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 $\overline{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$ に対して, 大数の強法則と中心極限定理を適用すると, $n \to \infty$ 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 $\frac{1}{n}\sum_{j=1}^n \widehat{g}_n(X_j)$ 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

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

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

6.2 経験過程の例

6.2.1 教育と雇用

個人からなる母集団からの標本

$$\boldsymbol{X}_1 = (Y_1,\, Z_1),\ \boldsymbol{X}_2 = (Y_2,\, Z_2),\ \ldots,\ \boldsymbol{X}_n = (Y_n,\, Z_n)$$

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

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

を考えることができる. 後に説明する道具を用いると

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

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

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

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,

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

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)

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

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

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

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 $\mathsf{p}_{\widehat{\theta}_n^{\mathrm{mle}}}$ は

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

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

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

となる. したがって

$$
\begin{aligned}
\mathsf{KL}\big(\mathsf{p}_{\theta_*},\, \mathsf{p}_{\widehat{\theta}_n^{\mathrm{mle}}}\big) &\leq \left|\frac{1}{n}\sum_{j=1}^n \log\left(\frac{\mathsf{p}_{\theta_*}(X_j)}{\mathsf{p}_{\widehat{\theta}_n^{\mathrm{mle}}}(X_j)}\right) - \mathsf{KL}\big(\mathsf{p}_{\theta_*},\, \mathsf{p}_{\widehat{\theta}_n^{\mathrm{mle}}}\big)\right| \\
&\leq \sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{j=1}^n \log\left(\frac{\mathsf{p}_{\theta_*}(X_j)}{\mathsf{p}_{\theta}(X_j)}\right) - \mathsf{KL}\big(\mathsf{p}_{\theta_*},\, \mathsf{p}_{\theta}\big)\right| \qquad (6.4)
\end{aligned}
$$

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.

命題 6.2(2) と (6.4) を合わせると

$$
\begin{aligned}
h^2\big(\mathsf{p}_{\theta^*},\, \mathsf{p}_{\widehat{\theta}_n^{\mathrm{mle}}}\big) &\leq \frac{1}{2}\mathsf{KL}\big(\mathsf{p}_{\theta^*},\, \mathsf{p}_{\widehat{\theta}_n^{\mathrm{mle}}}\big) \\
&\leq \frac{1}{2}\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{j=1}^n \log\left(\frac{\mathsf{p}_{\theta_*}(X_j)}{\mathsf{p}_{\theta}(X_j)}\right) - \mathsf{KL}\big(\mathsf{p}_{\theta_*},\, \mathsf{p}_{\theta}\big)\right|
\end{aligned}
$$

[4] 中心極限定理から用意に想像できる.

となる. ここで

$$
\begin{aligned}
\mathcal{G} &= \left\{ \frac{\mathsf{p}_{\theta^*}}{\mathsf{p}_\theta};\, \theta \in \Theta \right\}, \\
\widehat{\mathsf{P}}_n(B) &= \frac{1}{n} \#\{X_j \in B;\, j = 1,\, 2, \ldots,\, n\}, \\
\mathsf{P}(B) &= \int_B \mathsf{p}_{\theta_*}\, \mathrm{dm} \qquad (B \in \mathcal{B}(\mathbb{X}))
\end{aligned}
$$

と定める [5] と

$$
h^2\bigl(\mathsf{p}_{\theta^*},\, \mathsf{p}_{\widehat{\theta}_n^{\mathrm{mle}}}\bigr) \le \frac{1}{2} \sup_{g \in \mathcal{G}} \bigl| \bigl(\widehat{\mathsf{P}}_n - \mathsf{P}\bigr) g \bigr|
$$

と書き直すことができる. 上式の左辺の収束のスピードは関数族 $\mathcal{G}$ の複雑さに依存することがわかる.

6.3 計量エントロピー, 被覆数と ϵ 網

この章では, $\bigl(\mathbb{D},\, d\bigr)$ を一般の擬距離空間とする. ただし, 擬距離離関数 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 [6] 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

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

で定める.

定義 6.4. (1) $\bigl(\mathbb{D},\, d\bigr)$ 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.

[5] さらに, $\mathsf{P}g = \int g\, \mathrm{d}\mathsf{P}$ と $\widehat{\mathsf{P}}_n g = \frac{1}{n} \sum_{j=1}^n g(X_j)$ と約束している.

[6] 任意の $x,\, y,\, z \in \mathbb{D}$ に対して, ① $d(x,\, y) \ge 0$; ② $d(x,\, y) = d(y,\, x)$; ③ $d(x,\, z) \le d(x,\, y) + d(y,\, z)$, をみたす.

(2) 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.
(3) $H(\epsilon, A, d) = \log N(\epsilon, A, d)$ を部分集合 A の ϵ **エントロピー**ということにする.
(4) 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 $H_p(\epsilon, \mathcal{G}, d)$ を $H\bigl(\epsilon, \mathcal{G}, \|\cdot\|_{L_p(\mathsf{Q})}\bigr)$ または $H_p(\epsilon, \mathcal{G}, \mathsf{Q})$ とも記すことにする. この場合には, $\mathcal{G}$ は距離空間

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

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 $H_\infty(\epsilon, \mathcal{G}, \|\cdot\|_\infty)$ [7] と記すことにする.

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, $N_B(\epsilon, \mathcal{G}, d)$ [8] 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

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

[7] $H_\infty(\epsilon, \mathcal{G})$ とも記すことがある.
[8] $N_B\bigl(\epsilon, \mathcal{G}, \|\cdot\|_{L_p(\mathsf{Q})}\bigr)$ または $N_{B,p}(\epsilon, \mathcal{G}, \mathsf{Q})$ とも記す

- 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

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

  が成り立つ.

$H_B(\epsilon, \mathcal{G}, d) = \log N_B(\epsilon, \mathcal{G}, d)$ [9] 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. 以下が成立する.
(1)

$$H(\epsilon, \mathcal{G}, \|\cdot\|_{L_p(\mathsf{Q})}) \leq H_B(\epsilon, \mathcal{G}, \|\cdot\|_{L_p(\mathsf{Q})}).$$

(2)

$$H_B(\epsilon, \mathcal{G}, \|\cdot\|_{L_p(\mathsf{Q})}) \leq H\left(\frac{\epsilon}{2}, \mathcal{G}, \|\cdot\|_\infty\right).$$

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

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

が成立するとする. このとき

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

が成り立つ.

Proof. (1) $\epsilon > 0$ と $1 \leq p < \infty$ とし, 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. $\log N := H_B(\epsilon, \mathcal{G}, \|\cdot\|_{L_p(\mathsf{Q})})$ 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

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

なので

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

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

$$H_p(\epsilon, \mathcal{G}, \|\cdot\|_{L_p(\mathsf{Q})}) \leq H_{p,B}(\epsilon, \mathcal{G}, \|\cdot\|_{L_p(\mathsf{Q})})$$

[9] $H_B(\epsilon, \mathcal{G}, \|\cdot\|_{L_p(\mathsf{Q})})$ または $H_{B,p}(\epsilon, \mathcal{G}, \mathsf{Q})$ とも記す

がわかる.
(2) の証明: $\log N := H(\epsilon/2, \mathcal{G}, \|\cdot\|_\infty)$ 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

$$H_B(\epsilon, \mathcal{G}, \|\cdot\|_{L_p(\mathsf{Q})}) \le H\left(\frac{\epsilon}{2}, \mathcal{G}, \|\cdot\|_\infty\right)$$

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, $\{g_1, g_2, \ldots, g_N\}$ は距離関数 d に関しても ϵ 被覆となる. したがって, $e^{H(\epsilon, A, d)}$ は d 関する ϵ 被覆の最小数なので

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

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 $j \neq k$ 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 $\{c_1, c_2, \ldots, c_N\}$ 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 $\{c_1, c_2, \ldots, c_N, c_0\}$ も部分集合 A の ϵ 網となる. $\#\{c_1, c_2, \ldots, c_N, c_0\} = N + 1$ となるので, $\{c_1, c_2, \ldots, c_N\}$ 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

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

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

$$B_d\left(c_j,\,\frac{\epsilon}{2}\right)\bigcap B_d\left(c_k,\,\frac{\epsilon}{2}\right)=\varnothing \tag{6.5}$$

となる. また

$$\bigcup_{j=1}^N B_d\left(c_j,\,\frac{\epsilon}{2}\right)\subset B_d\left(0_d,\,R+\frac{\epsilon}{2}\right) \tag{6.6}$$

であることがわかる. すると (6.5) に注意して (6.6) の両辺の体積を比較すると

$$N\mathsf{vol}(B_d(0_1,\,1))\left(\frac{\epsilon}{2}\right)^d\le\mathsf{vol}(B_d(0_d,\,1))\left(R+\frac{\epsilon}{2}\right)^d \tag{6.7}$$

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. (6.7) 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. (6.5) と (6.6) を確認せよ.

