leq 2\sqrt{2\pi} \frac{K}{\sqrt{n}} + 24 \int_0^K \sqrt{\frac{H(\epsilon, \mathcal{G}, \|\cdot\|_{L_2(\hat{P}_n)})}{n}} d\epsilon \xrightarrow{n \rightarrow \infty} 0$$

が証明できた. よって, \mathcal{G} は P-GC 族であることがわかった. \square

6.7 定理 6.20 の証明

6.7.1 方針 ① の証明

証明は 2 津の段階 I と II で行う.

I. $\sup_{g \in \mathcal{G}} |g(x)| \leq K$ を仮定して, 定理 6.20 を証明: $\delta > 0, N = N(\delta, \mathcal{G}, \|\cdot\|_{L_1(\hat{P}_n)})$ とし, g_1, g_2, \dots, g_N を関数族 \mathcal{G} の δ 被覆数とする.
 $g \in \mathcal{G}$ を任意に取る. すると, ある $j \in \{1, 2, \dots, N\}$ に対して

$$P|g_j - g| := \frac{1}{n} \sum_{k=1}^n |g_j(X_k) - g(X_k)| < \delta$$

とできる. このことから

$$\begin{aligned} \left| \frac{1}{n} \sum_{k=1}^n \epsilon_k g(X_k) \right| &\leq \left| \frac{1}{n} \sum_{k=1}^n \epsilon_k g_j(X_k) \right| + \left| \frac{1}{n} \sum_{k=1}^n \epsilon_k \{g(X_k) - g_j(X_k)\} \right| \\ &\leq \left| \frac{1}{n} \sum_{k=1}^n \epsilon_k g_j(X_k) \right| + \frac{1}{n} \sum_{k=1}^n |g(X_k) - g_j(X_k)| \\ &\leq \left| \frac{1}{n} \sum_{k=1}^n \epsilon_k g_j(X_k) \right| + \delta \end{aligned}$$

である. したがって

$$\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{k=1}^n \epsilon_k g(X_k) \right| \leq \max_{j=1,2,\dots,N} \left| \frac{1}{n} \sum_{k=1}^n \epsilon_k g_j(X_k) \right| + \delta \quad (6.37)$$

を得る. $\mathbf{X} = (X_1, X_2, \dots, X_n)$ とする. Hoeffding の不等式 (補題 5.13) より, $t > 0$ に対して,

$$\Pr \left(\max_{j=1,2,\dots,N} \left| \frac{1}{n} \sum_{k=1}^n \epsilon_k g_j(X_k) \right| > K \sqrt{\frac{2(t + \log N)}{n}} \mid \mathbf{X} \right) \leq 2e^{-t} \quad (6.38)$$

となる. (6.37) と (6.38) を合わせると

$$\begin{aligned} \Pr \left(\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{k=1}^n \epsilon_k g(X_k) \right| > \delta + K \sqrt{\frac{2(t + \log N)}{n}} \mid \mathbf{X} \right) \\ \leq \Pr \left(\max_{j=1,2,\dots,N} \left| \frac{1}{n} \sum_{k=1}^n \epsilon_k g_j(X_k) \right| > K \sqrt{\frac{2(t + \log N)}{n}} \mid \mathbf{X} \right) \leq 2e^{-t} \end{aligned} \quad (6.39)$$

を得る. 条件付き期待値の性質 (towering property) を (6.39) に適用すると

$$\begin{aligned} \Pr \left(\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{k=1}^n \epsilon_k g(X_k) \right| > 2\delta + K \sqrt{\frac{2t}{n}} \right) \\ \leq 2e^{-t} + \Pr \left(K \sqrt{\frac{2H_1(\delta, \mathcal{G}, \|\cdot\|_{L_1(\hat{\mathbb{P}}_n)})}{n}} > \delta \right) \end{aligned} \quad (6.40)$$

となる. これは以下の議論からわかる.

$$A := \left\{ K \sqrt{\frac{2H(\delta, \mathcal{G}, \|\cdot\|_{L_1(\hat{\mathbb{P}}_n)})}{n}} > \delta \right\}$$

とおくと

$$A^c \text{ が起きると } \Rightarrow K\sqrt{\frac{2\log N}{n}} \leq \delta$$

となる.

$$\begin{aligned}
& \Pr \left(\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{k=1}^n \epsilon_k g(X_k) \right| > 2\delta + K\sqrt{\frac{2t}{n}} \right) \\
& \leq \Pr \left(\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{k=1}^n \epsilon_k g(X_k) \right| > 2\delta + K\sqrt{\frac{2t}{n}} \text{ かつ } A^c \right) \\
& \quad + \Pr \left(\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{k=1}^n \epsilon_k g(X_k) \right| > 2\delta + K\sqrt{\frac{2t}{n}} \text{ かつ } A \right) \\
& \leq \Pr \left(\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{k=1}^n \epsilon_k g(X_k) \right| > \delta + K\sqrt{\frac{2\log N}{n}} + K\sqrt{\frac{2t}{n}} \right) \\
& \quad + \Pr \left(K\sqrt{\frac{2H_1(\delta, \mathcal{G}, \|\cdot\|_{L_1(\hat{\mathbb{P}}_n)})}{n}} > \delta \right) \\
& \quad \left(\because A^c \Rightarrow K\sqrt{\frac{2\log N}{n}} \leq \delta \right. \\
& \quad \quad \left. \Rightarrow 2\delta + K\sqrt{\frac{2t}{n}} \geq \delta + K\sqrt{\frac{2\log N}{n}} + K\sqrt{\frac{2t}{n}} \right) \\
& \leq \Pr \left(\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{k=1}^n \epsilon_k g(X_k) \right| > \delta + K\sqrt{\frac{2(t + \log N)}{n}} \right) \\
& \quad + \Pr \left(K\sqrt{\frac{2H_1(\delta, \mathcal{G}, \|\cdot\|_{L_1(\hat{\mathbb{P}}_n)})}{n}} > \delta \right) \\
& \quad \left(\because \sqrt{a} + \sqrt{b} \geq \sqrt{a+b} (a, b \geq 0) \right) \\
& = \mathbb{E} \left[\Pr \left(\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{k=1}^n \epsilon_k g(X_k) \right| > \delta + K\sqrt{\frac{2(t + \log N)}{n}} \right) \middle| \mathbf{X} \right] \\
& \quad + \Pr \left(K\sqrt{\frac{2H_1(\delta, \mathcal{G}, \|\cdot\|_{L_1(\hat{\mathbb{P}}_n)})}{n}} > \delta \right) \\
& \leq 2e^{-t} + \Pr \left(K\sqrt{\frac{2H_1(\delta, \mathcal{G}, \|\cdot\|_{L_1(\hat{\mathbb{P}}_n)})}{n}} > \delta \right) \quad (\because (6.40))
\end{aligned}$$

からわかる. よって, (6.40) が確認できた.

