第6章　経験過程の収束

6.1　導入

$X, X_1, X_2, \ldots, X_n$ 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 $\overline{X}_n = \dfrac{1}{n}\sum_{j=1}^n X_j$ に対して, 大数の強法則と中心極限定理を適用すると, $n \to \infty$ のとき

$$\overline{X}_n \overset{\text{a.s.}}{\to} \mathsf{E}[X] \quad と \quad \sqrt{n}\big(\overline{X}_n - \mathsf{E}[X]\big) \rightsquigarrow \mathsf{N}\big(0, \sigma^2\big)$$

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 $\mathsf{E}[g^2(X)] < \infty$ なるものに対して, $n \to \infty$ のとき

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

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 $\widehat{g}_n$ に対する経験過程 $\dfrac{1}{n}\sum_{j=1}^n \widehat{g}_n(X_j)$ を考察する必要が出てくる. すなわち, $\mathcal{G}$ を実数値関数のある族としたとき, $g \in \mathcal{G}$ で添え字付けれらた経験過程

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

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

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

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

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

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

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

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

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

を観測したとする. ただし, $Y_j\in\{0,\,1\}\,(j=1,\,2,\,\ldots,\,n)$ 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

$$g^*(z)=\mathsf{Pr}\big(Y=1|\,Z=z\big)$$

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

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

[1] この定義については, 定義 8.18 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.

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

である.

ここで, 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

$$\bigl\|\widehat{g}_n-g^*\bigr\|_{L_2(\mathsf{Q})}=\left(\int_{\mathbb{R}}\bigl(\widehat{g}_n(z)-g^*(z)\bigr)^2\,\mathrm{d}\mathsf{Q}(z)\right)^{1/2}$$

を考えることができる. 後に説明する道具を用いると

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

となることがわかる.

目的関数の仮定として

$$\Lambda_2:=\left\{g:\mathbb{R}\longrightarrow(0,1);\ 0\le\frac{\mathrm{d}g}{\mathrm{d}z}\le M,\ g\text{ は concave}\right\}$$

を考えることもできる. この文脈では

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

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

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

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

$$\bigl\|\widehat{g}_n-g^*\bigr\|_{L_2(\mathsf{Q})}\le C|\widehat{\theta}_n-\theta^*|=O_P\left(\frac{1}{\sqrt{n}}\right)$$

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)

$$\{\mathsf{p}_\theta;\, \theta \in \Theta\}$$

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 $X_1, X_2, \ldots, X_n$ 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.

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

Hellinger 距離は

$$\mathsf{KL}(\mathsf{p}, \mathsf{q}) = \int_{\mathbb{X}} \log\left(\frac{\mathsf{p}(x)}{\mathsf{q}(x)}\right) \mathsf{p}(x)\, \mathrm{d}\mathsf{m}(x) \tag{6.3}$$

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, $\mathsf{KL}(\mathsf{p}, \mathsf{q}) = \infty$ と約束する. □

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

Proof. $v > 0$ に対して, 不等式

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

が成立することに注意する. (1) の証明:

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

から示せた.
(2) の証明:

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

から示せた. □

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

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

である.

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

ンダム標本としたとき, p_{θ^*} の最尤推定量は

$$\mathsf{p}_{\widehat{\theta}_n} \in \arg\min_{\theta\in\Theta}\left\{\sum_{j=1}^n \log\left(\frac{\mathsf{p}_{\theta^*}(X_j)}{\mathsf{p}_\theta(X_j)}\right)\right\}$$
$$\Leftrightarrow\ \mathsf{p}_{\widehat{\theta}_n} \in \arg\max_{\theta\in\Theta}\left\{\sum_{j=1}^n \log \mathsf{p}_\theta(X_j)\right\}$$

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

$$\begin{aligned}0 &\geq \frac{1}{n}\sum_{j=1}^n \log \mathsf{p}_{\theta^*}(X_j) - \frac{1}{n}\sum_{j=1}^n \log \mathsf{p}_{\widehat{\theta}_n}(X_j) = \frac{1}{n}\sum_{j=1}^n \log\left(\frac{\mathsf{p}_{\theta^*}(X_j)}{\mathsf{p}_{\widehat{\theta}_n}(X_j)}\right)\\ &= \frac{1}{n}\sum_{j=1}^n \log\left(\frac{\mathsf{p}_{\theta^*}(X_j)}{\mathsf{p}_{\widehat{\theta}_n}(X_j)}\right) - \mathsf{KL}\big(\mathsf{p}_{\theta^*},\, \mathsf{p}_{\widehat{\theta}_n}\big) + \mathsf{KL}\big(\mathsf{p}_{\theta^*},\, \mathsf{p}_{\widehat{\theta}_n}\big)\end{aligned}$$

となる. したがって

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

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

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

が成り立つこと [4] を知っている. このことから, 中心極限定理の一様バージョンを導出できれば, $\mathsf{KL}\big(\mathsf{p}_{\theta^*},\, \mathsf{p}_{\widehat{\theta}_n}\big)$ 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

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

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 $\mathrm{Cover}(A, \epsilon)$ 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 $\mathrm{Centers}(\mathcal{C}, \epsilon)$ 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

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

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 に含まれるかいなかは重要ではない. 実際, A の被覆 $\bigcup_{j=1}^N B_d(x_j, \epsilon)(\supset A)$ 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. □

6.3.1 関数族のエントロピー

$\mathbb{R}$ 上の確率測度を Q とし, $1 \le p < \infty$ 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

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

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 $H(\epsilon, \mathcal{G}, \mathsf{Q})$ や $H\bigl(\epsilon, \mathcal{G}, \|\cdot\|_{p,\mathsf{Q}}\bigr)$ とも記すことにする. この場合には, $\mathcal{G}$ は距離空間

$$L_p(\mathsf{Q}) := \left\{g : \mathbb{R} \longrightarrow \mathbb{R};\ g \text{ は可測関数で } \textstyle\int_{\mathbb{R}} |g|^p \,\mathrm{d}\mathsf{Q} < \infty\right\}$$

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.

$N(\epsilon, A,$ 距離関数$)$ と $N(\epsilon, A,$ 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, 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, $N_{p,B}(\epsilon, \mathcal{G}, \mathsf{Q})$ 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 $j = 1, 2, \ldots, N$ に対して
$$\|g_j^L - g_j^R\|_{L_p(\mathsf{Q})} \le \epsilon.$$
- すべての $g \in \mathcal{G}$ に対して, ある $j \in \{1, 2, \ldots, N\}$ が存在して
$$g_j^L(x) \le g(x) \le g_j^R(x) \qquad (x \in \mathbb{R})$$
が成り立つ.

$H_{p,B}(\epsilon, \mathcal{G}, \mathsf{Q}) = \log N_{p,B}(\epsilon, \mathcal{G}, \mathsf{Q})$ 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)
$$H_p(\epsilon, \mathcal{G}, \mathsf{Q}) \le H_{p,B}(\epsilon, \mathcal{G}, \mathsf{Q})$$
(2)
$$H_{p,B}(\epsilon, \mathcal{G}, \mathsf{Q}) \le H_\infty\left(\frac{\epsilon}{2}, \mathcal{G}.\right)$$
(3) $A \subset \mathbb{D}$ とし, d, d' を $\mathbb{D}$ 上の擬距離で
$$d(x, y) \le d'(x, y) \qquad (\forall x, y \in \mathbb{D})$$
が成立するとする. このとき
$$H(\epsilon, A, d) \le H(\epsilon, A, d')$$
が成り立つ.

Proof. (1) $\epsilon > 0$ と $1 \le p < \infty$ とし, 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. $\log N := H_{p,B}(\epsilon, \mathcal{G}, \mathsf{Q})$ 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, $\exists j_0 \in \{1, 2, \ldots, N\}$ があって
$$g_{j_0}^L(x) \le g(x) \le g_{j_0}^R(x) \qquad \text{かつ} \qquad \|g_{j_0}^L - g_{j_0}^R\|_{L_p(\mathsf{Q})} \le \epsilon$$

なので

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

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

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

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

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

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

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

となるので

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

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

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

となるので

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

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 $\{c_1, c_2, \ldots, c_N\}$ で

- 任意の $j \ne k$ に対して, $\|c_j - c_k\| \ge \epsilon$,
- 集合 $\{c_1, c_2, \ldots, c_N\}$ 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. $1 \le j \le N$ 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. □

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,

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

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

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

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

$$B_d\left(c_j, \frac{\epsilon}{2}\right) \bigcap B_d\left(c_k, \frac{\epsilon}{2}\right) = \varnothing \tag{6.4}$$

となる. また

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

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

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

を得る. ただし, $\mathsf{vol}(B_d(0_d, 1))$ は $\mathbb{R}^d$ の単位球の体積で $\dfrac{(2\pi)^{d/2}}{d\Gamma(d/2)}$ で与えられる. (6.6) の両辺を $\left(\dfrac{\epsilon}{2}\right)^d$ で割って整理すると

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

を得る. □

問 6.1. (6.4) と (6.5) を確認せよ.

例 6.11. □

例 6.12.

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

とする. このとき, ある $A > 0$ が存在して

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

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, $a_k = k\epsilon\,(k = 0, 1, \ldots, N-1)$, $a_N = 1$ とおく. $B_k = (a_{k-1}, k)\,(k = 1, 2, \ldots, N)$ とおき関数 $\widetilde{g} : [0, 1] \longrightarrow [0, 1]$ を

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

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

$$\sup_{x\in[0,1]} |g - \widetilde{g}| \leq 2\epsilon \tag{6.8}$$

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

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

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

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

となることがわかる. よって

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

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 $g : \mathcal{I} \to \mathbb{R}$ が Lipschitz 係数 $L\,(>0)$ の Lipschitz 関数とする. 任意の $x,\,x' \in \mathcal{I}$ に対し,

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

をみたすことである.

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

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

とする.

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

$$a_j - a_{j-1} \le \epsilon \qquad (j = 1,\,2,\,\ldots,\,N)$$

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

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

を得る.