例 6.11. □

例 6.12.

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

とする. このとき, ある $A>0$ が存在して

$$H(\epsilon, \mathcal{G}, \|\cdot\|_\infty) \leq \frac{A}{\epsilon} \qquad (\epsilon>0) \tag{6.8}$$

となる. (6.8) 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 $\widetilde{g}\colon[0,1]\longrightarrow[0,1]$ 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

$$\sup_{x\in[0,1]}|g-\widetilde{g}|\leq 2\epsilon \tag{6.9}$$

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

$$\begin{aligned}
&|\widetilde{g}(a_k)-\widetilde{g}(a_{k-1})|\\
&\quad\leq|\widetilde{g}(a_k)-g(a_k)|+|g(a_k)-g(a_{k-1})|+|\widetilde{g}(a_{k-1})-g(a_{k-1})|\\
&\quad\leq|\widetilde{g}(a_k)-g(a_k)|+\left|\int_{a_{k-1}}^{a_k}\overset{\bullet}{g}(t)\,\mathsf{dm}(t)\right|+|\widetilde{g}(a_{k-1})-g(a_{k-1})|\\
&\quad\leq|\widetilde{g}(a_k)-g(a_k)|+\int_{a_{k-1}}^{a_k}|\overset{\bullet}{g}(t)|\,\mathsf{dm}(t)+|\widetilde{g}(a_{k-1})-g(a_{k-1})|\\
&\quad\leq|\widetilde{g}(a_k)-g(a_k)|+\int_{a_{k-1}}^{a_k}\mathsf{dm}(t)+|\widetilde{g}(a_{k-1})-g(a_{k-1})|\\
&\quad\leq|\widetilde{g}(a_k)-g(a_k)|+(a_k-a_{k-1})+|\widetilde{g}(a_{k-1})-g(a_{k-1})|\\
&\quad\leq 3\epsilon
\end{aligned} \tag{6.10}$$

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. (6.10) 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 (6.9) から, ある $A'\,(>0)$ が存在して

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

となることがわかる. よって

$$H\big(\epsilon,\mathcal{G},\|\cdot\|_\infty\big)\leq\frac{2A'}{\epsilon}=:\frac{A}{\epsilon}$$

が示せた. □

例 6.13. □

例 6.14. $L > 0$ を定数とる. 区間 $\mathcal{I} \subset \mathbb{R}$ 上の実数値関数 $g : \mathcal{I} \to \mathbb{R}$ 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 $\left(1 + \dfrac{1}{\epsilon}\right)$ 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

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

よって，$0 < \epsilon < 1$ に対して，

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

□

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

定義 6.15. P を $\mathbb{R}$ 上の確率測度とし，$X, X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim}$ 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

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

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

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

となる. したがって

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

となる. 同様に

$$
\begin{aligned}
\int g\,\mathrm{d}\bigl(\widehat{\mathsf{P}}_n-\mathsf{P}\bigr) &= \int g\,\mathrm{d}\widehat{\mathsf{P}}_n-\int g\,\mathrm{d}\mathsf{P}\\
&\ge \int g_{j_0}^L\,\mathrm{d}\widehat{\mathsf{P}}_n-\int g\,\mathrm{d}\mathsf{P}\\
&= \int g_{j_0}^L\,\mathrm{d}\bigl(\widehat{\mathsf{P}}_n-\mathsf{P}\bigr)+\int (g_{j_0}^L-g)\,\mathrm{d}\mathsf{P}\\
&\ge \int g_{j_0}^L\,\mathrm{d}\bigl(\widehat{\mathsf{P}}_n-\mathsf{P}\bigr)-\int |g_{j_0}^L-g|\,\mathrm{d}\mathsf{P}\\
&\ge \int g_{j_0}^L\,\mathrm{d}\bigl(\widehat{\mathsf{P}}_n-\mathsf{P}\bigr)-\int |g_{j_0}^L-g_{j_0}^R|\,\mathrm{d}\mathsf{P}\\
&\ge \int g_{j_0}^L\,\mathrm{d}\bigl(\widehat{\mathsf{P}}_n-\mathsf{P}\bigr)-\epsilon
\end{aligned}
\tag{6.12}
$$

となる. (6.11) と (6.12) を合わせると

$$
\begin{aligned}
-\max_{j=1,2,\dots,N}\left|\int g_j^L\,\mathrm{d}\bigl(\widehat{\mathsf{P}}_n-\mathsf{P}\bigr)\right|-\epsilon &\le \int g\,\mathrm{d}\bigl(\widehat{\mathsf{P}}_n-\mathsf{P}\bigr)\\
&\le \max_{j=1,2,\dots,N}\left|\int g_j^R\,\mathrm{d}\bigl(\widehat{\mathsf{P}}_n-\mathsf{P}\bigr)\right|+\epsilon
\end{aligned}
\tag{6.13}
$$

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

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

となる. よって, 十分大きな n に対して

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

と

$$\mathsf{Pr}\left(\max_{j=1,\,2,\,\dots,\,N}\left|\int g_j^R\,\mathrm{d}\bigl(\widehat{\mathsf{P}}_n-\mathsf{P}\bigr)\right|\le\epsilon\right)=1 \tag{6.15}$$

となる. (6.13) – (6.15) を合わせると十分大きな n に対して

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

が成立することがわかる. すなわち

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

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, $H_B\bigl(\epsilon,\,\mathcal{G},\,\|\cdot\|_{L_1(\mathsf{P})}\bigr)<\infty$ ならば, $G\in L_1(\mathsf{P})$ である.

Proof. $\forall\epsilon>0$ に対して, $H_B(\epsilon,\,\mathcal{G},\,\|\cdot\|_{L_1(\mathsf{P})})<\infty$ なので, 命題 6.7(1) から, $H(\epsilon,\,\mathcal{G},\,\|\cdot\|_{L_1(\mathsf{P})})<\infty$ である. よって, $\bigl(\mathcal{G},\,\|\cdot\|_{L_1(\mathsf{P})}\bigr)$ は全有界となる. すなわち, $N:=H_B(\epsilon,\,\mathcal{G},\,\|\cdot\|_{L_1(\mathsf{P})})$ と書いた [10] とき, ある $\{g_1,\,g_2,\,\dots,\,g_N\}\subset\mathcal{G}$ が存在して,

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

とできる. ここで, $L_1(\mathsf{P})$ 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 [11] がわかる. さ

[10] 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.

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

(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.

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

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

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

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

となる.

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

と書き直すことができる. したがって

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

がわかる. よって

$$\mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\bigl|g(X_n)-\mathsf{P}g\bigr|\geq n, \text{i.o.}\right)=0 \tag{6.16}$$

となる. したがって, 補題 1.44 の対偶から

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

がわかる. よって

$$\begin{aligned}\mathsf{E}\bigl[\sup_{g\in\mathcal{G}}\bigl|g(X_n)-\mathsf{P}g\bigr|\bigr]&=\int_0^{\infty}\mathsf{Pr}\bigl(\sup_{g\in\mathcal{G}}\bigl|g(X_n)-\mathsf{P}g\bigr|>t\bigr)\,\mathrm{d}t && (\because \text{命題 } 1.32)\\ &\leq\sum_{n=1}^{\infty}\mathsf{Pr}\bigl(\sup_{g\in\mathcal{G}}\bigl|g(X_n)-\mathsf{P}g\bigr|\geq n\bigr)<\infty && (6.17)\end{aligned}$$

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 (6.17) から

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

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

$$\frac{1}{n}H(\delta, \mathcal{G}, \|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)})\overset{\mathrm{P}}{\to}0 \qquad (n\to\infty) \tag{6.18}$$

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 I と II を解説する. I は素朴な証明であり, II は chaining argument を用いている.