次に, 系 6.24 を (6.40) に適用するために, 以下の (6.28) が成立するた

めの条件を求める.

$$\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{k=1}^n g(X_k) - \int g \, d\mathbf{P} \right| \leq 2K$$

に注意して, Chebyshev の不等式 (系 1.33) を用いると

$$\begin{aligned} & \Pr \left(\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{k=1}^n g(X_k) - \int g \, d\mathbf{P} \right| > \frac{\delta}{2} \right) \\ & \leq \frac{4}{\delta^2} \mathbb{E} \left[\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{k=1}^n g(X_k) - \int g \, d\mathbf{P} \right|^2 \right] \leq \frac{8K^2}{n\delta^2} \leq \frac{1}{2} \end{aligned}$$

より

$$n \geq \frac{16K^2}{\delta^2}$$

が (6.28) が成立するための十分条件となることがわかる. よって, 系 6.24 と (6.40) を用いれば

$$\begin{aligned} & \Pr \left(\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{k=1}^n g(X_k) - \int g \, d\mathbf{P} \right| > 8\delta + 4K \sqrt{\frac{2t}{n}} \right) \\ & \leq 4\Pr \left(\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{k=1}^n \epsilon_k g(X_k) \right| > 2\delta + K \sqrt{\frac{2t}{n}} \right) \\ & \leq 8e^{-t} + 4\Pr \left(K \sqrt{\frac{2H_1(\delta, \mathcal{G}, \|\cdot\|_{L_1(\hat{\mathbf{P}}_n)})}{n}} > \delta \right) \end{aligned}$$

を得る. $\epsilon > 0$ に対して

$$8e^{-t} \leq \frac{\epsilon}{2} \text{ かつ } 4\Pr \left(K \sqrt{\frac{2H_1(\delta, \mathcal{G}, \|\cdot\|_{L_1(\hat{\mathbf{P}}_n)})}{n}} > \delta \right) \leq \frac{\epsilon}{2}$$

になるように t と n を取り, さらに

$$4K \sqrt{\frac{2t}{n}} \leq 2\delta$$

になるように n を大きく取り直せば, (6.28) が成立する. 以上のことから

$$\Pr \left(\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{k=1}^n g(X_j) - \int g \, d\mathbf{P} \right| > 10\delta \right) \leq \epsilon$$

となる. したがって

$$\sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{k=1}^n g(X_j) - \int g \, d\mathbf{P} \right| \xrightarrow{\mathbf{P}} 0 \quad (n \rightarrow \infty)$$

がわかる.

II. $\sup_{g \in \mathcal{G}} |g(x)| \leq K$ なしで定理 6.20 を証明: \mathcal{G} と \mathcal{G}_K の違いを評価するために, すべての $g \in \mathcal{G}$ と任意の $K > 0$ に対して,

$$\begin{aligned} & \left| \frac{1}{n} \sum_{j=1}^n g(X_j) - \int_{\mathbb{R}} g \, d\mathbf{P} \right| \\ & \leq \left| \frac{1}{n} \sum_{j=1}^n g(X_j) \mathbb{1}\{G(X_j) \leq K\} - \int_{\mathbb{R}} g \mathbb{1}\{G \leq K\} \, d\mathbf{P} \right| \\ & \quad + \left| \frac{1}{n} \sum_{j=1}^n g(X_j) \mathbb{1}\{G(X_j) > K\} - \int_{\mathbb{R}} g \mathbb{1}\{G > K\} \, d\mathbf{P} \right| \\ & =: E_1 + E_2 \end{aligned}$$

となることに注意する. 上の不等式の最右辺の E_1 は, \mathcal{G}_K は P-GC 族なので

$$E_1 \leq \sup_{g \in \mathcal{G}_K} \left| \frac{1}{n} \sum_{j=1}^n g(X_j) - \int_{\mathbb{R}} g \, d\mathbf{P} \right| \xrightarrow{\mathbf{P}} 0 \quad (n \rightarrow \infty)$$

となる. 評価できる. 一方,

$$\begin{aligned} E_2 & \leq \sup_{g \in \mathcal{G}} |(\widehat{\mathbf{P}}_n - \mathbf{P})g \mathbb{1}\{G > K\}| \\ & = \sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{j=1}^n g(X_j) \mathbb{1}\{G(X_j) > K\} - \int_{\mathbb{R}} g \mathbb{1}\{G > K\} \, d\mathbf{P} \right| \\ & = \frac{1}{n} \sum_{j=1}^n G(X_j) \mathbb{1}\{G(X_j) > K\} - \int_{\mathbb{R}} G \mathbb{1}\{G > K\} \, d\mathbf{P} \\ & \quad + 2 \int_{\mathbb{R}} G \mathbb{1}\{G > K\} \, d\mathbf{P} \end{aligned}$$

である. 任意の K に対して, 大数の法則より

$$\frac{1}{n} \sum_{j=1}^n G(X_j) \mathbb{1}\{G(X_j) > K\} - \int_{\mathbb{R}} G \mathbb{1}\{G > K\} \, d\mathbf{P} \xrightarrow{\mathbf{P}} 0 \quad (n \rightarrow \infty)$$

となる. また, $G \in L_1(\mathbf{P})$ なので

$$\lim_{K \rightarrow \infty} \int_{\mathbb{R}} G \mathbb{1}\{G > K\} \, d\mathbf{P} = 0$$

である. 以上から

$$E_2 \xrightarrow{\mathbf{P}} 0 \quad (n \rightarrow \infty)$$

がわかる. よって

$$\|\widehat{P}_n - P\|_{\mathcal{G}} \leq \|\widehat{P}_n - P\|_{\mathcal{G}_K} + \sup_{g \in \mathcal{G}} |(\widehat{P}_n - P)g \mathbb{1}\{G > K\}| \xrightarrow{P} 0 \quad (n \rightarrow \infty)$$

がわかる.