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

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

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

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, $g \in \mathcal{G}$ が動くとき, 関数 $\tilde{g}$ の個数の最大は

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

よって, $0 < \epsilon < 1$ に対して,

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

□

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, $X, X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim}$ P とする. 関数族 $\mathcal{G}$ は P-Glivenko-Cantelli 族であるとは

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

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, $H_B(\epsilon, \mathcal{G}, \|\cdot|_{L_1(\mathsf{P})}) < \infty$ 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 $N := N_{1,B}(\epsilon, \mathcal{G}, \mathsf{P}) < \infty$ 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

$$\max_{j=1,2,\ldots,N} \|g_j^L - g_j^R\|_{L_1(\mathsf{P})} \le \epsilon$$

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

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

となる. したがって

$$\begin{aligned}
\int g \, \mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big) &= \int g \, \mathrm{d}\widehat{\mathsf{P}}_n - \int g \, \mathrm{d}\mathsf{P} \\
&\le \int g_{j_0}^R \, \mathrm{d}\widehat{\mathsf{P}}_n - \int g \, \mathrm{d}\mathsf{P} \\
&= \int g_{j_0}^R \, \mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big) + \int (g_{j_0}^R - g) \, \mathrm{d}\mathsf{P} \\
&\le \int g_{j_0}^R \, \mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big) + \int |g_{j_0}^R - g| \, \mathrm{d}\mathsf{P} \\
&\le \int g_{j_0}^R \, \mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big) + \int |g_{j_0}^R - g_{j_0}^L| \, \mathrm{d}\mathsf{P} \\
&\le \int g_{j_0}^R \, \mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big) + \epsilon
\end{aligned} \tag{6.10}$$

となる. 同様に

$$\begin{aligned}
\int g \, \mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big) &= \int g \, \mathrm{d}\widehat{\mathsf{P}}_n - \int g \, \mathrm{d}\mathsf{P} \\
&\ge \int g_{j_0}^L \, \mathrm{d}\widehat{\mathsf{P}}_n - \int g \, \mathrm{d}\mathsf{P} \\
&= \int g_{j_0}^L \, \mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big) + \int (g_{j_0}^L - g) \, \mathrm{d}\mathsf{P} \\
&\ge \int g_{j_0}^L \, \mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big) - \int |g_{j_0}^L - g| \, \mathrm{d}\mathsf{P} \\
&\ge \int g_{j_0}^L \, \mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big) - \int |g_{j_0}^L - g_{j_0}^R| \, \mathrm{d}\mathsf{P} \\
&\ge \int g_{j_0}^L \, \mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big) - \epsilon
\end{aligned} \tag{6.11}$$

となる. (6.10) と (6.11) を合わせると

$$\begin{aligned}
-\max_{j=1,2,\ldots,N} \left| \int g_{j_0}^L \, \mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big) \right| - \epsilon &\le \int g \, \mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big) \\
&\le \max_{j=1,2,\ldots,N} \left| \int g_{j_0}^R \, \mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big) \right| + \epsilon
\end{aligned} \tag{6.12}$$

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

$$\max_{j=1,2,\ldots,N}\left|\int g_j^L \,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)\right| \overset{\text{a.s.}}{\to} 0 \qquad (n\to\infty)$$
$$\max_{j=1,2,\ldots,N}\left|\int g_j^R \,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)\right| \overset{\text{a.s.}}{\to} 0 \qquad (n\to\infty)$$

となる. よって, 十分大きな n に対して

$$\mathsf{Pr}\left(\max_{j=1,2,\ldots,N}\left|\int g_j^L \,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)\right| \le \epsilon\right) = 1 \tag{6.13}$$

と

$$\mathsf{Pr}\left(\max_{j=1,2,\ldots,N}\left|\int g_j^R \,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)\right| \le \epsilon\right) = 1 \tag{6.14}$$

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

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

が成立することがわかる. すなわち

$$\lim_{n\to\infty}\mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|\int g \,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)\right| \le 2\epsilon\right) = 1$$

がわかる. よって, 定理は証明された. □

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

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

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, $H_{1,B}(\epsilon, \mathcal{G}, \mathsf{P}) < \infty$ ならば, $G\in L_1(\mathsf{P})$ である.

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

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

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

[6] あとの都合で, $N := H_1(\epsilon, \mathcal{G}, \mathsf{P})$ 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 $\mathcal{G} \ni g \mapsto \|g\|_{L_1(\mathsf{P})}$ は連続なので, この写像による $\mathcal{G}$ の像 $\{\|g\|_{L_1(\mathsf{P})};\, g \in \mathcal{G}\}\,(\subset \mathbb{R})$ もコンパクトとなる. よって, ある $R > 0$ が存在して

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

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 $j \in \{1,\, 2,\, \ldots,\, N\}$ が存在して

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

とできる. このことから

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

がわかる. よって, 主張は証明された. □

命題 6.19. 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 P-GC 族で $L_1(\mathsf{P})$ 有界ならば, $G \in L_1(\mathsf{P})$ である.

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

$$\frac{1}{n} \sup_{g\in\mathcal{G}} \left|\widehat{\mathsf{P}}_n g - \mathsf{P}g\right| \overset{\text{a.s.}}{\to} 0$$

となる.

$$\begin{aligned}
\frac{1}{n}\left|g(X_n) - \mathsf{P}g\right| &= \frac{1}{n}\left|n\widehat{\mathsf{P}}_n - (n-1)\widehat{\mathsf{P}}_{n-1} - \mathsf{P}g\right| \\
&\le \left|\widehat{\mathsf{P}}_n - \mathsf{P}g\right| + \frac{n-1}{n}\left|\widehat{\mathsf{P}}_{n-1} - \mathsf{P}g\right|
\end{aligned}$$

(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

$$\sup_{g\in\mathcal{G}} \frac{1}{n}\bigl|g(X_n) - \mathsf{P}g\bigr| \le \sup_{g\in\mathcal{G}} \bigl|\widehat{\mathsf{P}}_n - \mathsf{P}g\bigr| + \frac{n-1}{n} \sup_{g\in\mathcal{G}} \bigl|\widehat{\mathsf{P}}_{n-1} - \mathsf{P}g\bigr| \overset{\text{a.s.}}{\to} 0$$

がわかる. よって

$$\mathsf{Pr}\Bigl(\sup_{g\in\mathcal{G}} \bigl|g(X_n) - \mathsf{P}g\bigr| \ge n, \text{i.o.}\Bigr) = 0 \tag{6.15}$$

となる. したがって, 補題 1.44 の対偶から

$$\sum_{n=1}^{\infty} \mathsf{Pr}\bigl(\sup_{g\in\mathcal{G}} \bigl|g(X_n) - \mathsf{P}g\bigr| \ge n\bigr) < \infty$$

がわかる. よって

$$\begin{aligned}\mathsf{E}\bigl[\sup_{g\in\mathcal{G}} \bigl|g(X_n) - \mathsf{P}g\bigr|\bigr] &= \int_0^{\infty} \mathsf{Pr}\bigl(\sup_{g\in\mathcal{G}} \bigl|g(X_n) - \mathsf{P}g\bigr| > t\bigr)\,\mathrm{d}t \qquad (\because \text{命題 } 1.31) \\ &\le \sum_{n=1}^{\infty} \mathsf{Pr}\bigl(\sup_{g\in\mathcal{G}} \bigl|g(X_n) - \mathsf{P}g\bigr| \ge n\bigr) < \infty \end{aligned} \tag{6.16}$$

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

$$\mathsf{E}[G] \le \mathsf{E}\Bigl[\sup_{g\in\mathcal{G}} \bigl|g(X_n) - \mathsf{P}g\bigr|\Bigr] + \sup_{g\in\mathcal{G}} \bigl|\mathsf{P}g\bigr| < \infty$$

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. 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. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P}$ に対して, $\widehat{\mathsf{P}}_n$ をこれらに基づく経験測度とする. すなわち

$$\widehat{\mathsf{P}}_n(B) = \frac{\#\{j \in \{1, 2, \ldots, n\};\ X_j \in B\}}{n} \qquad (B \in \mathcal{B}(\mathbb{R}))$$

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

$$\frac{1}{n} H(\delta, \mathcal{G}, \|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)}) \overset{\mathrm{P}}{\to} 0 \qquad (n \to \infty) \tag{6.17}$$

が成り立つとき, $\mathcal{G}$ は 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

$$\mathsf{Pr}\left(\left|\int g\,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)\right| > \frac{\delta}{2}\right) \le \frac{1}{2} \tag{6.18}$$

が成立すると

$$\mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|g\,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)\right| > \delta\right) \le 4\mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{j=1}^{n}\epsilon_j g_(X_j)\right| > \frac{\delta}{4}\right) \tag{6.19}$$

が成立する. ただし, $X_j\ (j = 1, 2, \ldots, n)$ とは独立は $\{\epsilon_j\}_{j=1}^n$ 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 $X_1, X_2, \ldots, X_n$ を条件付きにして, $\dfrac{1}{n}\sum_{j=1}^n \epsilon_j g_(X_j)$ に対して適用すると, $t > 0$ に対して

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

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

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

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) をみたす $\mathcal{G}$ に対して

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

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, $\mathcal{G}_K := \{g\mathbb{1}\{G \le K\}$ 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, $G \in L_1(\mathsf{P})$ だから, K_0 をうまく取ると

$$\int_{G>K_0} G\,\mathrm{d}\mathsf{P} < \delta$$

[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, ここで得らえた結果を用いる.