I 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

$$\mathsf{Pr}\left(\left|\int g\,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)\right| > \frac{\delta}{2}\right) \leq \frac{1}{2} \tag{6.19}$$

が成立すると

$$\mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|g\,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)\right| > \delta\right) \leq 4\mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{j=1}^{n}\epsilon_j g_(X_j)\right| > \frac{\delta}{4}\right) \tag{6.20}$$

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, (6.19) 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[12] 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

$$\mathsf{Pr}\left(\max_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{j=1}^{n} g(X_j)\right| > K\sqrt{\frac{2(\log N + t)}{n}}\right) \leq e^{-t} \tag{6.21}$$

と評価である. さらに, Chebyshev の不等式を用いて (6.19) が成立することが確認できる. このときから　(6.20) と (6.21) を用いて

$$\mathsf{Pr}\left(\max_{g\in\mathcal{G}}\left|\int g\,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)\right| > \frac{K}{4}\sqrt{\frac{2(\log N + t)}{n}}\right) \leq 8\,e^{-t} \tag{6.22}$$

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 (6.18) をみたす $\mathcal{G}$ に対して

$$G(x) := \sup_{g\in\mathcal{G}}|g(x)| \leq K \tag{6.23}$$

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, $\mathcal{G}_K := \big\{g\mathbb{1}_{(-\infty,\,K]}(G)\big\}$ 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

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

とでき, $\big\{G\mathbb{1}_{(-\infty,\,K]}(G)\big\}\bigcup\mathcal{G}_{K_0}$ も P-GC 族であることを利用して, (6.23) がなくとも $\mathcal{G}$ は P-GC 族であることを最後に示す.

[12]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.

II 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 6.29) 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

$$\frac{1}{n}H(\delta, \mathcal{G}, \|\cdot\|_{L_2(\widehat{\mathsf{P}}_n)}) \xrightarrow{\mathrm{P}} 0 \qquad (n\to\infty) \tag{6.24}$$

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, $\mathcal{G}_K := \{g\mathbb{1}\{G\le K;\, g\in\mathcal{G}\}$ とおく. 関数族 $\mathcal{G}$ は, 定理 6.20 の条件 (6.18) と条件 (6.23 をみたすとき, $\forall\delta>0$ に対して

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

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

$$\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{j=1}^n g(X_j) - \int g\,\mathrm{d}\mathsf{P}\right| \xrightarrow{\mathrm{P}} 0 \qquad (n\to\infty) \tag{6.25}$$

を証明する. すると (6.25) と逆マルチンゲールの収束定理の議論から

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

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 [13] とする. すなわち, $\mathsf{E}[X_{i,s}] = 0$ である. ただ

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

し, $\mathcal{T}$ 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

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

と

$$\mathsf{E}\left[\sup_{s\in\mathcal{T}}\sum_{j=1}^{n}X_{j,s}\right] \le 2\mathsf{E}\left[\sup_{s\in\mathcal{T}}\sum_{j=1}^{n}\epsilon_j X_{j,s}\right] \tag{6.27}$$

が成立する.

Proof. ① (6.26) 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 [14].

[14] 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 (6.26) の 1 番目の不等号の証明: ① と同じように示せばよい.
③ (6.27) 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

$$\mathsf{Pr}\left(\left|\int g \,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)\right| > \frac{\delta}{2}\right) \leq \frac{1}{2} \tag{6.19}$$

が成立する [15] 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

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

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

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

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. $\widehat{\mathsf{P}}_n$

[15] (6.19) から

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

となることに注意する.

と $\widehat{\mathsf{P}}'_n$ の独立性から

$$
\begin{aligned}
&\mathsf{Pr}\left(\boldsymbol{X} \in A_{g^*} \text{ かつ } \left|\int g^* \,\mathrm{d}\big(\widehat{\mathsf{P}}'_n - \mathsf{P}\big)\right| \le \frac{\delta}{2}\right) \\
&= \mathsf{E}\left[\mathbb{1}\left\{\{\boldsymbol{X} \in A_{g^*}\} \bigcap \left\{\left|\int g^* \,\mathrm{d}\big(\widehat{\mathsf{P}}'_n - \mathsf{P}\big)\right| \le \frac{\delta}{2}\right\}\right\}\right] \\
&= \mathsf{E}_{\boldsymbol{X}}\left[\mathbb{1}\left\{\{\boldsymbol{X} \in A_{g^*}\} \mathsf{E}\left[\mathbb{1}\left\{\left|\int g^* \,\mathrm{d}\big(\widehat{\mathsf{P}}'_n - \mathsf{P}\big)\right| \le \frac{\delta}{2}\right\}\middle|\boldsymbol{X}\right]\right]\right. \\
&= \mathsf{E}_{\boldsymbol{X}}\left[\mathbb{1}\left\{\{\boldsymbol{X} \in A_{g^*}\} \underbrace{\mathsf{Pr}\left(\left|\int g^* \,\mathrm{d}\big(\widehat{\mathsf{P}}'_n - \mathsf{P}\big)\right| \ge \frac{\delta}{2}\right\}\middle|\boldsymbol{X}\right)}_{\ge 1/2 \quad \because (6.19)}\right. \Bigg] \\
&\ge \frac{1}{2}\mathsf{Pr}\big(A_{g^*}\big) \\
&= \frac{1}{2}\mathsf{Pr}\left(\left|\int g^* \,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)\right| > \delta\right)
\end{aligned}
\tag{6.28}
$$

となる．この不等式を用いると

$$
\begin{aligned}
&\mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|\int g \,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)\right| > \delta\right) = \mathsf{Pr}\left(\boldsymbol{X} \in \bigcup_{g\in\mathcal{G}} A_g)\right) \\
&\le \mathsf{Pr}\big(\boldsymbol{X} \in A_{g^*}\big) \\
&= \mathsf{Pr}\left(\left|\int g_* \,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)\right| > \delta\right) \\
&\le 2\mathsf{Pr}\left(\boldsymbol{X} \in A_{g^*} \text{ かつ } \left|\int g^* \,\mathrm{d}\big(\widehat{\mathsf{P}}'_n - \mathsf{P}\big)\right| \le \frac{\delta}{2}\right) \qquad (\because (6.28)) \\
&= 2\mathsf{Pr}\left(\left\{\boldsymbol{X} \in A_{g^*}\right\} \bigcap \left\{\left|\int g^* \,\mathrm{d}\big(\widehat{\mathsf{P}}'_n - \mathsf{P}\big)\right| \le \frac{\delta}{2}\right\}\right) \\
&= 2\mathsf{Pr}\left(\left\{\left|\int g^* \,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)\right| > \delta\right\} \bigcap \left\{\left|\int g^* \,\mathrm{d}\big(\widehat{\mathsf{P}}'_n - \mathsf{P}\big)\right| \le \frac{\delta}{2}\right\}\right) \\
&\le 2\mathsf{Pr}\left(\left|\int g^* \,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \widehat{\mathsf{P}}'_n\big)\right| > \frac{\delta}{2}\right) \\
&\le 2\mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|\int g \,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \widehat{\mathsf{P}}'_n\big)\right| > \frac{\delta}{2}\right)
\end{aligned}
$$

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 [16] 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

$$\mathsf{Pr}\left(\left|\int g \,\mathrm{d}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)\right| > \frac{\delta}{2}\right) \le \frac{1}{2} \tag{6.29}$$

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.