6.7.2 方針 ② の証明

2 つの段階 I と II を踏んで, 定理は証明される.

I. 定理 6.20 の仮定 (6.17) のもとでも \mathcal{G}_K は P-GC 族であることを示そう. そのために, $g_1, g_2 \in \mathcal{G}$ に対して, $h_1 := g_1 \mathbb{1}\{G \leq K\}$ と $h_2 := g_2 \mathbb{1}\{G \leq K\}$ と記す. すると $h_1, h_2 \in \mathcal{G}_K$ であることに注意する. しかし

$$\int (h_1 - h_2)^2 d\widehat{P}_n = \int_{G \leq K} (g_1 - g_2)^2 d\widehat{P}_n \leq 2K \int |g_1 - g_2| d\widehat{P}_n$$

となる. 命題 6.7(3) に注意すると

$$H(2K\epsilon, \mathcal{G}_K, \|\cdot\|_{L_2(\widehat{P}_n)}) \leq H(\epsilon, \mathcal{G}, \|\cdot\|_{L_1(\widehat{P}_n)})$$

がわかる. よって

$$\frac{1}{n} H(\epsilon, \mathcal{G}, \|\cdot\|_{L_1(\widehat{P}_n)}) \xrightarrow{P} 0 \quad (n \rightarrow \infty)$$

ならば

$$\frac{1}{n} H(\epsilon, \mathcal{G}_K, \|\cdot\|_{L_2(\widehat{P}_n)}) \xrightarrow{P} 0 \quad (n \rightarrow \infty)$$

がわかる. 以上のことから補題 6.29 が適用できるので, \mathcal{G}_K は P-GC 族であることがわかる.

II. \mathcal{G} が P-GC 族であることの証明: $G \in L_1(P)$ なので, 任意の $\delta > 0$ に対して, ある $K_0 > 0$ が存在して

$$\int_{G > K_0} G dP \leq \delta$$

とできる. 前の段階の議論から, 関数族 $\{G \mathbb{1}\{G \geq K_0\} \cup \mathcal{G}_{K_0}\}$ は P-GC 族なので, 十分大きな n に対して

$$\sup_{g \in \mathcal{G}} \left| \int_{G \leq K_0} g d(\widehat{P}_n - P) \right| \leq \delta \quad \text{a.s.} \quad \text{かつ} \quad \int_{G > K_0} G d\widehat{P}_n \leq 2\delta \quad \text{a.s.}$$

とできる. 以上のことから

$$\begin{aligned} & \sup_{g \in \mathcal{G}} \left| \int g d(\widehat{\mathbb{P}}_n - \mathbb{P}) \right| \\ & \leq \sup_{g \in \mathcal{G}} \left| \int_{G \leq K_0} g d(\widehat{\mathbb{P}}_n - \mathbb{P}) \right| + \sup_{g \in \mathcal{G}} \left| \int_{G > K_0} g d(\widehat{\mathbb{P}}_n - \mathbb{P}) \right| \\ & \leq \sup_{g \in \mathcal{G}} \left| \int_{G \leq K_0} g d(\widehat{\mathbb{P}}_n - \mathbb{P}) \right| + \int_{G > K_0} G d\widehat{\mathbb{P}}_n + \int_{G > K_0} G d\mathbb{P} \\ & \leq 4\delta \quad \text{a.s.} \end{aligned}$$

がわかる. よって, 定理 6.20 は証明できた. \square

例 6.30. $\mathbb{X} = \mathbb{R}$ とし,

$$\mathcal{G} := \{g : \mathbb{R} \rightarrow \mathbb{R}; 0 \leq g(x) \leq 1 \text{ で } g \text{ は非減少関数}\}$$

とする. $N(\cdot, \mathcal{G}, \|\cdot\|_{L_\infty(\widehat{\mathbb{P}}_n)})$ を半ノルム

$$\max_{1 \leq j \leq n} |g(X_j)|$$

により誘導される擬距離に関する被覆数とする. $\delta > 0$ とする. 関数 $g \in \mathcal{G}$ を

$$\tilde{g} = \left\lceil \frac{g(x)}{\delta} \right\rceil \quad (x \in \mathbb{R})$$

とおく. ただし, $a \geq 0$ に対して, $\lceil a \rceil := \min\{b \in \mathbb{N}; b \geq a\}$ である. すると \tilde{g} は多くとも $m \leq 1 + 1/\delta$ の飛躍点 (X_1, X_2, \dots, X_n のいずれかの点) をもつ. したがって, \tilde{g} を $n-1$ の 0 と m の 1 の列で表現できる. そのような列の個数は

$$\binom{m+n-1}{m}$$

となる. したがって

$$N(\delta, \mathcal{G}, \|\cdot\|_{L_\infty(\widehat{\mathbb{P}}_n)}) \leq \binom{m+n-1}{m}$$

となる. これより

$$\begin{aligned} H(\delta, \mathcal{G}, \|\cdot\|_{L_1(\widehat{\mathbb{P}}_n)}) & \leq \log N_\infty(\delta, \mathcal{G}, \|\cdot\|_{L_\infty(\widehat{\mathbb{P}}_n)}) \\ & \leq m \log(m+n-1) \\ & \leq \left(1 + \frac{1}{\delta}\right) \log\left(n + \frac{1}{\delta}\right) \end{aligned}$$

となるので

$$\frac{1}{n}H(\delta, \mathcal{G}, \|\cdot\|_{L_1(\hat{P}_n)}) \xrightarrow{P} 0 \quad (n \rightarrow \infty)$$

がわかる. この主張と定理 6.20 より, \mathcal{G} は GC 族である.