とでき, $\{G\mathbb{1}\{G \geq K_0\} \bigcup \mathcal{G}_{K_0}$ 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 Dudely 積分で上から評価できることを証明 (定理 6.27) する.
Step 3. $X_1, X_2, \ldots, X_n$ 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

$$\frac{1}{n}H(\delta, \mathcal{G}, \|\cdot\|_{L_2(\widehat{\mathsf{P}}_n)}) \overset{\mathrm{P}}{\to} 0 \qquad (n \to \infty) \tag{6.23}$$

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

$$\frac{1}{n}H\big(\delta, \mathcal{G}_K, \|\cdot\|_{L_2(\widehat{\mathsf{P}})}\big) \overset{\mathrm{P}}{\to} 0 \qquad (n \to \infty)$$

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. □

注意 6.21. 以下の証明では, 定理の仮定のもと

$$\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{j=1}^{n} g(X_j) - \int g\,\mathrm{d}\mathsf{P}\right| \overset{\mathrm{P}}{\to} 0 \qquad (n \to \infty) \tag{6.24}$$

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

$$\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{j=1}^{n} g(X_j) - \int g\,\mathrm{d}\mathsf{P}\right| \overset{\mathrm{a.s.}}{\to} 0 \qquad (n \to \infty)$$

が成立することが知られている. □

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. $X_1, X_2, \ldots, X_n$ 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, $\mathsf{E}[X_{i,s}] = 0$ である. ただし,

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

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

$$\frac{1}{2}\mathsf{E}\left[\sup_{s\in\mathcal{T}}\left|\sum_{j=1}^n \epsilon_j X_{j,s}\right|\right] \le \mathsf{E}\left[\sup_{s\in\mathcal{T}}\left|\sum_{j=1}^n X_{j,s}\right|\right] \le 2\mathsf{E}\left[\sup_{s\in\mathcal{T}}\left|\sum_{j=1}^n \epsilon_j X_{j,s}\right|\right] \tag{6.25}$$

と

$$\mathsf{E}\left[\sup_{s\in\mathcal{T}}\sum_{j=1}^n X_{j,s}\right] \le 2\mathsf{E}\left[\sup_{s\in\mathcal{T}}\sum_{j=1}^n \epsilon_j X_{j,s}\right] \tag{6.26}$$

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

$$\begin{aligned}
\mathsf{E}\left[\sup_{s\in\mathcal{T}}\left|\sum_{j=1}^n X_{j,s}\right|\right] &= \mathsf{E}\left[\sup_{s\in\mathcal{T}}\left|\sum_{j=1}^n \left\{X_{j,s} - \mathsf{E}\left[X_{j,s}'\right]\right\}\right|\right] \\
&= \mathsf{E}\left[\sup_{s\in\mathcal{T}}\left|\mathsf{E}\left[\sum_{j=1}^n \left\{X_{j,s} - X_{j,s}'\right\}\middle| X_j'\right]\right|\right] \\
&\le \mathsf{E}\left[\sup_{s\in\mathcal{T}}\left|\sum_{j=1}^n \left\{X_{j,s} - X_{j,s}'\right\}\right|\right] \\
&\qquad (\because \text{Jensen の不等式と towering propaty}) \\
&= \mathsf{E}\left[\sup_{s\in\mathcal{T}}\left|\sum_{j=1}^n \epsilon_j\left\{X_{j,s} - X_{j,s}'\right\}\right|\right] \\
&\qquad (\because X_{j,s} - X_{j,s}' \text{ の分布の対称性}) \\
&\le 2\mathsf{E}\left[\sup_{s\in\mathcal{T}}\left|\sum_{j=1}^n \epsilon_j X_{j,s}\right|\right]
\end{aligned}$$

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

$$\begin{aligned}
\mathsf{E}\left[\sup_{s\in\mathcal{T}}\left|\mathsf{E}\left[\sum_{j=1}^n \left\{X_{j,s} - X_{j,s}'\right\}\middle| X_j'\right]\right|\right] &\le \mathsf{E}\left[\sup_{s\in\mathcal{T}}\mathsf{E}\left[\left|\sum_{j=1}^n \left\{X_{j,s} - X_{j,s}'\right\}\right|\middle| X_j'\right]\right] \\
&\le \mathsf{E}\left[\mathsf{E}\left[\sup_{s\in\mathcal{T}}\left|\sum_{j=1}^n \left\{X_{j,s} - X_{j,s}'\right\}\right|\middle| X_j'\right]\right] \\
&= \mathsf{E}\left[\sup_{s\in\mathcal{T}}\left|\sum_{j=1}^n \left\{X_{j,s} - X_{j,s}'\right\}\right|\right]
\end{aligned}$$

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. □

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

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

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

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

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

Proof. $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ 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

$$A_g := \left\{\boldsymbol{X};\ \left|\int g\,\mathrm{d}\left(\widehat{\mathsf{P}}_n - \mathsf{P}\right)\right| > \delta\right\}$$

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

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

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

[11] (6.18) から

$$\frac{1}{2} \leq 1 - \mathsf{Pr}\left(\left|\int g\,\mathrm{d}\left(\widehat{\mathsf{P}}_n - \mathsf{P}\right)\right| > \frac{\delta}{2}\right) = \mathsf{Pr}\left(\left|\int g\,\mathrm{d}\left(\widehat{\mathsf{P}}_n - \mathsf{P}\right)\right| \leq \frac{\delta}{2}\right)$$

となることに注意する.

と $\widehat{\mathsf{P}}'_n$ の独立性から

$$
\begin{aligned}
&\mathsf{Pr}\left(\boldsymbol{X} \in A_{g^*} \text{ かつ } \left|\int g^* \,\mathrm{d}\big(\widehat{\mathsf{P}}'_n - \mathsf{P}\big)\right| \le \frac{\delta}{2}\right) \\
&= \mathsf{E}\left[\mathbb{1}\left\{\{\boldsymbol{X} \in A_{g^*}\}\bigcap\left\{\left|\int g^* \,\mathrm{d}\big(\widehat{\mathsf{P}}'_n - \mathsf{P}\big)\right| \le \frac{\delta}{2}\right\}\right\}\right] \\
&= \mathsf{E}_{\boldsymbol{X}}\left[\mathbb{1}\left\{\{\boldsymbol{X} \in A_{g^*}\}\mathsf{E}\left[\mathbb{1}\left\{\left|\int g^* \,\mathrm{d}\big(\widehat{\mathsf{P}}'_n - \mathsf{P}\big)\right| \le \frac{\delta}{2}\right\}\middle|\boldsymbol{X}\right]\right]\right. \\
&= \mathsf{E}_{\boldsymbol{X}}\left[\mathbb{1}\left\{\{\boldsymbol{X} \in A_{g^*}\}\underbrace{\mathsf{Pr}\left(\left|\int g^* \,\mathrm{d}\big(\widehat{\mathsf{P}}'_n - \mathsf{P}\big)\right| \ge \frac{\delta}{2}\right\}\middle|\boldsymbol{X}\right)}_{\ge 1/2 \quad \because (6.18)}\right] \\
&\ge \frac{1}{2}\mathsf{Pr}\big(A_{g^*}\big) \\
&= \frac{1}{2}\mathsf{Pr}\left(\left|\int g^* \,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)\right| > \delta\right)
\end{aligned}
\tag{6.27}
$$

となる．この不等式を用いると

$$
\begin{aligned}
&\mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|\int g \,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)\right| > \delta\right) = \mathsf{Pr}\left(\boldsymbol{X} \in \bigcup_{g\in\mathcal{G}} A_g)\right) \\
&\le \mathsf{Pr}\big(\boldsymbol{X} \in A_{g^*}\big) \\
&= \mathsf{Pr}\left(\left|\int g^* \,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)\right| > \delta\right) \\
&\le 2\mathsf{Pr}\left(\boldsymbol{X} \in A_{g^*} \text{ かつ } \left|\int g^* \,\mathrm{d}\big(\widehat{\mathsf{P}}'_n - \mathsf{P}\big)\right| \le \frac{\delta}{2}\right) \qquad (\because (6.27)) \\
&= 2\mathsf{Pr}\left(\left\{\boldsymbol{X} \in A_{g^*}\right\}\bigcap\left\{\left|\int g^* \,\mathrm{d}\big(\widehat{\mathsf{P}}'_n - \mathsf{P}\big)\right| \le \frac{\delta}{2}\right\}\right) \\
&= 2\mathsf{Pr}\left(\left\{\left|\int g^* \,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)\right| > \delta\right\}\bigcap\left\{\left|\int g^* \,\mathrm{d}\big(\widehat{\mathsf{P}}'_n - \mathsf{P}\big)\right| \le \frac{\delta}{2}\right\}\right) \\
&\le 2\mathsf{Pr}\left(\left|\int g^* \,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \widehat{\mathsf{P}}'_n\big)\right| > \frac{\delta}{2}\right) \\
&\le 2\mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|\int g \,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \widehat{\mathsf{P}}'_n\big)\right| > \frac{\delta}{2}\right)
\end{aligned}
$$

がわかる．最初の不等号は

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

なので

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

からわかる. また, 最後の不等号は

$$\left|\int g^* \,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)\right| > \delta \text{ かつ } \left|\int g^* \,\mathrm{d}\big(\widehat{\mathsf{P}}'_n - \mathsf{P}\big)\right| \le \frac{\delta}{2}$$

ならば

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

であること [12] からわかる. □

系 6.24. $\mathcal{G}$ を関数族, P を $\mathbb{R}$ 上の確率測度, $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P}$ とする. $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ を Rademacher 確率変数列で $X_j\,(j = 1, 2, \ldots, n)$ 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

$$\mathsf{Pr}\left(\left|\int g \,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)\right| > \frac{\delta}{2}\right) \le \frac{1}{2} \tag{6.28}$$

とする. このとき

$$\mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|\int g \,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)\right| > \delta\right) \le 4\mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{j=1}^{n} \epsilon g(X_j)\right| > \frac{\delta}{4}\right)$$

が成り立つ.

[12] 三角不等式と条件から

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

から

$$\left|\int g^* \,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \widehat{\mathsf{P}}'_n\big)\right| \le \frac{\delta}{2}$$

がわかる.

Proof. 補題 6.23 から

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

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

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

ならば

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

となる．したがって

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

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

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

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

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

と仮定する．このとき, $t>0$ に対して

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

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

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

が成り立つ．

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

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

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

$$|\epsilon_j g(x)|\le K \qquad (x\in\mathbb{R})$$

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

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

を得る．さらに, towering property から

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

を得る. 次に, 上の不等式と union bound を用いると

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

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.