[16] 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

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

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

$$\max_{g\in\mathcal{G}}\|g\|_\infty\le K \tag{6.30}$$

と仮定する．このとき, $t>0$ に対して

$$\mathsf{Pr}\left(\max_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{j=1}^{n}g(X_j)\right|>K\sqrt{\frac{2(\log N+t)}{n}}\right)\leq e^{-t} \tag{6.21}$$

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

$$\mathsf{Pr}\left(\max_{g\in\mathcal{G}}\left|\int g\,\mathrm{d}\big(\widehat{\mathsf{P}}_n-\mathsf{P}\big)\right|>\frac{K}{4}\sqrt{\frac{2(\log N+t)}{n}}\right)\leq 8\,e^{-t} \tag{6.22}$$

が成り立つ．

Proof. (6.21) の証明: $\epsilon_1,\,\epsilon_2,\,\dots,\,\epsilon_n$ を $X_1,\,X_2,\,\dots,\,X_n$ とは独立な Rademacher 列とする．$g\in\mathcal{G}$ と $j=1,\,2,\dots,\,n$ に対して

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

となる．また, (6.30) と $\epsilon_j\in\{-1,\,1\}\,(j=1,\,2,\,\dots\,,N)$ から

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

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

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

を得る．さらに, towering property から

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

を得る. 次に, 上の不等式と union bound を用いると

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

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, (6.21) は証明できた.

<u>(6.22) の証明</u>: (6.21) と系 6.24 を用いて証明する. そのために条件 (6.29) を確認する. まず, (6.30) 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 1.34）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, (6.29) 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 縮小写像の原理

関数 $f : \mathbb{R} \longrightarrow \mathbb{R}$ は**縮小写像**であるとは

$$\bigl|f(y) - f(x)\bigr| \leq |y - x| \qquad (\forall x,\, y \in \mathbb{R})$$

をみたすときをいう.

定理 6.26. $\Phi : [0,\, \infty) \longrightarrow [0,\, \infty)$ を非減少凸関数関数とし, $T \subset \mathbb{R}^n$ を空でない有界部分集合とする. $f_j : \mathbb{R} \longrightarrow \mathbb{R}\,(j = 1,\, 2,\, \ldots,\, n)$ を縮小写像で $f_j(0) = 0$ をみたすとする. $\{\epsilon_j\}_{j=1}^n$ を独立な Rademacher 列としたとき

$$\mathsf{E}\left[\Phi\left(\frac{1}{2}\sup_{\boldsymbol{t}\in T}\left|\sum_{j=1}^n f_j(t_j)\epsilon_j\right|\right)\right] \leq \mathsf{E}\left[\sup_{\boldsymbol{t}\in T}\left|\sum_{j=1}^n \epsilon_j t_j\right|\right] \tag{6.31}$$

が成り立つ. ただし, $\boldsymbol{t} = (t_1,\, t_2,\, \ldots,\, t_n)$ とした.

Proof. □

系 6.27. $\mathbb{X} \subset \mathbb{R}$, $\mathcal{G}$ を $\mathbb{X}$ 上の実数値関数のある族, P を $\mathbb{X}$ 上の確率分布とする. $X_1,\, X_2,\, \ldots,\, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P}$ とし, $\{\epsilon_j\}_{j=1}^n$ を $\{X_j\}_{j=1}^n$ とは独立に分布する独立な Rademacher 列とする. $\sigma^2\,(\sigma > 0)$ を任意の正の定数で

$$\sup_{g\in\mathcal{G}} \mathsf{P}g^2 \leq \sigma^2$$

とする. このとき

$$\mathsf{E}\left[\sup_{g\in\mathcal{G}}\left|\sum_{j=1}^n g^2(X_j)\right|\right] \leq n\sigma^2 + 8\mathsf{E}\left[\left\{\max_{j=1,\,2,\,\ldots,\,n} G(X_j)\right\}\sup_{g\in\mathcal{G}}\bigl|\epsilon_j g(X_j)\bigr|\right] \tag{6.32}$$

が成り立つ. ただし, $G(x) = \sup_{g\in\mathcal{G}}|g(x)|\,(x \in \mathbb{X})$ である. が成立する.

Proof. 三角不等式から

$$\begin{aligned}\left|\sum_{j=1}^n g^2(X_j)\right| &\leq \left|\sum_{j=1}^n \bigl\{g^2(X_j) - \mathsf{P}g^2\bigr\} + n\mathsf{P}g^2\right| \\ &\leq n\mathsf{P}g^2 + \left|\sum_{j=1}^n \bigl\{g^2(X_j) - \mathsf{P}g^2\bigr\}\right| \\ &\leq n\sigma^2 + \left|\sum_{j=1}^n \bigl\{g^2(X_j) - \mathsf{P}g^2\bigr\}\right|\end{aligned}$$

となる．上の不等式の両辺の sup を取り, 期待値を取ったものの最右辺の第 2 項目に補題 6.22(6.27) を適用すると

$$\mathsf{E}\left[\sup_{g\in\mathcal{G}}\left|\sum_{j=1}^{n} g^2(X_j)\right|\right] \leq n\sigma^2 + 2\mathsf{E}\left[\sup_{g\in\mathcal{G}}\left|\epsilon_j g^2(X_j)\right|\right]$$

となる．$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ を固定して, $M := \max_{j=1,2,\ldots,n} G(X_j)$ とおく．関数 $\varphi\colon \mathbb{R} \longrightarrow \mathbb{R}$ を

$$\varphi(x) = \begin{cases} M^2 & (x > M) \\ x^2 & (|x| < M) \\ M^2 & (x < -M) \end{cases}$$

で定める．すると φ は Lipschitz 連続な関数で, その Lipschitz 定数は $2M$ となる．すなわち

$$\bigl|\varphi(y) - \varphi(x)\bigr| \leq 2M|y - x| \qquad (x,\, y \in \mathbb{R})$$

となる．すると定理 6.26 を適用する[17]と

$$\mathsf{E}\left[\sup_{g\in\mathcal{G}}\Bigl|\sum_{j=1}^{n} \epsilon_j g^2(X_j)\Bigr|\,\Big|\,\boldsymbol{X}\right] \leq 4M\mathsf{E}\left[\sup_{g\in\mathcal{G}}\Bigl|\sum_{j=1}^{n} \epsilon_j g(X_j)\Bigr|\,\Big|\,\boldsymbol{X}\right]$$

を得る．towering property を用いると

$$\begin{aligned}
\mathsf{E}\left[\sup_{g\in\mathcal{G}}\left|\sum_{j=1}^{n} g^2(X_j)\right|\right] &= \mathsf{E}\left[\mathsf{E}\left[\sup_{g\in\mathcal{G}}\Bigl|\sum_{j=1}^{n} \epsilon_j g^2(X_j)\Bigr|\,\Big|\,\boldsymbol{X}\right]\right] \\
&\leq \mathsf{E}\left[n\sigma^2 + 8\mathsf{E}\left[\left\{\max_{j=1,2,\ldots,n} G(X_j)\right\}\sup_{g\in\mathcal{G}}\Bigl|\sum_{j=1}^{n} \epsilon_j g(X_j)\Bigr|\,\Big|\,\boldsymbol{X}\right]\right] \\
&\leq n\sigma^2 + 8\mathsf{E}\left[\left\{\max_{j=1,2,\ldots,n} G(X_j)\right\}\sup_{g\in\mathcal{G}}\left|\epsilon_j g(X_j)\right|\right]
\end{aligned}$$

から系は証明された． □

6.6.3 Dudley のエントロピー積分

補題 6.28. $X_1, X_2, \ldots, X_N$ 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, $\mathsf{E}\bigl[e^{\lambda X_j}\bigr] \leq e^{\lambda^2\nu/2}\,(j = 1, 2, \ldots, N)$ が成立す

[17] $\boldsymbol{X}$ は与えられているので, 定理 6.26 において

$$\Phi(x) = x; \quad \varphi_j(\,\cdot\,) = \frac{\varphi_j(\,\cdot\,)}{2M}, \quad t_j = g(X_j)\,(j = 1, 2, \ldots, n)$$