さらに, $\mathcal{F} := \{\mathbb{1}_{(-\infty, x]}; x \in \mathbb{R}\}$ に対して, $-\mathcal{F} + 1 := \{-g + 1; g \in \mathcal{F}\}$ とおくと $-\mathcal{F} + 1 \subset \mathcal{G}$ となる. さらに

$$\hat{F}_n(x) = \int \mathbb{1}_{(-\infty, x]} d\hat{P}_n, \quad F_n(x) = \int \mathbb{1}_{(-\infty, x]} dP$$

となることに注意する. すると

$$\begin{aligned} \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F| &= \sup_{g \in \mathcal{F}} \left| \int g d(\hat{P}_n - P) \right| \\ &= \sup_{g \in -\mathcal{F} + 1} \left| \int g d(\hat{P}_n - P) \right| \\ &\leq \sup_{g \in \mathcal{G}} \left| \int g d(\hat{P}_n - P) \right| \xrightarrow{P} 0 \quad (n \rightarrow \infty) \end{aligned}$$

となる. したがって, Grevenko-Cantelli の定理

$$\sup_{x \in \mathbb{R}} |\hat{F}(x) - F| \xrightarrow{P} 0 \quad (n \rightarrow \infty)$$

がわかる.

6.8 VC 集合族

\mathcal{D} を \mathbb{X} の部分集合の集まりとする. $n \in \mathbb{N}$ とし, $\{\xi_1, \xi_2, \dots, \xi_n\}$ を \mathbb{X} の点の集合とする.

定義 6.31.

$$\Delta^{\mathcal{D}}(\xi_1, \xi_2, \dots, \xi_n) = \#\{D \cap \{\xi_1, \xi_2, \dots, \xi_n\}; D \in \mathcal{D}\}$$

とする. すなわち, $D \in \mathcal{D}$ によって, $\{\xi_1, \xi_2, \dots, \xi_n\}$ から抽出できる部分集合の個数である. したがって,

$$\Delta^{\mathcal{D}}(\xi_1, \xi_2, \dots, \xi_n) \leq 2^n$$

であり, 上の式で等号が成立する ($\Delta^{\mathcal{D}}(\xi_1, \xi_2, \dots, \xi_n) = 2^n$) とき, \mathcal{D} は $\{\xi_1, \xi_2, \dots, \xi_n\}$ を完全に分離するという. さらに,

$$VC^{\mathcal{D}}(n) = \sup\{\Delta^{\mathcal{D}}(\xi_1, \xi_2, \dots, \xi_n); \xi_1, \xi_2, \dots, \xi_n \in \mathcal{X}\}$$

と書く.

例 6.32. $\mathbb{X} = \mathbb{R}$ とし,

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

とする. すべての $\{\xi_1, \xi_2, \dots, \xi_n\} \subset \mathbb{R}$ に対して,

$$\Delta^{\mathcal{D}}(\xi_1, \xi_2, \dots, \xi_n) \leq n + 1$$

定理 6.33. 次の 2 つは同値である:

(1)

$$\frac{1}{n} \log \Delta^{\mathcal{D}}(X_1, X_2, \dots, X_n) \xrightarrow{P} 0 \quad (n \rightarrow \infty).$$

(2)

$$\sup_{D \in \mathcal{D}} |\hat{P}_n(D) - P(D)| \xrightarrow{P} 0 \quad (n \rightarrow \infty).$$

Proof. (1) \Rightarrow (2) の証明:

$$\mathcal{G} := \{\mathbb{1}_D(x) : \mathcal{X} \rightarrow \mathbb{R} : D \in \mathcal{D}\}$$

とおき, 定理 6.20 を用いる.

まず, 任意の $g \in \mathcal{G}$ に対して

$$|g(x)| \leq 1 \quad (\forall x \in \mathbb{X})$$

であることに注意する. $\delta > 0$ に対して, $N(\delta, \mathcal{G}, L^\infty(P_n))$ を疑ノルム

$$\max_{1 \leq i \leq n} |g(X_i)|$$

から誘導された半ノルムに関する δ 被覆数とする. このとき

$$N(\delta, \mathcal{G}, \|\cdot\|_{L_1(\hat{P}_n)}) \leq N(\delta, \mathcal{G}, \|\cdot\|_{L_\infty(\hat{P}_n)})$$

であることが

$$\|g\|_{L_1(\hat{P})} = \frac{1}{n} \sum_{j=1}^n |g(X_j)| \leq \max_{1 \leq j \leq n} |g(X_j)| = \|g\|_{L_\infty(\hat{P}_n)}$$

からすぐにわかる. しかし, $0 < \delta < 1$ に対して

$$N(\delta, \{\mathbb{1}_D; D \in \mathcal{D}\}, \|\cdot\|_{L_\infty(\hat{P}_n)}) = \Delta^{\mathcal{D}}(X_1, X_2, \dots, X_n)$$

である.

(2) \Rightarrow (1) の証明: もし

$$\frac{1}{n} \log \Delta^{\mathcal{D}}(X_1, X_2, \dots, X_n) \xrightarrow{P} 0 \quad (n \rightarrow \infty)$$

ならば,

$$\frac{1}{n}H(\delta, \{\mathbb{1}_D; D \in \mathcal{D}\}, \|\cdot\|_{L_1(\hat{\mathbb{P}}_n)}) \leq \frac{1}{n} \log \Delta^{\mathcal{D}}(X_1, X_2, \dots, X_n)$$

である. よって

$$\frac{1}{n}H(\delta, \{\mathbb{1}_D; D \in \mathcal{D}\}, \|\cdot\|_{L_1(\hat{\mathbb{P}}_n)}) \xrightarrow{\mathbb{P}} 0 \quad (n \rightarrow \infty)$$

がわかる. □

定義 6.34. $n \in \mathbb{N}$ に対して,

$$\text{VC}^{\mathcal{D}}(n) := \sup\{\Delta^{\mathcal{D}}(x_1, x_2, \dots, x_n); x_1, x_2, \dots, x_n \in \mathbb{X}\}$$

とおく. \mathcal{D} がバプニツク・チェルヴオネンキス (Vapnik-Cheronenkis) 族または VC 族であるとは, ある定数 $C > 0$ と $V > 0$ が存在して, すべての n に対して,

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

が成り立つときをいう.

定理 6.35. \mathcal{D} が VC 族ならば, GC 族.