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

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

$$
\mathsf{Var}[g(X_1)] \le \mathsf{E}[\{g(X_1)\}^2] \le K^2
$$

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

$$
\begin{aligned}
\mathsf{Pr}\left(\left|\int g\,\mathrm{d}\big(\widehat{\mathsf{P}}_n-\mathsf{P}\big)\right| > \delta\right) &= \mathsf{Pr}\left(\left|\frac{1}{n}\sum_{j=1}^n\Big\{g(X_j)-\mathsf{E}[g(X_j)]\Big\}\right| > \delta\right) \\
&\le \frac{\mathsf{Var}[g(X_1)]}{n\delta^2} \le \frac{K^2}{n\delta^2} \le \frac{1}{2}
\end{aligned}
$$

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

$$
\begin{aligned}
&\mathsf{Pr}\left(\max_{g\in\mathcal{G}}\left|\int g\,\mathrm{d}\big(\widehat{\mathsf{P}}_n-\mathsf{P}\big)\right| > K\sqrt{\frac{2(t+\log N)}{n}}\right) \\
&\quad\le 4\mathsf{Pr}\left(\max_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{j=1}^n \epsilon_j g(X_j)\right| > \frac{K}{4}\sqrt{\frac{2(t+\log N)}{n}}\right) \le 8e^{-t}
\end{aligned}
$$

を得る. □

6.6.2 Dudley のエントロピー積分

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

$$\mathsf{E}\left[\max_{j=1,2,\ldots,N} X_j\right] \le \sqrt{2\nu\log N}$$

が成立する.

Proof. Jensen の不等式 (定理 1.36) を用いると

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

を得る. 両辺の対数を取ると

$$\mathsf{E}\left[\max_{j=1,2,\ldots,N} X_j\right] \le \frac{\log N}{\lambda} + \frac{\lambda\nu}{2}$$

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

$$\mathsf{E}\left[\max_{j=1,2,\ldots,N} X_j\right] \le \sqrt{2\nu\log(N)}$$

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

$$\log\left(\mathsf{E}\big[\exp\{\lambda(X_{t_1} - X_{t_2})\}\big]\right) \le \frac{\lambda^2 d^2(t_1, t_2)}{2} \tag{6.30}$$

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

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

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

$$\mathsf{E}\left[\sup_{t_1,\, t_2\in\mathbb{T}} \bigl|X_{t_1} - X_{t_2}\bigr|\right] \le 24 \int_0^{\delta/2} \sqrt{H(\epsilon,\, \mathbb{T},\, d)}\, \mathrm{d}\epsilon \tag{6.31}$$

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

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

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, $\#\{j \in \{1, 2, \dots, N_j;\, d(t_j,\, t) < \delta_j\} \ge 2$ 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

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

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

$$\mathsf{E}\left[\sup_{t\in\mathbb{T}} \bigl|X_t - X_{t_0}\bigr|\right] \le \sum_{j=0}^{\infty} \mathsf{E}\left[\sup_{t\in\mathbb{T}} \bigl|X_{\Pi_{j+1}(t)} - X_{\Pi_j(t)}\bigr|\right] \tag{6.32}$$

である．また

$$\begin{aligned}\#\bigl\{\bigl(\Pi_j(t),\, \Pi_{j+1}(t)\bigr);\, t \in \mathbb{T}\bigr\} &\le N(\delta_j,\, \mathbb{T},\, d) \times N(\delta_{j+1},\, \mathbb{T},\, d)\\ &\le \{N(\delta_{j+1},\, \mathbb{T},\, d)\}^2\\ &= \exp\Bigl(\log\bigl(\{N(\delta_{j+1},\, \mathbb{T},\, d)\}^2\bigr)\Bigr)\\ &= \exp\Bigl(2H(\delta_{j+1},\, \mathbb{T},\, d)\Bigr) \end{aligned} \tag{6.33}$$

となることがわかる. 三角不等式から

$$d\big(\Pi_j(t),\,\Pi_{j+1}(t)\big) \le d\big(\Pi_j(t),\,t\big) + d\big(t,\,\Pi_{j+1}(t)\big) \le \delta_j + \delta_{j+1} \le 3\delta_{j+1} \tag{6.34}$$

となる. (6.30) と (6.34) から

$$\begin{aligned}\log\Big(\mathsf{E}\big[\exp(\lambda(X_{\Pi_{j+1}(t)} - X_{\Pi_j(t)}))\big]\Big) &\le \frac{\lambda^2 d^2\big(\Pi_{j+1}(t),\,\Pi_j(t)\big)}{2} \\ &\le \frac{9\lambda^2\delta_{j+1}^2}{2}\end{aligned}$$

であるので

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

となることがわかる. よって, 補題 6.26 と (6.33) から

$$\begin{aligned}\mathsf{E}\Big[\sup_{t\in\mathbb{T}}\big|X_{\Pi_{j+1}(t)} - X_{\Pi_j(t)}\big|\Big] &\le \sqrt{2\times 9\delta_{j+1}^2\log\big(\#\big\{\big(\Pi_j(t),\,\Pi_{j+1}(t)\big);\,t\in\mathbb{T}\big\}\big)} \\ &\le \sqrt{18\delta_{j+1}^2 \times 2H(\delta_{j+1},\,\mathbb{T},\,d)} \\ &= 6\delta_{j+1}\sqrt{H(\delta_{j+1},\,\mathbb{T},\,d)} \end{aligned} \tag{6.35}$$

を得る. (6.32) の右辺に (6.35) を代入すると

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

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. $X_1,\,X_2,\,\ldots\,X_n$ を i.i.d. 確率変数列とし, $\epsilon_1,\,\epsilon_2,\,\ldots,\,\epsilon_n$ 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

$$\mathsf{E}\Big[\sup_{g\in\mathcal{G}}\Big|\frac{1}{n}\sum_{j=1}^{n}\epsilon_j g(X_j)\Big|\,\Big|\,X_1,\,X_2,\,\ldots,\,X_n\Big] \le \sqrt{2\pi}\,\frac{\sup_{g\in\mathcal{G}}\|g\|_{L_2(\widehat{\mathsf{P}}_2)}}{\sqrt{n}} + 12\int_0^{D_n}\frac{H(\epsilon,\,\mathcal{G},\,\|\cdot\|_{L_2(\widehat{\mathsf{P}}_n)})}{n}\,\mathrm{d}\epsilon$$

が成立する. ただし

$$D_n := \sup_{g\in\mathcal{G}}\|g\|_{L_\infty(\widehat{\mathsf{P}}_n)};\qquad \|g\|_{L_\infty(\widehat{\mathsf{P}}_n)} := \max_{j=1,\,2,\,\ldots,\,n}|g(X_j)|$$

$$\|g\|_{L_2(\widehat{\mathsf{P}}_n)} := \frac{1}{n}\sum_{j=1}^{n}g^2(X_j)$$

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 $\big\{\sum_{j=1}^{n}\epsilon_j g(X_j)\big\}_{g\in\mathcal{G}}$ 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. $j=1,\,2,\ldots,\,n$ に対して

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

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

$$\mathsf{E}\big[Y_j\big|\,\boldsymbol{X}\big]=0,\qquad Y_j\in\Bigg[\underbrace{-\frac{\big|g_1(X_j)-g_2(X_j)\big|}{n}}_{a},\,\underbrace{\frac{\big|g_1(X_j)-g_2(X_j)\big|}{n}}_{=b}\Bigg]$$

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

$$\mathsf{E}\Big[\exp\big(\lambda Y_j\big)\Big|\,\boldsymbol{X}\Big] \le \exp\left(\frac{\lambda^2\big\{g_1(X_j)-g_2(X_j)\big\}^2}{2n^2}\right)$$

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

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

となる．よって, $d(g_1, g_2) = \|g_1 - g_2\|_{L_2(\widehat{\mathsf{P}}_n)}/\sqrt{n}$ 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

$$
\begin{aligned}
&\mathsf{E}\biggl[\sup_{g\in\mathcal{G}}\Bigl|n^{-1}\sum_{j=1}^{n}\epsilon_j\bigl\{g(X_j) - g_0(X_j)\bigr\}\Bigr|\,\bigg|\,\boldsymbol{X}\biggr] \\
&\quad\le 12\int_0^{\delta_n/2}\sqrt{H\bigl(\epsilon, \mathcal{G}, n^{-1/2}\|\cdot\|_{L_2(\widehat{\mathsf{P}}_n)}\bigr)}\,\mathrm{d}\epsilon \\
&\quad= 12\int_0^{\delta_n/2}\sqrt{H\bigl(\sqrt{n}\epsilon, \mathcal{G}, \|\cdot\|_{L_2(\widehat{\mathsf{P}}_n)}\bigr)}\,\mathrm{d}\epsilon \\
&\quad= 12\int_0^{\sqrt{n}\delta_n/2}\sqrt{\frac{H\bigl(\epsilon', \mathcal{G}, \|\cdot\|_{L_2(\widehat{\mathsf{P}}_n)}\bigr)}{n}}\,\mathrm{d}\epsilon' \\
&\qquad\quad(\epsilon' = \sqrt{n}\epsilon \text{ と変換}) \\
&\quad= 12\int_0^{D_n}\sqrt{\frac{H\bigl(\epsilon, \mathcal{G}, \|\cdot\|_{L_2(\widehat{\mathsf{P}}_n)}\bigr)}{n}}\,\mathrm{d}\epsilon
\end{aligned}
\tag{6.36}
$$