として, この定理を適用する．

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 **??**) を用いると

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

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, $\lambda = \sqrt{2\nu^{-1} \log N}$ 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 6.29. $(\mathbb{T},\, d)$ 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

$$\log\left(\mathsf{E}\bigl[\exp\{\lambda(X_{t_1} - X_{t_2})\}\bigr]\right) \le \frac{\lambda^2 d^2(t_1,\, t_2)}{2} \tag{6.33}$$

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

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

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

$$\mathsf{E}\left[\sup_{t_1,\, t_2\in\mathbb{T}} \bigl|X_{t_1} - X_{t_2}\bigr|\right] \le 24 \int_0^{\delta/2} \sqrt{H(\epsilon,\, \mathbb{T},\, d)}\, \mathrm{d}\epsilon \tag{6.34}$$

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 $\mathbb{T}_j := \{t_1, t_2, \ldots, t_{N_j}\} \subset \mathbb{T}$ 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 $t_j\,(\exists j \in \{1, 2, \ldots, N_j)$ 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, $\#\{j \in \{1, 2, \ldots, N_j;\, d(t_j, t) < \delta_j\} \geq 2$ の場合には, そのうちのどれかを対応させればよい. ここで, $\mathbb{T}_0 = \{t_0\}$ とし, $\Pi_0(t) = t_0$ と定義する.
Step 1: つぎに

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

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

$$\mathsf{E}\left[\sup_{t\in\mathbb{T}}\bigl|X_t - X_{t_0}\bigr|\right] \leq \sum_{j=0}^{\infty}\mathsf{E}\left[\sup_{t\in\mathbb{T}}\bigl|X_{\Pi_{j+1}(t)} - X_{\Pi_j(t)}\bigr|\right] \tag{6.35}$$

である. また

$$\begin{aligned}
\#\bigl\{\bigl(\Pi_j(t), \Pi_{j+1}(t)\bigr);\, t \in \mathbb{T}\bigr\} &\leq N(\delta_j, \mathbb{T}, d) \times N(\delta_{j+1}, \mathbb{T}, d)\\
&\leq \{N(\delta_{j+1}, \mathbb{T}, d)\}^2\\
&= \exp\Bigl(\log\bigl(\{N(\delta_{j+1}, \mathbb{T}, d)\}^2\bigr)\Bigr)\\
&= \exp\Bigl(2H(\delta_{j+1}, \mathbb{T}, d)\Bigr) \qquad (6.36)
\end{aligned}$$

となることがわかる. 三角不等式から

$$d\bigl(\Pi_j(t), \Pi_{j+1}(t)\bigr) \leq d\bigl(\Pi_j(t), t\bigr) + d\bigl(t, \Pi_{j+1}(t)\bigr) \leq \delta_j + \delta_{j+1} \leq 3\delta_{j+1} \tag{6.37}$$

となる. (6.33) と (6.37) から

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

であるので

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

となることがわかる. よって, 補題 6.28 と (6.36) から

$$\begin{aligned}
\mathsf{E}\Big[\sup_{t\in\mathbb{T}}\big|X_{\Pi_{j+1}(t)} - X_{\Pi_j(t)}\big|\Big] &\le \sqrt{2\times 9\delta_{j+1}^2\log\big(\#\{\big(\Pi_j(t),\,\Pi_{j+1}(t)\big);\,t\in\mathbb{T}\}\big)} \\
&\le \sqrt{18\delta_{j+1}^2 \times 2H(\delta_{j+1},\,\mathbb{T},\,d)} \\
&= 6\delta_{j+1}\sqrt{H(\delta_{j+1},\,\mathbb{T},\,d)} \qquad (6.38)
\end{aligned}$$

を得る. (6.35) の右辺に (6.38) 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 6.30. $X_1,\,X_2,\,\ldots\,X_n$ 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

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

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 6.29 を用いるために, 条件 (6.33) を確認する. すなわち, 過程 $\left\{\sum_{j=1}^{n}\epsilon_j g(X_j)\right\}_{g\in\mathcal{G}}$ 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

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

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, (6.33) が成立することが示せた. Dudley のエントロピー上限 (補題 6.30) を適用するた

めに, $g_0 \in \mathcal{G}$ を取る. すると

$$
\begin{aligned}
&\mathsf{E}\Bigl[\sup_{g\in\mathcal{G}}\Bigl|n^{-1}\sum_{j=1}^{n}\epsilon_j\bigl\{g(X_j)-g_0(X_j)\bigr\}\Bigr|\,\Big|\,\boldsymbol{X}\Bigr] \\
&\le 12\int_0^{\delta_n/2}\sqrt{H\bigl(\epsilon,\,\mathcal{G},\,n^{-1/2}\|\cdot\|_{L_2(\widehat{\mathsf{P}}_n)}\bigr)}\,\mathrm{d}\epsilon \\
&= 12\int_0^{\delta_n/2}\sqrt{H\bigl(\sqrt{n}\epsilon,\,\mathcal{G},\,\|\cdot\|_{L_2(\widehat{\mathsf{P}}_n)}\bigr)}\,\mathrm{d}\epsilon \\
&= 12\int_0^{\sqrt{n}\delta_n/2}\sqrt{\frac{H\bigl(\epsilon',\,\mathcal{G},\,\|\cdot\|_{L_2(\widehat{\mathsf{P}}_n)}\bigr)}{n}}\,\mathrm{d}\epsilon' \\
&\qquad (\epsilon'=\sqrt{n}\epsilon \text{ と変換}) \\
&= 12\int_0^{D_n}\sqrt{\frac{H\bigl(\epsilon,\,\mathcal{G},\,\|\cdot\|_{L_2(\widehat{\mathsf{P}}_n)}\bigr)}{n}}\,\mathrm{d}\epsilon
\end{aligned}
\tag{6.39}
$$

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

$$
\begin{aligned}
&\mathsf{E}\Bigl[\Bigl|n^{-1}\sum_{j=1}^{n}\epsilon_j g_0(X_j)\Bigr|\,\Big|\,\boldsymbol{X}\Bigr] \\
&= \int_0^\infty \mathsf{Pr}\Bigl(\Bigl|n^{-1}\sum_{j=1}^{n}\epsilon_j g_0(X_j)\Bigr| > t\,\Big|\,\boldsymbol{X}\Bigr)\,\mathrm{d}t \\
&\qquad (\because \text{命題 } 1.32) \\
&\le \int_0^\infty 2\exp\Bigl(-\frac{2n^2t^2}{4\sum_{j=1}^{n}\{g_0(X_j)\}^2}\Bigr)\,\mathrm{d}t \\
&\qquad (\because \text{命題 } 5.14) \\
&= \int_{-\infty}^\infty \exp\Bigl(-\frac{n^2t^2}{2\sum_{j=1}^{n}\{g_0(X_j)\}^2}\Bigr)\,\mathrm{d}t \\
&= \int_{-\infty}^\infty \exp\Bigl(-\frac{t^2}{2\{\|g_0\|_{L_2(\widehat{\mathsf{P}})}/n\}^2}\Bigr)\,\mathrm{d}t \\
&= \sqrt{2\pi}\,\frac{\|g_0\|_{L_2(\widehat{\mathsf{P}}_n)}}{4\sqrt{n}} \\
&\le \sqrt{2\pi}\,\frac{\sup_{g\in\mathcal{G}}\|g\|_{L_2(\widehat{\mathsf{P}})}}{\sqrt{n}}
\end{aligned}
\tag{6.40}
$$

を得る. (6.39) と (6.40) 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 6.31. 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

$$
\frac{1}{n}H(\epsilon,\mathcal{G},\|\cdot\|_{L_2(\widehat{\mathsf{P}}_n)})\xrightarrow{\mathrm{P}}0\qquad(n\to\infty)\tag{6.24}
$$

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 6.30 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 2.7(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, (6.24) と定理 2.9(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, $g_1, g_2, \ldots, g_N$ 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 $j \in \{1, 2, \ldots, N\}$ に対して

$$
\widehat{\mathsf{P}}_n|g_j - g| := \frac{1}{n}\sum_{k=1}^n \bigl|g_j(X_k) - g(x)\bigr| < \delta \tag{6.41}
$$