Proof.

$$\mathcal{G} := \{\mathbb{1}_D(x); D \in \mathcal{D}\}$$

とおく. すると, $\delta > 0$ に対して,

$$\begin{aligned} \frac{1}{n}H(\delta, \mathcal{G}, \|\cdot\|_{L_1(\hat{\mathbb{P}}_n)}) &\leq \frac{1}{n} \log \Delta^{\mathcal{D}}(X_1, X_2, \dots, X_n) \leq \log \text{VC}^{\mathcal{D}}(n) \\ &\leq \frac{1}{n} \log\{C n^V\} = V \frac{\log n}{n} + \frac{C}{n} \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

となる. □

例 6.36. (1) $\mathbb{X} = \mathbb{R}$, $\mathcal{D} = \{(-\infty, r]; r \in \mathbb{R}\}$ は VC 族である. なぜならば, $\text{VC}^{\mathcal{D}}(n) \leq n + 1$ よりわかる.

(2) $p \geq 2$ を自然数とする. $\mathbb{X} = \mathbb{R}^p$, $\mathcal{D} = \{\mathbb{1}_{(-\infty, \mathbf{r}]}(x); \mathbf{r} \in \mathbb{R}^p\}$ は VC 族である. なぜならば, $\text{VC}^{\mathcal{D}}(n) \leq (n + 1)^p$ よりわかる. ただし, $\mathbf{r} = (r_1, r_2, \dots, r_p)$ に対して,

$$(-\infty, \mathbf{r}] = (-\infty, r_1] \times (-\infty, r_2] \times \dots \times (-\infty, r_p]$$

とした.

(3) $\mathbb{X} = \mathbb{R}^p$ とする.

$$\mathcal{D} = \left\{ \mathbf{x} \in \mathbb{R}^p; \boldsymbol{\theta}^\top \mathbf{x} > t, \begin{pmatrix} \boldsymbol{\theta} \\ t \end{pmatrix} \in \mathbb{R}^{p+1} \right\}$$

は VC 族である. なぜならば

$$\text{VC}^{\mathcal{D}}(n) \leq 2^p \binom{n}{p}$$

よりわかる. ただし, $\binom{n}{p}$ は n 個から p 個の組を取り出す組み合わせ数である.

補題 6.37. $\mathcal{D}, \mathcal{D}_1, \mathcal{D}_2$ は VC 族とする. このとき, 以下の族も VC 族である.

- (1) $\mathcal{D}^c := \{D^c; D \in \mathcal{D}\}$.
- (2) $\mathcal{D}_1 \cap \mathcal{D}_2 := \{D_1 \cap D_2; D_1 \in \mathcal{D}_1, D_2 \in \mathcal{D}_2\}$.
- (3) $\mathcal{D}_1 \cup \mathcal{D}_2 := \{D_1 \cup D_2; D_1 \in \mathcal{D}_1, D_2 \in \mathcal{D}_2\}$.
- (4) p を自然数とする. \mathbb{R}^p の開球

$$\{\mathbf{x} \in \mathbb{R}^p; |\mathbf{x} - \boldsymbol{\mu}|^2 \leq c, \boldsymbol{\mu} \in \mathbb{R}^p, c \geq 0\}.$$

Proof. 証明は略. □

定義 6.38. 集合族 \mathcal{D} の VC 次元を

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

で定義する. 上記の式をみたす n が存在しないとき, $\text{VC}^{\mathcal{D}} = \infty$ とする.

補題 6.39. 次の 2 つは同値である:

- (1) \mathcal{D} は VC 族.
- (2) $\text{VC}^{\mathcal{D}} < \infty$.

Proof. $V = \text{VC}^{\mathcal{D}}$ としたとき,

$$\text{VC}^{\mathcal{D}} \leq \sum_{k=0}^V \binom{n}{k}.$$

□

6.9 VC 関数族

定義 6.40. 関数 $g; \mathbb{X} \rightarrow \mathbb{R}$ のサブグラフとは

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

である. 関数族 \mathcal{G} が VC 族であるとは, そのサブグラフの族 $\{\text{subgraph}(g); g \in \mathcal{G}\}$ が VC 族のときをいう.

例 6.41. \mathcal{D} が VC 族ならば, $\mathcal{G} = \{\mathbf{1}_D; D \in \mathcal{D}\}$ も VC 族.

注意 6.42. 自然数 p に対して, $\varphi_k : \mathbb{X} \rightarrow \mathbb{R} (k = 1, 2, \dots, p)$ を固定した関数とする. このとき,

$$\mathcal{G} := \{\theta_1 \varphi_1 + \dots + \theta_p \varphi_p; \theta_1, \dots, \theta_p \in \mathbb{R}\}$$

は VC 族となる. 実際,

$$\mathcal{D} := \{\text{subgraph}(g); g \in \mathcal{G}\}$$

としたとき,

$$\text{VC}^{\mathcal{D}} \leq p + 2$$

である. Pollard (1984) を参照.

定義 6.43. \mathcal{S} は距離空間 (\mathcal{M}, d) の空でない部分集合とする. $\delta > 0$ に対して, \mathcal{S} の δ パッキング数 $D(\delta, \mathcal{S}, \|\cdot\|_d)$ は, 次をみたす最大の自然数 N である. \mathcal{S} の元 s_1, s_2, \dots, s_N が存在して,

$$d(s_k, s_j) > \delta, \quad (\forall k \neq j, k, j = 1, 2, \dots, N).$$

補題 6.44. 任意の $\delta > 0$ に対して以下が成り立つ.

(i)

$$N(\delta, \mathcal{S}, \|\cdot\|_d) \leq D(\delta, \mathcal{S}, \|\cdot\|_d).$$

(ii)

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

Proof. (i) は明らか. (ii) を示すために,

$$\{s_1, s_2, \dots, s_N\} \subset \mathcal{S}, \quad N = D(\delta, \mathcal{S}, \|\cdot\|_d)$$

とする. 明らかに

$$N\left(\frac{\delta}{2}, \mathcal{S}, \|\cdot\|_d\right) \geq N\left(\frac{\delta}{2}, \{s_1, s_2, \dots, s_N\}, \|\cdot\|_d\right).$$

しかし,

$$N\left(\frac{\delta}{2}, \{s_1, s_2, \dots, s_N\}, \|\cdot\|_d\right) = N.$$

□

定理 6.45. \mathbb{Q} を \mathbb{X} 上の確率測度, \mathcal{G} を \mathbb{X} 上の実数値関数の族, G を関数族 \mathcal{G} の封筒関数, $N_1(\delta, \mathcal{G}, \|\cdot\|_{L_1(\mathbb{Q})})$ ($\delta > 0$) を $L_1(\mathbb{Q})$ ノルムから誘導される距離に関する δ 被覆数とする. 族 \mathcal{G} の VC 次元 $VC^{\mathcal{G}} =: V$ が $V < \infty$ のとき, V に依存する定数 $A > 0$ が存在して,

$$N(\delta \mathbb{Q} \mathcal{G}, \mathcal{G}, \|\cdot\|_{L_1(\mathbb{Q})}) \leq \max\{A\delta^{-2V}, e^{\delta/4}\}$$

となる.