を得る．ただし, $\delta_n = n^{-1/2}\sup_{g\in\mathcal{G}}\|g - g_0\|_{L_2(\widehat{\mathsf{P}})}$ と $D_n := \sup_{g\in\mathcal{G}}\|g - g_0\|_{L_2(\widehat{\mathsf{P}})}$ 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

$$
\begin{aligned}
&\mathsf{E}\biggl[\Bigl|n^{-1}\sum_{j=1}^{n}\epsilon_j g_0(X_j)\Bigr|\,\bigg|\,\boldsymbol{X}\biggr] \\
&\quad= \int_0^{\infty}\mathsf{Pr}\biggl(\Bigl|n^{-1}\sum_{j=1}^{n}\epsilon_j g_0(X_j)\Bigr| > t\,\bigg|\,\boldsymbol{X}\biggr)\,\mathrm{d}t \\
&\qquad\quad(\because \text{ 命題 } 1.31) \\
&\quad\le \int_0^{\infty}2\exp\biggl(-\frac{2n^2t^2}{4\sum_{j=1}^{n}\{g_0(X_j)\}^2}\biggr)\,\mathrm{d}t \\
&\qquad\quad(\because \text{ 命題 } 5.14) \\
&\quad= \int_{-\infty}^{\infty}\exp\biggl(-\frac{n^2t^2}{2\sum_{j=1}^{n}\{g_0(X_j)\}^2}\biggr)\,\mathrm{d}t \\
&\quad= \int_{-\infty}^{\infty}\exp\biggl(-\frac{t^2}{2\{\|g_0\|_{L_2(\widehat{\mathsf{P}})}/n\}^2}\biggr)\,\mathrm{d}t \\
&\quad= \sqrt{2\pi}\,\frac{\|g_0\|_{L_2(\widehat{\mathsf{P}}_n)}}{4\sqrt{n}} \\
&\quad\le \sqrt{2\pi}\,\frac{\sup_{g\in\mathcal{G}}\|g\|_{L_2(\widehat{\mathsf{P}})}}{\sqrt{n}}
\end{aligned}
\tag{6.37}
$$

を得る. (6.36) と (6.37) から

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

を得る. よって, 補題は示された. □

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

$$
\frac{1}{n}H(\epsilon,\mathcal{G},\|\cdot\|_{L_2(\widehat{\mathsf{P}}_n)})\xrightarrow{\mathrm{P}}0\qquad(n\to\infty)\tag{6.23}
$$

ならば, $\mathcal{G}$ は P-GC 族である.

Proof.

$$
\sup_{g\in\mathcal{G}}\bigl|\widehat{\mathsf{P}}_n g-\mathsf{P}g\bigr|\xrightarrow{\mathrm{P}}0\qquad(n\to\infty)
$$

を示すために

$$
\mathsf{E}\left[\sup_{g\in\mathcal{G}}\bigl|\widehat{\mathsf{P}}_n g-\mathsf{P}g\bigr|\right]\xrightarrow{n\to\infty}0
$$

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

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

となる. 命題 6.7(1)(2) から

$$
H\bigl(\epsilon,\mathcal{G},\|\cdot\|_{L_2(\widehat{\mathsf{P}}_n)}\bigr)\le H_B\bigl(\epsilon,\mathcal{G},\|\cdot\|_{L_2(\widehat{\mathsf{P}}_n)}\bigr)\le H\left(\frac{\epsilon}{2},\mathcal{G}\right)\le\left(\frac{2K}{\epsilon}\right)^n
$$

となる. すると

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

なので, 命題 2.6(1) から $\left\{\sqrt{\frac{H\left(\epsilon,\mathcal{G},\|\cdot\|_{L_2(\widehat{\mathsf{P}}_n)}\right)}{n}}\right\}_{n=1}^{\infty}$ 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

$$
\int_0^K \sqrt{\frac{H\bigl(\epsilon,\mathcal{G},\|\cdot\|_{L_2(\widehat{\mathsf{P}}_n)}\bigr)}{n}}\,\mathrm{d}\epsilon \overset{n\to\infty}{\longrightarrow} 0
$$

がわかる. 以上から

$$
\begin{aligned}
\mathsf{E}\left[\sup_{g\in\mathcal{G}}\bigl|\widehat{\mathsf{P}}_n g - \mathsf{P}g\bigr|\right] &\le 2\sqrt{2\pi}\frac{K}{\sqrt{n}} + 24\int_0^K \sqrt{\frac{H\bigl(\epsilon,\mathcal{G},\|\cdot\|_{L_2(\widehat{\mathsf{P}}_n)}\bigr)}{n}}\,\mathrm{d}\epsilon \\
&\overset{n\to\infty}{\longrightarrow} 0
\end{aligned}
$$

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. □

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. $\sup_{g\in\mathcal{G}}|g(x)| \le K$ を仮定して, 定理 6.20 を証明: $\delta > 0$, $N = N(\delta, \mathcal{G}, \|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)})$ 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

$$
\widehat{\mathsf{P}}_n|g_j - g| := \frac{1}{n}\sum_{k=1}^n \bigl|g_j(X_k) - g(x)\bigr| < \delta \tag{6.38}
$$

とできる．このことから

$$\begin{aligned}
\left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k g(X_k)\right| &\le \left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k g_j(X_k)\right| + \left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k\bigl\{g(X_k)-g_j(X_k)\bigr\}\right| \\
&\le \left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k g_j(X_k)\right| + \underbrace{\frac{1}{n}\sum_{k=1}^{n}\bigl|g(X_k)-g_j(X_k)\bigr|}_{<\delta \quad \because (6.38)} \\
&\le \left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k g_j(X_k)\right| + \delta
\end{aligned}$$

である．したがって

$$\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k g(X_k)\right| \le \max_{j=1,\,2,\,\dots,\,N}\left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k g_j(X_k)\right| + \delta \tag{6.39}$$

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

$$\mathsf{Pr}\left(\max_{j=1,\,2,\,\dots,\,N}\left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k g_j(X_k)\right| > K\sqrt{\frac{2(t+\log N)}{n}}\,\middle|\,\boldsymbol{X}\right) \le 2e^{-t} \tag{6.40}$$

となる．(6.39) と (6.40) を合わせると

$$\begin{aligned}
&\mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k g(X_k)\right| > \delta + K\sqrt{\frac{2(t+\log N)}{n}}\,\middle|\,\boldsymbol{X}\right) \\
&\qquad \le \mathsf{Pr}\left(\max_{j=1,\,2,\,\dots,\,N}\left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k g_j(X_k)\right| > K\sqrt{\frac{2(t+\log N)}{n}}\,\middle|\,\boldsymbol{X}\right) \le 2e^{-t}
\end{aligned} \tag{6.41}$$

を得る．条件付き期待値の性質 (towering property) を (6.41) に適用すると

$$\begin{aligned}
&\mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k g(X_k)\right| > 2\delta + K\sqrt{\frac{2t}{n}}\right) \\
&\qquad \le 2e^{-t} + \mathsf{Pr}\left(K\sqrt{\frac{2H_1(\delta,\,\mathcal{G},\,\|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)})}{n}} > \delta\right)
\end{aligned} \tag{6.42}$$

となる．これは以下の議論からわかる．

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

とおくと

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

となる.

$$\begin{aligned}
&\mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{k=1}^n \epsilon_k g(X_k)\right| > 2\delta + K\sqrt{\frac{2t}{n}}\right)\\
&\quad\le \mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{k=1}^n \epsilon_k g(X_k)\right| > 2\delta + K\sqrt{\frac{2t}{n}} \text{ かつ } A^c\right)\\
&\qquad + \mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{k=1}^n \epsilon_k g(X_k)\right| > 2\delta + K\sqrt{\frac{2t}{n}} \text{ かつ } A\right)\\
&\quad\le \mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{k=1}^n \epsilon_k g(X_k)\right| > \delta + K\sqrt{\frac{2\log N}{n}} + K\sqrt{\frac{2t}{n}}\right)\\
&\qquad + \mathsf{Pr}\left(K\sqrt{\frac{2H_1(\delta,\mathcal{G},\|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)})}{n}} > \delta\right)\\
&\qquad\left(\because A^c \Rightarrow K\sqrt{\frac{2\log N}{n}} \le \delta\right.\\
&\qquad\qquad\left.\Rightarrow 2\delta + K\sqrt{\frac{2t}{n}} \ge \delta + K\sqrt{\frac{2\log N}{n}} + K\sqrt{\frac{2t}{n}}\right)\\
&\quad\le \mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{k=1}^n \epsilon_k g(X_k)\right| > \delta + K\sqrt{\frac{2(t+\log N)}{n}}\right)\\
&\qquad + \mathsf{Pr}\left(K\sqrt{\frac{2H_1(\delta,\mathcal{G},\|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)})}{n}} > \delta\right)\\
&\qquad\qquad\left(\because \sqrt{a}+\sqrt{b} \ge \sqrt{a+b}\,(a,\,b\ge 0)\right)\\
&\quad= \mathsf{E}\left[\mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{k=1}^n \epsilon_k g(X_k)\right| > \delta + K\sqrt{\frac{2(t+\log N)}{n}}\right)\middle|\boldsymbol{X}\right]\\
&\qquad + \mathsf{Pr}\left(K\sqrt{\frac{2H_1(\delta,\mathcal{G},\|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)})}{n}} > \delta\right)\\
&\quad\le 2e^{-t} + \mathsf{Pr}\left(K\sqrt{\frac{2H_1(\delta,\mathcal{G},\|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)})}{n}} > \delta\right) \qquad (\because (6.42))
\end{aligned}$$

からわかる. よって, (6.42) が確認できた.

次に, 系 6.24 を (6.42) 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.

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

に注意して, Chebyshev の不等式 (系 1.33) を用いると

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

より

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

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 (6.42) を用いれば

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

を得る. $\epsilon>0$ に対して

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

になるように t と n を取り, さらに

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

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

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

となる. したがって

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

がわかる.