とできる．このことから

$$\begin{aligned}\left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k g(X_k)\right| &\le \left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k g_j(X_k)\right| + \left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k\big\{g(X_k)-g_j(X_k)\big\}\right|\\ &\le \left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k g_j(X_k)\right| + \underbrace{\frac{1}{n}\sum_{k=1}^{n}\big|g(X_k)-g_j(X_k)\big|}_{<\delta \quad \because (6.41)}\\ &\le \left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k g_j(X_k)\right| + \delta\end{aligned}$$

である．したがって

$$\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k g(X_k)\right| \le \max_{j=1,2,\ldots,N}\left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k g_j(X_k)\right| + \delta \tag{6.42}$$

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

$$\mathsf{Pr}\left(\max_{j=1,2,\ldots,N}\left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k g_j(X_k)\right| > K\sqrt{\frac{2(t+\log N)}{n}}\,\middle|\,\boldsymbol{X}\right) \le 2e^{-t} \tag{6.43}$$

となる．(6.42) と (6.43) を合わせると

$$\begin{aligned}&\mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k g(X_k)\right| > \delta + K\sqrt{\frac{2(t+\log N)}{n}}\,\middle|\,\boldsymbol{X}\right)\\ &\quad\le \mathsf{Pr}\left(\max_{j=1,2,\ldots,N}\left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k g_j(X_k)\right| > K\sqrt{\frac{2(t+\log N)}{n}}\,\middle|\,\boldsymbol{X}\right) \le 2e^{-t}\end{aligned} \tag{6.44}$$

を得る．条件付き期待値の性質 (towering property) を (6.44) に適用すると

$$\begin{aligned}&\mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k g(X_k)\right| > 2\delta + K\sqrt{\frac{2t}{n}}\right)\\ &\quad\le 2e^{-t} + \mathsf{Pr}\left(K\sqrt{\frac{2H_1(\delta, \mathcal{G}, \|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)})}{n}} > \delta\right)\end{aligned} \tag{6.45}$$

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

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

となる.

$$\begin{aligned}
&\mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k g(X_k)\right| > 2\delta + K\sqrt{\frac{2t}{n}}\right)\\
&\quad\leq \mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k g(X_k)\right| > 2\delta + K\sqrt{\frac{2t}{n}} \text{ かつ } A^c\right)\\
&\qquad + \mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k g(X_k)\right| > 2\delta + K\sqrt{\frac{2t}{n}} \text{ かつ } A\right)\\
&\quad\leq \mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k g(X_k)\right| > \delta + K\sqrt{\frac{2\log N}{n}} + K\sqrt{\frac{2t}{n}}\right)\\
&\qquad + \mathsf{Pr}\left(K\sqrt{\frac{2H_1(\delta,\mathcal{G},\|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)})}{n}} > \delta\right)\\
&\qquad\left(\because A^c \Rightarrow K\sqrt{\frac{2\log N}{n}} \leq \delta\right.\\
&\qquad\qquad\left.\Rightarrow 2\delta + K\sqrt{\frac{2t}{n}} \geq \delta + K\sqrt{\frac{2\log N}{n}} + K\sqrt{\frac{2t}{n}}\right)\\
&\quad\leq \mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k g(X_k)\right| > \delta + K\sqrt{\frac{2(t+\log N)}{n}}\right)\\
&\qquad + \mathsf{Pr}\left(K\sqrt{\frac{2H_1(\delta,\mathcal{G},\|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)})}{n}} > \delta\right)\\
&\qquad\qquad\left(\because \sqrt{a}+\sqrt{b} \geq \sqrt{a+b}\,(a,\,b\geq 0)\right)\\
&\quad= \mathsf{E}\left[\mathsf{Pr}\left(\sup_{g\in\mathcal{G}}\left|\frac{1}{n}\sum_{k=1}^{n}\epsilon_k g(X_k)\right| > \delta + K\sqrt{\frac{2(t+\log N)}{n}}\right)\middle|\boldsymbol{X}\right]\\
&\qquad + \mathsf{Pr}\left(K\sqrt{\frac{2H_1(\delta,\mathcal{G},\|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)})}{n}} > \delta\right)\\
&\quad\leq 2e^{-t} + \mathsf{Pr}\left(K\sqrt{\frac{2H_1(\delta,\mathcal{G},\|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)})}{n}} > \delta\right) \qquad (\because (6.45))
\end{aligned}$$

からわかる. よって, (6.45) が確認できた.

次に, 系 6.24 を (6.45) に適用するために, 以下の (6.29) 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 1.34) を用いると

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

より

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

が (6.29) 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 (6.45) を用いれば

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

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, (6.29) 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. $\sup_{g\in\mathcal{G}}|g(x)|\le K$ 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,

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

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

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

となる. 評価できる. 一方,

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

である. 任意の K に対して, 大数の法則より

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

となる. また, $G\in L_1(\mathsf{P})$ なので

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

である. 以上から

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

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 (6.18) 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 6.31 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 6.32. $\mathbb{X}=\mathbb{R}$ 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

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

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, $\{\xi_1, \xi_2, \dots, \xi_n\}$ を $\mathbb{X}$ の点の集合とする.

定義 6.33.

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

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

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

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

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

と書く.

例 6.34. $\mathbb{X} = \mathbb{R}$ 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,

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

定理 6.35. 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) $\Rightarrow$ (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

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

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

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

から誘導された半ノルムに関する δ 被覆数とする. このとき

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

であることが

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

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) $\Rightarrow$ (1) の証明: 省略. □

定義 6.36. $n \in \mathbb{N}$ 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,

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

が成り立つときをいう.

定理 6.37. $\mathcal{D}$ 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,

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

となる. □

例 6.38. (1) $\mathbb{X} = \mathbb{R}$, $\mathcal{D} = \{(-\infty,\, r];\, r \in \mathbb{R}\}$ は VC 族である. なぜならば, $\mathrm{VC}^{\mathcal{D}}(n) \le n + 1$ よりわかる.
(2) $p \ge 2$ 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, $\mathrm{VC}^{\mathcal{D}}(n) \le (n+1)^p$ 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

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

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 6.39. $\mathcal{D}$, $\mathcal{D}_1$, $\mathcal{D}_2$ 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

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

Proof. 証明は略. □

定義 6.40. 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 6.41. 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,

$$\mathrm{VC}^{\mathcal{D}} \leq \sum_{k=0}^{V} \begin{pmatrix} n \\ k \end{pmatrix}.$$

□

6.9 VC 関数族

定義 6.42. 関数 $g;\ \mathbb{X} \to \mathbb{R}$ のサブグラフとは

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

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

例 6.43. $\mathcal{D}$ が VC 族ならば, $\mathcal{G} = \{\mathbb{1}_D;\ D \in \mathcal{D}\}$ も VC 族.

注意 6.44. 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,

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

としたとき,

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

である. Pollard (1984) を参照.

定義 6.45. 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 6.46. 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 6.47. 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[18], $N(\delta, \mathcal{G}, \|\cdot\|_{L_{(\mathsf{Q})}})\,(\delta > 0)$ 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 $\text{VC}^{\mathcal{G}} =: V$ が $V < \infty$ のとき, V に依存する定数 $A > 0$ が存在して

$$N(\delta \mathsf{Q} G, \mathcal{G}, \|\cdot\|_{L_1(\mathsf{Q})}) \leq \max\{A\delta^{-2V}, e^{\delta/4}\} \tag{6.46}$$

となる.