Proof. 一般性を失うことなく, $G > 0$ で

$$\mathbb{Q}G := \int_{\mathbb{X}} G(x) d\mathbb{Q}(x) = 1$$

としてよい. \mathbb{X} 値確率要素 S を取り, その分布を

$$\mathbb{Q}_S(A) = \Pr(S \in A) = \int_A G(x) d\mathbb{Q}(x) \quad (A \in \mathcal{B}(\mathbb{X}))$$

とする. $S = s$ を与えたとき, 確率変数 T は区間 $[-G(s), G(s)]$ 上の一様分布に従うとする. $\{g_1, g_2, \dots, g_N\}$ は $g_j: \mathbb{X} \rightarrow \mathbb{R}$ ($j = 1, 2, \dots, N$) で

$$\mathbb{Q}(|g_j - g_k|) > \delta \quad (j \neq k, j, k = 1, 2, \dots, N)$$

をみたす最大の N を持つ集合とする. $S = s$ を与えたとき, 区間

$$[\max\{g_j(s), g_k(s)\}, \min\{g_j(s), g_k(s)\}]$$

上に T が落ちる確率は

$$\frac{|g_j(s) - g_k(s)|}{2G(s)}$$

である. したがって, (S, T) が g_j と g_k のグラフの間に落ちる確率は

$$\begin{aligned} \int_{\mathbb{X}} \frac{|g_j(s) - g_k(s)|}{2G(s)} d\mathbb{Q}_S(x) &= \int_{\mathbb{X}} \frac{|g_j(s) - g_k(s)|}{2} d\mathbb{Q}(x) \\ &= \frac{\mathbb{Q}|g_j - g_k|}{2} > \frac{\delta}{2} \end{aligned}$$

となる. $\{(S_i, T_i)\}_{i=1}^n$ を (S, T) の i.i.d. 複製とする. $(S_1, T_1), (S_2, T_2), \dots, (S_n, T_n)$ のすべてが g_j と g_k のグラフの間に落ちる確率は, 大きくとも

$$\left(1 - \frac{\delta}{2}\right)^n$$

である. ある $j \neq k$ に対して, $(S_1, T_1), (S_2, T_2), \dots, (S_n, T_n)$ のすべてが g_j と g_k のグラフの間に落ちない確率は大きくとも

$$\binom{N}{2} \left(1 - \frac{\delta}{2}\right)^n \leq \frac{N^2}{2} e^{-n\delta/2} = \frac{1}{2} \exp\left(2 \log N - \frac{n\delta}{2}\right)$$

となる. n を大きくして

$$n \geq \frac{4 \log N}{\delta}$$

とすれば,

$$\binom{N}{2} \left(1 - \frac{\delta}{2}\right)^n \leq \frac{1}{2} < 1.$$

そのような n に対して, 任意の $j \neq k (j, k = 1, 2, \dots, N)$ に対して, $(S_i, T_i) (i = 1, 2, \dots, n)$ は g_j と g_k のグラフの間に落ちる確率は正なので,

$$\begin{aligned} D(\delta, \mathcal{G}, \|\cdot\|_{L_1(\mathcal{Q})}) &= N \leq \sup \left\{ \Delta^{\mathcal{D}} \left((S_1, T_1), (S_2, T_2), \dots, (S_n, T_n) \right) \right. \\ &\quad \left. ; (S_i, T_i) \in \mathcal{X} \times \mathbb{R} (i = 1, 2, \dots, n) \right\} \\ &\leq Cn^V \end{aligned}$$

となる. ただし, C は定義 6.34 の定数であり

$$\mathcal{D} = \{\text{subgraph}(g); g \in \mathcal{G}\}$$

である.

しかし, $N \geq \exp\left\{\frac{\delta}{4}\right\}$ のとき

$$\begin{aligned} N &\leq cn^V \leq c \left(\frac{4 \log N}{\delta} + 1 \right)^V \\ &\leq c \left(\frac{\delta \log N}{\delta} \right)^V \\ &= c \left(\frac{16V \log N^{1/(2V)}}{\delta} \right)^V \\ &\leq c \left(\frac{16V}{\delta} \right)^V N^{1/2} \end{aligned}$$

である. よって

$$\sqrt{N} \leq c(16V)^V \delta^{-V}$$

より

$$N \geq e^{\delta/4} \Rightarrow N \leq c^2(16V)^{2V} \delta^{-2V}$$

がわかる. これらを合わせると

$$N(\delta, \mathcal{G}, L_1(\mathcal{Q})) \leq D(\delta, \mathcal{G}, L_1(\mathcal{Q})) \leq \max\{A\delta^{-2V}, e^{\delta/4}\}$$

を得る. $QG = 1$ に規準化していたので,

$$N(\delta QG, \mathcal{S}, \|\cdot\|_{L_1(\mathcal{Q})}) \leq \max\{A\delta^{-2V}, e^{\delta/4}\}$$

がわかる. □

系 6.46. \mathcal{G} を関数族とし, G を関数族 \mathcal{G} の封筒関数とする. 関数族 \mathcal{G} は VC 族で, $PG < \infty$ ならば, \mathcal{G} は GC 族である.