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

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

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

$$E_1\le\sup_{g\in\mathcal{G}_K}\left|\frac{1}{n}\sum_{j=1}^n g(X_j)-\int g\,\mathrm{dP}\right|\stackrel{\mathrm{P}}{\to}0\qquad(n\to\infty)$$

となる. 評価できる. 一方,

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

である. 任意の K に対して, 大数の法則より

$$\frac{1}{n}\sum_{j=1}^n G(X_j)\mathbb{1}\{G(X_j)>K\}-\int_{\mathbb{R}} G\mathbb{1}\{G>K\}\,\mathrm{dP}\stackrel{\mathrm{P}}{\to}0\qquad(n\to\infty)$$

となる. また, $G\in L_1(\mathrm{P})$ なので

$$\lim_{K\to\infty}\int_{\mathbb{R}} G\mathbb{1}\{G>K\}\,\mathrm{dP}=0$$

である. 以上から

$$E_2\stackrel{\mathrm{P}}{\to}0\qquad(n\to\infty)$$

がわかる. よって

$$\|\widehat{\mathsf{P}}_n - \mathsf{P}\|_{\mathcal{G}} \ \le \ \|\widehat{\mathsf{P}}_n - \mathsf{P}\|_{\mathcal{G}_K} + \sup_{g\in\mathcal{G}} |(\widehat{\mathsf{P}}_n - \mathsf{P})g\mathbb{1}\{G > K\}| \xrightarrow{\mathrm{P}} 0 \qquad (n\to\infty)$$

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, $h_1 := g_1\mathbb{1}\{G \le K\}$ と $h_2 := g_2\mathbb{1}\{G \le K\}$ 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

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

となる. 命題 6.7(3) に注意すると

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

がわかる. よって

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

ならば

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

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: $G \in L_1(\mathsf{P})$ なので, 任意の $\delta > 0$ に対して, ある $K_0 > 0$ が存在して

$$\int_{G>K_0} G \,\mathrm{d}\mathsf{P} \le \delta$$

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

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

とできる．以上のことから

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

がわかる．よって, 定理 6.20 は証明できた． □

例 6.30. $\mathbb{X}=\mathbb{R}$ とし,

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

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

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

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

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

とおく．ただし, $a\ge 0$ に対して, $\lceil a\rceil := \min\{b\in\mathbb{N};\, b\ge a\}$ である．すると $\tilde{g}$ は多くとも $m\le 1+1/\delta$ の飛躍点（$X_1, X_2, \ldots, X_n$ 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

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

となる． これより

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

となるので

$$\frac{1}{n}H(\delta, \mathcal{G}, \|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)}) \xrightarrow{\mathrm{P}} 0 \qquad (n \to \infty)$$

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

$$\widehat{\mathsf{F}}_n(x) = \int \mathbb{1}_{(-\infty, x]}\,\mathrm{d}\widehat{\mathsf{P}}_n, \qquad \mathsf{F}_n(x) = \int \mathbb{1}_{(-\infty, x]}\,\mathrm{d}\mathsf{P}$$

となることに注意する. すると

$$\begin{aligned}\sup_{x\in\mathbb{R}}\bigl|\widehat{\mathsf{F}}_n(x) - \mathsf{F}\bigr| &= \sup_{g\in\mathcal{F}}\left|\int g\,\mathrm{d}\bigl(\widehat{\mathsf{P}}_n - \mathsf{P}\bigr)\right| \\ &= \sup_{g\in-\mathcal{F}+1}\left|\int g\,\mathrm{d}\bigl(\widehat{\mathsf{P}}_n - \mathsf{P}\bigr)\right| \\ &\leq \sup_{g\in\mathcal{G}}\left|\int g\,\mathrm{d}\bigl(\widehat{\mathsf{P}}_n - \mathsf{P}\bigr)\right| \xrightarrow{\mathrm{P}} 0 \qquad (n\to\infty)\end{aligned}$$

となる. したがって, Grevenko-Cantelli の定理

$$\sup_{x\in\mathbb{R}}\bigl|\widehat{\mathsf{F}}(x) - \mathsf{F}\bigr| \xrightarrow{\mathrm{P}} 0 \qquad (n\to\infty)$$

がわかる.

6.8 VC 集合族

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

定義 6.31.

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

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

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

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

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

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 $\{\xi_1, \xi_2, \ldots, \xi_n\} \subset \mathbb{R}$ に対して,

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

定理 6.33. 次の 2 つは同値である:
(1)

$$\frac{1}{n} \log \Delta^{\mathcal{D}}(X_1, X_2, \ldots, X_n) \xrightarrow{\mathrm{P}} 0 \qquad (n \to \infty).$$

(2)

$$\sup_{D \in \mathcal{D}} |\widehat{\mathsf{P}}_n(D) - \mathsf{P}(D)| \xrightarrow{\mathrm{P}} 0 \qquad (n \to \infty).$$

Proof. (1) ⇒ (2) の証明:

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

とおき, 定理 6.20 を用いる.
　まず, 任意の $g \in \mathcal{G}$ に対して

$$|g(x)| \le 1 \qquad (\forall x \in \mathbb{X})$$

であることに注意する. $\delta > 0$ に対して, $N(\delta, \mathcal{G}, L_\infty(\widehat{\mathsf{P}}_n))$ を疑ノルム

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

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

$$N(\delta, \mathcal{G}, \|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)}) \le N(\delta, \mathcal{G}, \|\cdot\|_{L_\infty(\widehat{\mathsf{P}}_n)})$$

であることが

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

からすぐにわかる. しかし, $0 < \delta < 1$ に対して

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

である.
(2) ⇒ (1) の証明: 省略. □

定義 6.34. $n \in \mathbb{N}$ に対して,

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

とおく. $\mathcal{D}$ がバプニィック・チェルヴォネンキス (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,

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

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,

$$\begin{aligned}\frac{1}{n}H(\delta, \mathcal{G}, \|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)}) &\leq \frac{1}{n}\log \Delta^{\mathcal{D}}(X_1, X_2, \ldots, X_n) \leq \log \mathrm{VC}^{\mathcal{D}}(n) \\ &\leq \frac{1}{n}\log\{C\, n^V\} = V\frac{\log n}{n} + \frac{C}{n} \overset{n\to\infty}{\longrightarrow} 0\end{aligned}$$

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

$$(-\infty, \boldsymbol{r}] = (-\infty, r_1] \times (-\infty, r_2] \times \cdots \times (-\infty, r_p]$$

とした.
(3) $\mathbb{X} = \mathbb{R}^p$ とする.

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

は VC 族である. なぜならば

$$\mathrm{VC}^{\mathcal{D}}(n) \leq 2^p \begin{pmatrix} n \\ p \end{pmatrix}$$

よりわかる. ただし, $\begin{pmatrix} n \\ p \end{pmatrix}$ 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

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

Proof. 証明は略. □

定義 6.38. 集合族 $\mathcal{D}$ の VC 次元を

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

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

補題 6.39. 次の 2 つは同値である:
(1) $\mathcal{D}$ は VC 族.
(2) $\mathrm{VC}^{\mathcal{D}} < \infty$.

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

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

□

6.9 VC 関数族

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

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

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 $\{\mathrm{subgraph}(g);\ g \in \mathcal{G}\}$ が VC 族のときをいう．

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

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

$$\mathcal{G} := \{\theta_1\varphi_1 + \cdots + \theta_p\varphi_k;\ \theta_1,\ \ldots,\ \theta_p \in \mathbb{R}\}$$

は VC 族となる. 実際,

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

としたとき,

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

である. Pollard (1984) を参照.

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

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

補題 6.44. S は距離空間 $(\mathbb{M}, d)$ の空でない部分集合とする. 任意の $\delta > 0$ に対して以下が成り立つ.

(i)

$$N(\delta, S\, d) \leq D(\delta, S, d).$$

(ii)

$$D(\delta, S, d) \leq N\left(\frac{\delta}{2}, S, d\right).$$

Proof. (i) は明らか. (ii) を示すために

$$\{s_1, s_2, \ldots, s_N\} \subset S, \qquad N = D(\delta, S, d)$$

とする. 明らかに

$$N\left(\frac{\delta}{2}, S, d\right) \geq N\left(\frac{\delta}{2}, \{s_1, s_2, \ldots, s_N\}, d\right)$$

である. しかし

$$N\left(\frac{\delta}{2}, \{s_1, s_2, \ldots, s_N\}, d\right) = N$$

となる. □

定理 6.45. 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 [13], $N(\delta, \mathcal{G}, \|\cdot\|_{L_(\mathsf{Q})})\,(\delta > 0)$ を $L_1(\mathsf{Q})$ 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 $\mathrm{VC}^{\mathcal{G}} =: V$ が $V < \infty$ のとき, V に依存する定数 $A > 0$ が存在して

$$N(\delta \mathsf{Q}G, \mathcal{G}, \|\cdot\|_{L_1(\mathsf{Q})}) \leq \max\{A\delta^{-2V}, e^{\delta/4}\} \tag{6.43}$$

となる.

[13] $G(x) = \sup_{g \in \mathcal{G}} |g(x)|$ である.