[18] $G(x) = \sup_{g \in \mathcal{G}} |g(x)|$ 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

$$n \geq \frac{4\log N}{\delta} \tag{6.47}$$

をみたす最小の n を取ると

$$\binom{N}{2}\left(1 - \frac{\delta}{2}\right)^n \leq \frac{1}{2} < 1 \tag{6.48}$$

とできる．また, n は (6.47) をみたす最小の n なので

$$n \le \frac{4\log N}{\delta} + 1 \tag{6.49}$$

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, (6.48) 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

$$\begin{aligned} D(\delta,\,\mathcal{G},\,\|\cdot\|_{L_1(\mathsf{Q})}) = N &\le \sup\left\{\Delta^{\mathcal{D}}\Big((S_1,\,T_1),\,(S_2,\,T_2),\,\ldots,\,(S_n,\,T_n)\Big)\right. \\ &\qquad\qquad \left.;\,(S_i,\,T_i) \in \mathcal{X}\times\mathbb{R}\,(i = 1,\,2,\,\ldots,\,n)\right\} \\ &\le Cn^V \end{aligned} \tag{6.50}$$

となる．ただし, C は定義 6.36 の定数であり

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

である．

$N \ge \exp\left\{\dfrac{\delta}{4}\right\}$ の場合を考える [19]．すると, (6.47) は

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

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

$$\begin{aligned} N &\le Cn^V \qquad (\because (6.50)) \\ &\le C\left(\frac{4\log N}{\delta} + 1\right)^V \qquad (\because (6.49)) \\ &\le C\left(\frac{8\log N}{\delta}\right)^V \quad \left(N \ge e^{\delta/4} \Rightarrow \frac{4\log N}{\delta} \ge 1 \because\right) \\ &= C\left(\frac{16V\log N^{1/(2V)}}{\delta}\right)^V \\ &\le C\left(\frac{16V}{\delta}\right)^V \sqrt{N} \end{aligned}$$

となる．この不等式の両辺を $\sqrt{N}$ で割ると

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

[19] $N \ge e^{\delta/4}$ でないときは, $N < e^{\delta/4}$ となるので, (6.46 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 6.48. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P}$ 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

$$\frac{1}{n}H\big(\widetilde{\delta}, \mathcal{G}, \|\cdot\|_{L_1(\widehat{\mathsf{P}}_n)}\big) \xrightarrow{\mathrm{P}} 0 \qquad (n \to \infty) \tag{6.51}$$

である．すなわち，(6.51) を確認すればよい．そのために，定理 6.47 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

$$N(\delta \mathsf{Q}G, \mathcal{G}, \|\cdot\|_{L_1(\mathsf{Q})}) \le \max\big\{Ae^{-2V}, e^{\delta/4}\big\} \tag{6.52}$$

であった．ここで，$\mathsf{Q} = \widehat{\mathsf{P}}_n$ とおくと，(6.52) 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 [20]．最後に, $\widetilde{\delta} = 2\delta\mathsf{P}G$ とおくと (6.51) 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.

$$\mathsf{p}_{\mathsf{F}_0}(x) := \int \mathsf{k}(x|\,y)\,\mathrm{d}\mathsf{F}_0(y) \qquad (x \in \mathbb{R}) \tag{6.53}$$

を持つことがわかる．

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

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

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 $\widehat{F}_n^{\mathrm{mle}}$ 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 6.49. $p \geq 1$ とし, P を距離空間 $\mathbb{R}$ 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

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

は以下をみたすとする.

- すべての $x \in \mathbb{R}$ に対して, 写像

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

  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

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

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 1.41) から

$$\lim_{\rho \to 0} \mathsf{P}\omega_{\theta\,\rho}^p = 0$$

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

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

とできる．つぎに

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

とおくと $\left\{B_\theta;\, \theta \in \Theta\right\}$ 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 $\theta \in B_{\theta_j}$ と $x \in \mathbb{X}$ に対して

$$g_\theta(x) := g_{\theta_j}(x) + \omega_{\theta_j,\rho_{\theta_j}}(x) =: g_j^{\mathrm{R}}(x),$$
$$g_\theta(x) := g_{\theta_j}(x) - \omega_{\theta_j,\rho_{\theta_j}}(x) =: g_j^{\mathrm{L}}(x),$$

とおけば

$$g^L_{\theta_j}(x) \le g_\theta(x) \le g^R_{\theta_j}(x)\,(\forall x \in \mathbb{R}) \quad \text{かつ} \quad \mathsf{P}|g_j^{\mathrm{R}} - g_j^{\mathrm{L}}|^p \le \mathsf{P}\big(2\omega_{\theta_j,\rho_{\theta_j}}\big)^p \le (2\delta)^p$$

となる．よって

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

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, 実数値確率変数 X は, (6.53) で与えれる p.d.f. $\mathsf{p}_{\mathsf{F}_0}$ を持つとする．また, X の確率分布を P_0 と書くことにする．$\mathcal{F}$ 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）を

$$\widehat{F}_n^{\mathrm{mle}} \in \arg\max_{F \in \mathcal{F}} \widehat{\mathsf{P}}_n \log \mathsf{p}_F \tag{6.54}$$

で定める．ただし

$$\widehat{\mathsf{P}}_n(B) = \frac{1}{n}\sum_{j=1}^n \mathbb{1}_B(X_j)\,\big(\forall B \in \mathbb{B}(\mathbb{R})\big), \tag{6.55}$$

である．また, 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．ただし, m は $\mathbb{R}$ 上の Lebesgue 測度である．

注意 6.50. F_0 の最尤推定量 $\widehat{F}_n^{\mathrm{mle}}$ は F_0 の経験分布関数 $\widehat{\mathsf{F}}_n$ と一致するとは限らない. ただし

$$\widehat{\mathsf{F}}_n(x) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{(-\infty,\, x]}(X_j)\ \bigl(\forall x \in \mathbb{R}\bigr)$$

である. □

補題 6.51. $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P}_0$ とする. ただし, P_0 は, (6.53) で与えられる p.d.f. $\mathsf{p}_{\mathsf{F}_0}$ から得られる確率分布である. 分布関数 F_0 の最尤推定量 $\widehat{F}_n^{\mathrm{mle}}$ (すなわち, (6.54) で与えられるもの) に対して

$$h(\mathsf{p}_{\widehat{F}_n^{\mathrm{mle}}},\, \mathsf{p}_{\mathsf{F}_0}) \overset{\text{a.s.}}{\to} 0 \qquad (n \to \infty)$$

が成立する.

Proof. $\mathbb{R}$ 上の分布関数と $\mathbb{R}$ の Borel 測度とは対応関係があるので, 分布関数全体の集合 $\mathcal{F}$ を $\mathbb{R}$ 上の Borel 確率測度全体の集合 $\mathcal{M}$ と同一視する. すると $\mathcal{M}$ の弱収束に対応する距離関数 d が取れること [21] が知られている. すなわち, $\{\mathsf{P}_n\}_{n=0}^{\infty} \subset \mathcal{M}$ に対して

$$\mathsf{P}_n \rightsquigarrow \mathsf{P}_0\,(n \to \infty) \quad \Leftrightarrow \quad d(\mathsf{P}_n,\, \mathsf{P}_0) \overset{n\to\infty}{\longrightarrow} 0$$

となる. さらに, $\mathcal{F}$ は距離関数 d から誘導される位相に関してコンパクトであること [22] が知られている. 以下では, すこし記号を乱用して, d を $\mathcal{F}$ 上の距離関数ともみなすことにする.

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

$$\mathcal{F} \ni F \mapsto \mathsf{p}_{\mathsf{F}}(x) = \int_{\mathbb{R}} \mathsf{k}(x|\, y)\, \mathrm{d}F(y) \tag{6.56}$$

[21] 確率空間上の有界 Lipschitz 距離関数である. [55, p.73] を参照のこと.
[22] 下記の注意 6.52 を参照のこと.

は d に関して連続であること [23] がわかる. ただし, $\mathcal{F}$ はすべての分布関数の集まりである.