Proof. $PG < \infty$ より, $\mathcal{G} \subset L_1(P)$. 関数族 \mathcal{G} の VC 次元を $V < \infty$ としたとき, 定理 6.45 より, $\delta > 0$ に対して

$$N(\delta PG, \mathcal{G}, \|\cdot\|_{L_1(P)}) \leq \max\{A\delta^{-2\delta}, e^{\delta/4}\}$$

となる. したがって

$$\frac{1}{n} \log N(\delta \hat{P}_n G, \mathcal{G}, \|\cdot\|_{L_1(\hat{P}_n)}) \leq \frac{1}{n} \log (\max\{A\delta^{-2\delta}, e^{\delta/4}\}) \longrightarrow 0 \quad (n \rightarrow \infty)$$

である. さらに

$$|\hat{P}_n G - PG| \xrightarrow{P} 0 \quad (n \rightarrow \infty)$$

に注意すれば, 上の式の左辺の $\delta \hat{P}_n G$ 項の δPG に入れ替えたもので

$$\frac{1}{n} \log N(\delta PG, \mathcal{G}, \|\cdot\|_{L_1(\hat{P}_n)}) \xrightarrow{P} 0 \quad (n \rightarrow \infty)$$

となる. よって, 定理 6.20 より, 主張は証明された. \square

6.10 混合型モデルの最尤推定量の一致性

ここでは, 大数の一様法則を用いて, 混合型モデルの最尤推定量の一致性を示そう.

混合型モデルはつぎのように表現できるものである. Y は実数値確率変数で, 分布関数 F_0 を持つ. $Y = y (y \in \mathbb{R})$ を与えたときに, 実数値確率変数 X は p.d.f. $k(x|y)$ を持つ既知の分布に従うものとする. このとき, X は混合 p.d.f.

$$p_{F_0}(x) := \int k(x|y) dF_0(y) \quad (x \in \mathbb{R}) \quad (6.41)$$

を持つ.

まず, 関数族を添え字づける母数空間がコンパクトならば, その関数族のブラケット付き計量エントロピーは有界であることを示す.

補題 6.47. $p \geq 1$, 距離空間 \mathbb{X} 上の確率測度を P , (Θ, d) をコンパクトな距離空間,

$$\mathcal{G} := \{g_\theta : \mathbb{X} \rightarrow \mathbb{R}; \theta \in \Theta\}$$

とする. 以下を仮定する:

- すべての $x \in \mathbb{X}$ に対して, 写像

$$\Theta \ni \theta \mapsto g_\theta(x) \in \mathcal{G} \subset L_p(\mathbf{P})$$

は連続である. ただし, \mathcal{G} のノルムは $L_p(\mathbf{P})$ とする.

- $G(x) := \sup_{\theta \in \Theta} |g_\theta(x)| \in L_p(\mathbf{P})$ とする.

このとき, $0 < \delta \leq 1$ に対して,

$$H_{p,B}(\delta, \mathcal{G}, \|\cdot\|_{L_p(\mathbf{P})}) < \infty$$

となる.

Proof. $\theta \in \Theta$ と $\rho > 0$ に対して

$$\omega_{\theta,\rho}(x) := \sup_{\tilde{\theta}: d(\theta, \tilde{\theta}) < \rho} |g_\theta(x) - g_{\tilde{\theta}}(x)|$$

と定める. すべての x に対して, 関数

$$\Theta \ni \theta \mapsto g_\theta(x) \in \mathbb{R}$$

は連続なので

$$\lim_{\rho \rightarrow 0} \omega_{\theta,\rho}(x) = 0 \quad (x \in \mathbb{X})$$

となる. $G \in L_p(\mathbf{P})$ なので, 優収束定理 (定理 1.40) より

$$\lim_{\rho \rightarrow 0} \mathbf{P} \omega_{\theta,\rho}^p = 0$$

がわかる. このことから, 任意の $\delta > 0$ に対して, $\rho_\theta > 0$ をうまく取り

$$\mathbf{P} \omega_{\theta,\rho_\theta}^p < \delta^p$$

とできる. つぎに

$$B_\theta := \{\tilde{\theta} \in \Theta : d(\theta, \tilde{\theta}) < \rho_\theta\}$$

とおくと $\{B_\theta; \theta \in \Theta\}$ は Θ の開被覆となる. Θ はコンパクトなので, $N \in \mathbb{N}$ と $\theta_1, \theta_2, \dots, \theta_N \in \Theta$ をうまくとり, 有限部分開被覆

$$\{B_j := \{\tilde{\theta} \in \Theta; d(\theta_j, \tilde{\theta}) < \rho_{\theta_j}\}\}_{j=1}^N$$

を構成できる. 各 $\theta \in B_j$ ($j = 1, 2, \dots, N$) と $x \in \mathcal{X}$ に対して

$$\begin{aligned} g_\theta(x) &\leq g_{\theta_j}(x) + w_{\theta_j, \rho_{\theta_j}}(x) =: g_j^U(x), \\ g_\theta(x) &\geq g_{\theta_j}(x) - w_{\theta_j, \rho_{\theta_j}}(x) =: g_j^L(x), \end{aligned}$$

とおけば

$$P(g_j^U - g_j^L)^p \leq P(2\omega_{\theta_j, \rho_{\theta_j}})^p \leq (2\delta)^p$$

となる. よって

$$H_{p,B}(2\delta, \mathcal{G}, \|\cdot\|_{L_p(P)}) \leq \log N < \infty$$

が示せた. □

いま, 観測 X_1, X_2, \dots, X_n は X の i.i.d. 複製とする. ただし, X は p.d.f. (6.41) をもつ. \mathcal{F} はすべての分布関数の集まりとする. F_0 の最尤推定量 (存在するならば) は

$$\begin{aligned} p_{\hat{F}_n}(x) &:= \arg \max_{F \in \mathcal{F}} \hat{F}_n \log p_F, \\ \hat{F}_n(x) &:= \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{(-\infty, x]}(X_j) \end{aligned}$$

である. 任意の p.d.f. p と \tilde{p} の Hellinger 距離は

$$h(p, \tilde{p}) = \left\{ \frac{1}{2} \int_{\mathbb{R}} (\sqrt{p(x)} - \sqrt{\tilde{p}(x)})^2 dm(x) \right\}^{1/2}$$

であった.

補題 6.48.

$$h(p_{\hat{F}_n}, p_{F_0}) \xrightarrow{P} 0 \quad (n \rightarrow \infty)$$

である.