Proof. 一般性を失うことなく, $G > 0$ で

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

としてよい. $\mathbb{X}$ 値確率要素 S を取り, その分布を

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

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. $\{g_1,\, g_2,\, \ldots,\, g_N\}$ は $g_j : \mathbb{X} \longrightarrow \mathbb{R}\,(j = 1,\, 2,\, \ldots,\, N)$ で

$$Q(|g_j - g_k|) > \delta \qquad (j \neq k,\, j,\, k = 1,\, 2,\, \ldots,\, N)$$

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

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

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

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

となる. $\{(S_i,\, T_i)\}_{i=1}^n$ を $(S,\, T)$ の i.i.d. 複製とする. $(S_1,\, T_1),\, (S_2,\, T_2),\, \ldots,\, (S_n,\, T_n)$ 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

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

である. ある $j \neq k$ に対して, $(S_1,\, T_1),\, (S_2,\, T_2),\, \ldots,\, (S_n,\, T_n)$ 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. ここで

$$n \geq \frac{4\log N}{\delta} \tag{6.44}$$

をみたす最小の n を取ると

$$\binom{N}{2}\left(1 - \frac{\delta}{2}\right)^n \leq \frac{1}{2} < 1 \tag{6.45}$$

とできる．また, n は (6.44) をみたす最小の n なので

$$n \leq \frac{4\log N}{\delta} + 1 \tag{6.46}$$

となっていることに注意せよ． そのような n に対して, 任意の $j \neq k\,(j,\,k = 1,\,2,\,\ldots,\,N)$ に対して, $(S_i,\,T_i)\,(i = 1,\,2,\,\ldots,\,n)$ は g_j と g_k のグラフの間に落ちる確率は正である．なぜならば, (6.45) から落ちない確率が 1 より小さいことからわかる．すなわち, n 個の $\{(S_j,\,T_j\}_{j=1}^n$ は完全に分離されないことがわかる．よって

$$\begin{aligned}D(\delta,\,\mathcal{G},\,\|\cdot\|_{L_1(\mathsf{Q})}) = N &\leq \sup\Big\{\Delta^{\mathcal{D}}\Big((S_1,\,T_1),\,(S_2,\,T_2),\,\ldots,\,(S_n,\,T_n)\Big)\\ &\qquad ;\,(S_i,\,T_i)\in\mathcal{X}\times\mathbb{R}\,(i=1,\,2,\,\ldots,\,n)\Big\}\\ &\leq Cn^V \end{aligned}\tag{6.47}$$

となる．ただし, C は定義 6.34 の定数であり

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

である．

$N \geq \exp\left\{\dfrac{\delta}{4}\right\}$ の場合を考える[14]．すると, (6.44) は

$$n \geq \frac{4\log N}{\delta} \geq 4\times\frac{1}{4}$$

となるので, $\{(S_j,\,T_j\}_{j=1}^n$ は完全に分離されない確率がすべての n に対して正になる．したがって

$$\begin{aligned}N &\leq Cn^V \qquad (\because (6.47))\\ &\leq C\left(\frac{4\log N}{\delta}+1\right)^V \qquad (\because (6.46))\\ &\leq C\left(\frac{8\log N}{\delta}\right)^V \quad \left(N\geq e^{\delta/4} \Rightarrow \frac{4\log N}{\delta}\geq 1 \because\right)\\ &= C\left(\frac{16V\log N^{1/(2V)}}{\delta}\right)^V\\ &\leq C\left(\frac{16V}{\delta}\right)^V\sqrt{N}\end{aligned}$$

となる．この不等式の両辺を $\sqrt{N}$ で割ると

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

[14] $N \geq e^{\delta/4}$ でないときは, $N < e^{\delta/4}$ となるので, (6.43 のもう一方の上限が出てくる．

がわかる. よって

$$\sqrt{N} \le C(16V)^V \delta^{-V}$$

となる. このことから

$$N \ge e^{\delta/4} \Rightarrow N \le C^2(16V)^{2V}\delta^{-2V} =: A\delta^{-2V}$$

がわかる. すなわち

$$N \ge e^{\delta/4} \Rightarrow N \le A\delta^{-2V}$$

がわかる. これらを合わせると

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

を得る. $\mathsf{Q}G = 1$ に規準化していたので

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

がわかる. □

系 6.46. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P}$ とする. ただし, P は $\mathbb{X}$ 上の確率測度である. $\mathcal{G}$ を関数族とし, G を関数族 $\mathcal{G}$ の封筒関数とする. $\mathsf{P}G < \infty$ かつ関数族 $\mathcal{G}$ は VC 族ならば, $\mathcal{G}$ は P-GC 族である.

Proof. まず, 定理 6.20 を思い出す. 関数族 $\mathcal{G}$ が P-GC 族であるための十分条件は, $\mathsf{P}G \in L_1(\mathsf{P})$ かつ, 任意の $\widetilde{\delta} > 0$ に対して

$$\frac{1}{n} H\big(\widetilde{\delta}, \mathcal{G}, \|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)}\big) \xrightarrow{\mathrm{P}} 0 \qquad (n \to \infty) \tag{6.48}$$

である. すなわち, (6.48) を確認すればよい. そのために, 定理 6.45 を用いる. $\mathsf{Q}G < \infty$ なる任意の確率測度 Q と $\delta > 0$ に対して

$$N(\delta \mathsf{Q}G, \mathcal{G}, \|\cdot\|_{L_1(\mathsf{Q})}) \le \max\big\{Ae^{-2V}, e^{\delta/4}\big\} \tag{6.49}$$

であった. ここで, $\mathsf{Q} = \widehat{\mathsf{P}}_n$ とおくと, (6.49) から

$$\begin{aligned}\frac{1}{n}\log N\big(\delta\widehat{\mathsf{P}}_n G, \mathcal{G}, \|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)}\big) &\le \frac{1}{n}\log\Big(\max\big\{Ae^{-2V}, e^{\delta/4}\big\}\Big) \\ &\overset{n\to\infty}{\longrightarrow} 0\end{aligned}$$

がわかる. よって

$$\frac{1}{n} H\big(\widetilde{\delta}, \mathcal{G} \|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)}\big) \xrightarrow{\mathrm{P}} 0 \qquad (n \to \infty)$$

が示せた. さらに

$$|\widehat{\mathsf{P}}_n G - \mathsf{P}G| \xrightarrow{\mathrm{P}} 0 \qquad (n \to \infty)$$

に注意すれば, 上の式の左辺の $\delta\widehat{\mathsf{P}}_n G$ 項の $2\delta\mathsf{P}G$ に入れ替えたもので

$$\frac{1}{n}\log N(2\delta\mathsf{P}G, \mathcal{G}, \|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)}) \xrightarrow{\mathrm{P}} 0 \qquad (n \to \infty)$$

となる[15]. 最後に, $\widetilde{\delta} = 2\delta\mathsf{P}G$ とおくと (6.48) がわかる. □

6.10 混合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, 混合モデルの最尤推定量の一致性を示そう.

混合モデルは次のように表現できるものである. Y を実数値確率変数とし, 分布関数 F_0 を持つ. $Y = y\,(y \in \mathbb{R})$ を与えたとき, 実数値確率変数 X は p.d.f. $\mathsf{k}(x|\,y)$ を持つとする. このとき, X は混合 p.d.f.

$$\mathsf{p}_{\mathsf{F}_0} := \int \mathsf{k}(x|\,y)\,\mathrm{d}\mathsf{F}_0(y) \qquad (x \in \mathbb{R}) \tag{6.50}$$

を持つことがわかる.

[15] $|(\widehat{\mathsf{P}}_n - \mathsf{P})G| \le \frac{\mathsf{P}G}{2}$ のとき

$$2\mathsf{P}G - \widehat{\mathsf{P}}_n G = \mathsf{P}G + (\mathsf{P} - \widehat{\mathsf{P}}_n)G \ge \mathsf{P}G - |(\widehat{\mathsf{P}}_n - \mathsf{P})G| > \mathsf{P}G/2 > 0$$

となるので

$$N(2\delta\mathsf{P}G, \mathcal{G}, \|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)}) \le N(\delta\widehat{\mathsf{P}}_n G, \mathcal{G}, \|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)})$$

が成り立つことに注意する. すなわち

$$|(\widehat{\mathsf{P}}_n - \mathsf{P})G| \le \frac{\mathsf{P}G}{2} \Rightarrow N(2\delta\mathsf{P}G, \mathcal{G}, \|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)}) \le N(\delta\widehat{\mathsf{P}}_n G, \mathcal{G}, \|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)})$$

である. このことから, 任意の $\epsilon > 0$ に対して

$$\begin{aligned}
&\mathsf{Pr}\left(\frac{1}{n}\log N(2\delta\mathsf{P}G, \mathcal{G}, \|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)}) > \epsilon\right) \\
&\quad \le \mathsf{Pr}\left(\left\{\frac{1}{n}\log N(2\delta\mathsf{P}G, \mathcal{G}, \|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)}) > \epsilon\right\} \bigcap \left\{|(\widehat{\mathsf{P}}_n - \mathsf{P})G| \le \frac{\mathsf{P}G}{2}\right\}\right) \\
&\qquad + \mathsf{Pr}\left(|(\widehat{\mathsf{P}}_n - \mathsf{P})G| > \frac{\mathsf{P}G}{2}\right) \\
&\quad \le \mathsf{Pr}\left(\frac{1}{n}\log N(\delta\widehat{\mathsf{P}}_n G, \mathcal{G}, \|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)}) > \epsilon\right) + \mathsf{Pr}\left(|(\widehat{\mathsf{P}}_n - \mathsf{P})G| > \frac{\mathsf{P}G}{2}\right) \\
&\quad \overset{n\to\infty}{\longrightarrow} 0
\end{aligned}$$

がわかる.

ここでは, k は既知とし, 未知の F_0 の最尤推定量の一致性を示す.

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エントロピーは有界であることを示す.

補題 6.47. $p \geq 1$ とし, P を距離空間 $\mathbb{X}$ 上の確率測度とし, (Θ, d) をコンパクトな距離空間とする. 関数族

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

は以下をみたすとする.

- すべての $x \in \mathbb{X}$ に対して, 写像

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

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

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

このとき, 任意の $0 < \delta \leq 1$ に対して

$$H_B\big(\delta, \mathcal{G}, \|\cdot\|_{L_p(\mathsf{P})}\big) < \infty$$

となる.