Proof.

$$\mathcal{H}^* = \{\mu; \mu \text{ は } \mathbb{R} \text{ 上の測度で } \mu(\mathbb{R}) < \infty\}$$

とする. $\mu, \mu_n \in \mathcal{H}^* (n = 1, 2, \dots)$ に対して, 任意の有界連続関数 g に対して,

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}} g(x) d\mu_n(x) = \int_{\mathbb{R}} g(x) d\mu(x)$$

が成立するとき, μ_n は μ に弱収束 (漠収束) するといひ,

$$\mu_n \xrightarrow{w} \mu \quad (n \rightarrow \infty)$$

と書くことにする. すると, 弱位相によって定義される \mathcal{H}^* 上の距離関数 d が存在して

$$d(\mu_n, \mu) \xrightarrow{n \rightarrow \infty} 0 \quad \Leftrightarrow \quad \mu_n \xrightarrow{w} \mu (n \rightarrow \infty)$$

とできる. 加藤 (2016, pp. 264–267) を参照のこと. さらに, (\mathcal{H}^*, d) はコンパクトとなることが知られている.

もし, ほとんどすべての $x \in \mathbb{R}$ に対して,

$$\mathbb{R} \ni y \mapsto k(x|y) \in \mathbb{R}$$

は連続で

$$\lim_{y \rightarrow \pm\infty} k(x|y) = 0$$

であれば, 写像

$$\mathcal{F} \ni F \mapsto p_F(x) = \int_{\mathbb{R}} k(x|y) dF(y)$$

は d に関して連続である. ただし, \mathcal{F} はすべての分布関数の集まりである. さらに, 距離関数 d は

$$d(\hat{F}_n, F_0) \xrightarrow{n \rightarrow \infty} 0 \iff \hat{F}_n \xrightarrow{w} F_0 (n \rightarrow \infty)$$

をみたす \mathcal{H} 上の距離関数である. なぜならば

$$y \mapsto k(x|y)$$

は有界連続関数なので

$$\hat{F}_n \xrightarrow{w} F_0 (n \rightarrow \infty) \Rightarrow \int_{\mathbb{R}} k(x|y) d\hat{F}_n(y) \xrightarrow{n \rightarrow \infty} \int_{\mathbb{R}} k(x|y) dF_0(y) \quad \text{a.e. } x$$

となる. したがって, 優収束定理 (定理 1.40) より

$$\lim_{n \rightarrow \infty} h(p_{\hat{F}_n}, p_{F_0}) = \left\{ \frac{1}{2} \int_{\mathbb{R}} \{ \sqrt{p_{\hat{F}_n}} - \sqrt{p_{F_0}} \}^2 dm \right\}^{1/2} = 0$$

よりわかる. すると, 写像

$$\mathcal{F} \ni F \mapsto \frac{2p_F}{p_F + p_{F_0}}$$

も d に関して連続となる.

いま, つぎの関数族を考える:

$$\mathcal{G} := \left\{ \frac{2p_F}{p_F + p_{F_0}}; F \in \mathcal{F} \right\}.$$

この関数族の包絡関数 G は

$$G(x) \leq 2$$

をみだし, \mathcal{H}^* は弱位相に関してコンパクトであり, 写像

$$\mathcal{F} \ni F \mapsto \frac{2p_F}{p_F + p_{F_0}} \in \mathcal{G}$$

は連続なので, 関数族 \mathcal{G} はコンパクトになる. 補題 6.47 より \mathcal{G} は GC 族である. よって

$$1 - P \left(\frac{2p_{\hat{F}_n}}{p_{\hat{F}_n} + p_{F_0}} \right) \leq (\hat{P}_n - P) \left(\frac{2p_{\hat{F}_n}}{p_{\hat{F}_n} + p_{F_0}} \right). \quad (6.42)$$

これは

$$\begin{aligned} 0 &\leq \hat{P}_n \left(\log \frac{2p_{\hat{F}_n}}{p_{\hat{F}_n} + p_{F_0}} \right) \\ &\leq \hat{P}_n \left(\frac{2p_{\hat{F}_n}}{p_{\hat{F}_n} + p_{F_0}} \right) - 1 \\ &= (\hat{P}_n - P) \left(\frac{2p_{\hat{F}_n}}{p_{\hat{F}_n} + p_{F_0}} \right) + P \left(\frac{2p_{\hat{F}_n}}{p_{\hat{F}_n} + p_{F_0}} \right) - 1. \end{aligned}$$

からわかる. しかし

$$h^2(p, p_0) \leq 1 - P \left(\frac{2p}{p + p_0} \right) \quad (6.43)$$

である. これは

$$\begin{aligned} h^2(p, p_0) &= \frac{1}{2} \int_{\mathbb{R}} \{\sqrt{p} - \sqrt{p_0}\}^2 dm \\ &= \frac{1}{2} \int_{\mathbb{R}} \frac{(p - p_0)^2}{\{\sqrt{p} + \sqrt{p_0}\}^2} dm \\ &\leq \frac{1}{2} \int_{\mathbb{R}} \frac{(p - p_0)^2}{\{p + p_0\}^2} dm \\ &= \frac{1}{2} \int_{\mathbb{R}} \frac{(p - p_0)^2}{\{p + p_0\}^2} dm + \frac{1}{2} \int_{\mathbb{R}} \frac{p - p_0}{\{p + p_0\}^2} dm \\ &= \frac{1}{2} \int_{\mathbb{R}} \frac{p_0(p - p_0)}{\{p + p_0\}^2} dm \\ &= \frac{1}{2} \int_{\mathbb{R}} \left(1 - \frac{2p}{p + p_0} \right)^2 dm \\ &= 1 - P \left(\frac{2p}{p + p_0} \right) \end{aligned}$$

からわかる. (6.42) と (6.43) より

$$\begin{aligned} h^2(p_{\hat{F}_n}, p_{F_0}) &\leq (\hat{P}_n - P) \left(\frac{2p_{\hat{F}_n}}{p_{\hat{F}_n} + p_{F_0}} \right) \\ &\leq \sup_{g \in \mathcal{G}} |(\hat{P}_n - P)g| = \|\hat{P}_n - P\|_{\mathcal{G}} \xrightarrow{P} 0 \quad (n \rightarrow \infty) \end{aligned}$$

となる. \mathcal{G} が GC 族なので, 最後は保証される. したがって

$$h(\mathbf{p}_{\hat{F}_n}, \mathbf{p}_{F_0}) \xrightarrow{P} 0 \quad (n \rightarrow \infty)$$

を得る.

□