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

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

と定める. すべての $x \in \mathbb{X}$ に対して, 関数

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

は連続なので

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

となる.

$$|g_\theta(x) - g_{\widetilde{\theta}}(x)| \leq 2G(x)$$

かつ $G \in L_p(\mathsf{P})$ なので, 優収束定理 (定理 1.40) から

$$\lim_{\rho\to 0} \mathsf{P}\omega^p_{\theta\,\rho} = 0$$

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

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

とできる. つぎに

$$B_\theta := \bigl\{\widetilde{\theta} \in \Theta;\, d(\theta, \widetilde{\theta}) < \rho_\theta)\bigr\}$$

とおくと $\bigl\{B_\theta;\, \theta \in \Theta\bigr\}$ は Θ の開被覆となる. Θ はコンパクトなので, $\exists N \in \mathbb{N}$ と $\exists\{\theta_1, \theta_2, \ldots, \theta_N\} \subset \Theta$ が存在して

$$\bigcup_{j=1}^{N} B_{\theta_j} = \Theta$$

とできる. 各 $\theta \in B_j$ と $x \in \mathbb{X}$ に対して

$$\begin{aligned} g_\theta(x) &:\le g_{\theta_j}(x) + \omega_{\theta_j,\,\rho_{\theta_j}}(x) =: g_j^{\mathrm{R}}(x), \\ g_\theta(x) &:\ge g_{\theta_j}(x) - \omega_{\theta_j,\,\rho_{\theta_j}}(x) =: g_j^{\mathrm{L}}(x), \end{aligned}$$

とおけば

$$\mathsf{P}|g_j^{\mathrm{R}} - g_j^{\mathrm{L}}|^p \le \mathsf{P}\bigl(2\omega_{\theta_j,\,\rho_{\theta_j}}\bigr)^p \le (2\delta)^p$$

となる. よって

$$H_B\bigl(2\delta, \mathcal{G}, \|\cdot\|_{L_p(\mathsf{P})}\bigr) \le \log N < \infty$$

がわかる. □

いま, 観測 $X_1, X_2, \ldots, X_n$ は X の i.i.d. 複製とする. ただし, X は p.d.f. (6.50) 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

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

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

$$h(\mathsf{p}, \widetilde{\mathsf{p}}) = \left\{\frac{1}{2}\int_{\mathbb{R}} (\sqrt{\mathsf{p}(x)} - \sqrt{\widetilde{\mathsf{p}}(x)})^2 \,\mathrm{d}\mathsf{m}(x)\right\}^{1/2}$$

であった.

補題 6.48.

$$h(\mathsf{p}_{\widehat{\mathsf{F}}_n},\, \mathsf{p}_{\mathsf{F}_0}) \xrightarrow{\mathrm{P}} 0 \qquad (n \to \infty)$$

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

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

が成立するとき, μ_n は μ に弱収束（漠収束）するといい,

$$\mu_n \xrightarrow{w} \mu \qquad (n \to \infty)$$

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

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

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,

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

は連続で

$$\lim_{y\to\pm\infty} \mathsf{k}(x|\, y) = 0$$

であれば, 写像

$$\mathcal{F} \ni \mathsf{F} \mapsto \mathsf{p}_{\mathsf{F}}(x) = \int_{\mathbb{R}} \mathsf{k}(x|\, y)\,\mathrm{d}\mathsf{F}(y)$$

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

$$d(\widehat{\mathsf{F}}_n,\, \mathsf{F}_0) \overset{n\to\infty}{\longrightarrow} 0 \quad \Leftrightarrow \quad \widehat{\mathsf{F}}_n \xrightarrow{w} \mathsf{F}_0\,(n \to \infty)$$

をみたす $\mathcal{H}$ 上の距離関数である. なぜならば

$$y \mapsto \mathsf{k}(x|\, y)$$

は有界連続関数なので

$$\widehat{\mathsf{F}}_n \xrightarrow{w} \mathsf{F}_0\,(n \to \infty) \;\; \Rightarrow \;\; \int_{\mathbb{R}} \mathsf{k}(x|\, y)\,\mathrm{d}\widehat{\mathsf{F}}_n(y) \overset{n\to\infty}{\longrightarrow} \int_{\mathbb{R}} \mathsf{k}(x|\, y)\,\mathrm{d}\mathsf{F}_0(y) \qquad \text{a.e. } x$$

となる．したがって, 優収束定理（定理 1.40）より

$$\lim_{n\to\infty} h(\mathsf{p}_{\widehat{\mathsf{F}}_n}, \mathsf{p}_{\mathsf{F}_0}) = \left\{\frac{1}{2}\int_{\mathbb{R}} \{\sqrt{\mathsf{p}_{\widehat{\mathsf{F}}_n}} - \sqrt{\mathsf{p}_{\mathsf{F}_0}}\}^2 \,\mathrm{d}\mathsf{m}\right\}^{1/2} = 0$$

よりわかる．すると, 写像

$$\mathcal{F} \ni \mathsf{F} \mapsto \frac{2\mathsf{p}_{\mathsf{F}}}{\mathsf{p}_{\mathsf{F}} + \mathsf{p}_{\mathsf{F}_0}}$$

も d に関して連続となる．

いま, つぎの関数族を考える:

$$\mathcal{G} := \left\{\frac{2\mathsf{p}_{\mathsf{F}}}{\mathsf{p}_{\mathsf{F}} + \mathsf{p}_{\mathsf{F}_0}};\, \mathsf{F} \in \mathcal{F}\right\}.$$

この関数族の包絡関数 G は

$$G(x) \leq 2$$

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

$$\mathcal{F} \ni \mathsf{F} \mapsto \frac{2\mathsf{p}_{\mathsf{F}}}{\mathsf{p}_{\mathsf{F}} + \mathsf{p}_{\mathsf{F}_0}} \in \mathcal{G}$$

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

$$1 - \mathsf{P}\left(\frac{2\mathsf{p}_{\widehat{\mathsf{F}}_n}}{\mathsf{p}_{\widehat{\mathsf{F}}_n} + \mathsf{p}_{\mathsf{F}_0}}\right) \leq (\widehat{\mathsf{P}}_n - \mathsf{P})\left(\frac{2\mathsf{p}_{\widehat{\mathsf{F}}_n}}{\mathsf{p}_{\widehat{\mathsf{F}}_n} + \mathsf{p}_{\mathsf{F}_0}}\right). \tag{6.51}$$

これは

$$\begin{aligned} 0 &\leq \widehat{\mathsf{P}}_n\left(\log\frac{2\mathsf{p}_{\widehat{\mathsf{F}}_n}}{\mathsf{p}_{\widehat{\mathsf{F}}} + \mathsf{p}_{\mathsf{F}_0}}\right) \\ &\leq \widehat{\mathsf{P}}_n\left(\frac{2\mathsf{p}_{\widehat{\mathsf{F}}_n}}{\mathsf{p}_{\widehat{\mathsf{F}}_n} + \mathsf{p}_{\mathsf{F}_0}}\right) - 1 \\ &= (\widehat{\mathsf{P}}_n - \mathsf{P})\left(\frac{2\mathsf{p}_{\widehat{F}}}{\mathsf{p}_{\widehat{F}_n} + \mathsf{p}_{F_0}}\right) + \mathsf{P}\left(\frac{2\mathsf{p}_{\widehat{\mathsf{F}}_n}}{\mathsf{p}_{\widehat{\mathsf{F}}_n} + \mathsf{p}_{\mathsf{F}_0}}\right) - 1. \end{aligned}$$

からわかる．しかし

$$h^2(\mathsf{p}, \mathsf{p}_0) \leq 1 - \mathsf{P}\left(\frac{2\mathsf{p}}{\mathsf{p} + \mathsf{p}_0}\right) \tag{6.52}$$

である. これは

$$
\begin{aligned}
h^2(\mathsf{p}, \mathsf{p}_0) &= \frac{1}{2}\int_{\mathbb{R}} \{\sqrt{\mathsf{p}} - \sqrt{\mathsf{p}_0}\}^2 \, \mathrm{d}\mathsf{m} \\
&= \frac{1}{2}\int_{\mathbb{R}} \frac{(\mathsf{p}-\mathsf{p}_0)^2}{\{\sqrt{\mathsf{p}} + \sqrt{\mathsf{p}_0}\}^2} \, \mathrm{d}\mathsf{m} \\
&\le \frac{1}{2}\int_{\mathbb{R}} \frac{(\mathsf{p}-\mathsf{p}_0)^2}{\{\mathsf{p}+\mathsf{p}_0\}^2} \, \mathrm{d}\mathsf{m} \\
&= \frac{1}{2}\int_{\mathbb{R}} \frac{(\mathsf{p}-\mathsf{p}_0)^2}{\{\mathsf{p}+\mathsf{p}_0\}^2} \, \mathrm{d}\mathsf{m} + \frac{1}{2}\int_{\mathbb{R}} \frac{\mathsf{p}-\mathsf{p}_0}{\{\mathsf{p}+\mathsf{p}_0\}^2} \, \mathrm{d}\mathsf{m} \\
&= \frac{1}{2}\int_{\mathbb{R}} \frac{\mathsf{p}_0(\mathsf{p}-\mathsf{p}_0)}{\{\mathsf{p}+\mathsf{p}_0\}^2} \, \mathrm{d}\mathsf{m} \\
&= \frac{1}{2}\int_{\mathbb{R}} \left(1 - \frac{2\mathsf{p}}{\mathsf{p}+\mathsf{p}_0}\right)^2 \, \mathrm{d}\mathsf{m} \\
&= 1 - \mathsf{P}\left(\frac{2\mathsf{p}}{\mathsf{p}+\mathsf{p}_0}\right)
\end{aligned}
$$

からわかる. (6.51) と (6.52) より

$$
\begin{aligned}
h^2(\mathsf{p}_{\widehat{\mathsf{F}}_n}, \mathsf{p}_{\mathsf{F}_0}) &\le (\widehat{\mathsf{P}}_n - \mathsf{P})\left(\frac{2\mathsf{p}_{\widehat{\mathsf{F}}_n}}{\mathsf{p}_{\widehat{\mathsf{F}}_n} + \mathsf{p}_{\mathsf{F}_0}}\right) \\
&\le \sup_{g\in\mathcal{G}} |(\widehat{\mathsf{P}}_n - \mathsf{P})g| = \|\widehat{\mathsf{P}}_n - \mathsf{P}\|_{\mathcal{G}} \xrightarrow{\mathrm{P}} 0 \qquad (n\to\infty)
\end{aligned}
$$

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

$$
h(\mathsf{p}_{\widehat{\mathsf{F}}_n}, \mathsf{p}_{\mathsf{F}_0}) \xrightarrow{\mathrm{P}} 0 \qquad (n\to\infty)
$$

を得る. □