いま, 関数族

$$\mathcal{G} := \left\{ \frac{2\mathsf{p}_F}{\mathsf{p}_F + \mathsf{p}_{\mathsf{F}_0}};\, F \in \mathcal{F} \right\} \tag{6.57}$$

を考える. この関数族の封筒関数 G は

$$G(x) := \sup_{F\in\mathcal{F}} \left| \frac{2\mathsf{p}_F(x)}{\mathsf{p}_F(x) + \mathsf{p}_{\mathsf{F}_0}(x)} \right| \le 2$$

となる. 分布関数全体の集合 $\mathcal{F}$ は弱位相 (距離関数 d によって誘導される位相) に関してコンパクトであった. さらに, 写像 (6.58) の連続性から, 写像

$$\mathcal{F} \ni F \mapsto \frac{2\mathsf{p}_F}{\mathsf{p}_F + \mathsf{p}_{\mathsf{F}_0}} \in \mathcal{G} \tag{6.58}$$

も連続 [24] である. したがって, $G \in L_1(\mathsf{P})$ であり, $\mathcal{F}$ はコンパクトで, 写像 (6.58) は連続であることに注意して, 補題 6.49 を用いると, $\forall\delta > 0$ に対して

$$H_B\big(\delta, \mathcal{G}, \|\cdot\|_{L_1(\mathsf{P})}\big) < \infty$$

となる. ただし, 関数族 $\mathcal{G}$ は (6.57) で与えられたものある. このことと定理 6.16 から $\mathcal{G}$ は P-GC 族となることがわかる. よって

$$\begin{aligned}\left| (\widehat{\mathsf{P}}_n - \mathsf{P}) \left(\frac{2\mathsf{p}_{\widehat{F}_n^{\mathrm{mle}}}}{\mathsf{p}_{\widehat{F}_n^{\mathrm{mle}}} + \mathsf{p}_{\mathsf{F}_0}} \right) \right| &\le \sup_{F\in\mathcal{F}} \left| (\widehat{\mathsf{P}}_n - \mathsf{P}) \left(\frac{2\mathsf{p}_F}{\mathsf{p}_F + \mathsf{p}_{\mathsf{F}_0}} \right) \right| \\ &= \sup_{g\in\mathcal{G}} \big| (\widehat{\mathsf{P}}_n - \mathsf{P}) g \big| \overset{\text{a.s.}}{\to} 0 \quad (n \to \infty)\end{aligned} \tag{6.59}$$

がわかる. しかし

$$\underbrace{1 - \mathsf{P} \left(\frac{2\mathsf{p}_{\widehat{F}_n^{\mathrm{mle}}}}{\mathsf{p}_{\widehat{F}_n^{\mathrm{mle}}} + \mathsf{p}_{\mathsf{F}_0}} \right)}_{\ge 0} \le \left| (\widehat{\mathsf{P}}_n - \mathsf{P}) \left(\frac{2\mathsf{p}_{\widehat{F}_n^{\mathrm{mle}}}}{\mathsf{p}_{\widehat{F}_n^{\mathrm{mle}}} + \mathsf{p}_{\mathsf{F}_0}} \right) \right| \tag{6.60}$$

---

[23] なぜならば, $\mathcal{F}$ の点列 $\{F_n\}_{n=0}^{\infty}$ に対して, 距離関数 d は

$$d(F_n, F_0) \overset{n\to\infty}{\longrightarrow} 0 \quad \Leftrightarrow \quad F_n \rightsquigarrow F_0\,(n\to\infty)$$

をみたす $\mathcal{F}$ 上の距離関数である. さらに, 関数 $y \mapsto \mathsf{k}(x|\,y)$ は有界連続なので, Portmanteau の定理 (定理 2.3) から

$$F_n \rightsquigarrow F_0\,(n\to\infty) \quad \Rightarrow \quad \int_{\mathbb{R}} \mathsf{k}(x|\,y)\,\mathrm{d}F_n(y) \overset{n\to\infty}{\longrightarrow} \int_{\mathbb{R}} \mathsf{k}(x|\,y)\,\mathrm{d}F_0(y) \qquad \text{a.e. } x$$

となることからわかる.

[24] (6.56) は連続写像であった. また, $c > 0$ に対して, 写像 $[0, \infty) \ni t \mapsto \dfrac{2t}{t+c}$ は連続であるので, その合成写像 (6.58) も連続である.

が成り立つこと [25] に注意する. また

$$h^2(\mathsf{p},\,\mathsf{p}_0) \le 1 - \mathsf{P}\left(\frac{2\mathsf{p}}{\mathsf{p}+\mathsf{p}_0}\right) \tag{6.61}$$

であること [26] に注意する. 最後に, (6.59) – (6.61) を合わせると

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

となる. したがって

$$h(\mathsf{p}_{\widehat{F}_n^{\mathrm{mle}}},\,\mathsf{p}_{\mathsf{F}_0}) \overset{\text{a.s.}}{\to} 0 \qquad (n\to\infty)$$

を得る. □

注意 6.52. $\mathbb{R}$ 上の分布関数と $\mathbb{R}$ の Borel 測度とは対応関係があるので, 分布関数全体の集合 $\mathcal{F}$ を $\mathbb{R}$ 上の Borel っ確率測度全体の集合 $\mathcal{M}$ と同一視する. すると $\mathcal{M}$ の弱収束に対応する距離関数 d が取れること [27] が知られている. すなわち, $\{\mathsf{P}_n\}_{n=0}^{\infty} \subset \mathcal{M}$ に対して

$$\mathsf{P}_n \rightsquigarrow \mathsf{P}_0\,(n\to\infty) \quad\Leftrightarrow\quad d(\mathsf{P}_n,\,\mathsf{P}_0) \overset{n\to\infty}{\longrightarrow} 0$$

---

[25] これは

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

からわかる.

[26] これは

$$\begin{aligned} h^2(\mathsf{p},\,\mathsf{p}_0) &= \frac{1}{2}\int_{\mathbb{R}} \{\sqrt{\mathsf{p}} - \sqrt{\mathsf{p}_0}\}^2\,\mathrm{dm} = \frac{1}{2}\int_{\mathbb{R}} \frac{(\mathsf{p}-\mathsf{p}_0)^2}{\{\sqrt{\mathsf{p}} + \sqrt{\mathsf{p}_0}\}^2}\,\mathrm{dm} \le \frac{1}{2}\int_{\mathbb{R}} \frac{(\mathsf{p}-\mathsf{p}_0)^2}{\mathsf{p}+\mathsf{p}_0}\,\mathrm{dm} \\ &= \frac{1}{2}\int_{\mathbb{R}} \frac{\{(\mathsf{p}+\mathsf{p}_0)^2 - 4\mathsf{p}\mathsf{p}_0}{\mathsf{p}+\mathsf{p}_0}\,\mathrm{dm} = \frac{1}{2}\left\{\underbrace{\int_{\mathbb{R}} (\mathsf{p}+\mathsf{p}_0)\,\mathrm{dm}}_{=2} - 4\int_{\mathbb{R}} \frac{\mathsf{p}}{\mathsf{p}+\mathsf{p}_0}\,\mathsf{p}_0\,\mathrm{dm}\right\} \\ &= 1 - \mathsf{P}\left(\frac{2\mathsf{p}}{\mathsf{p}+\mathsf{p}_0}\right) \end{aligned}$$

からわかる.

[27] 確率空間上の有界 Lipschitz 距離関数である. [55, p.73] を参照のこと.

となる．さらに, $\mathcal{M}' = \{\mathsf{P}_1 - \mathsf{P}_2;\ \mathsf{P}_1, \mathsf{P}_2 \in \mathcal{M}\}$ とすると, $\mathcal{M}'$ の双対空間を $\mathbb{R}$ 上の有界 Lipschitz 関数全体の集合 $BL(\mathbb{R})$ と同一視できる．すなわち, 有界 Lipschitz 関数 $h\colon \mathbb{R} \longrightarrow \mathbb{R}$ と $(\mathcal{M}', d)$ の双対空間の元を

$$h\colon \mathcal{M}' \ni \mu \mapsto \int h \,\mathrm{d}\mu$$

で同一視する．これは, $BL(\mathbb{R})$ の双対空間を $\mathcal{M}'$ と同一視することである．さらに, Alaoglu の定理[28] から, $(\mathcal{M}', d)$ の部分空間 $\{\mu \in \mathcal{M}';\ \|\mu\|_d \le 1\} =: \mathcal{M}_0$ はコンパクトなることがわかる．したがって, $\mathcal{F}$ は $\mathcal{M}_0$ の閉部分集合と同一視できるので, d に関してコンパクトであることがわかる．
□

6.11 最大不等式

最大不等式は, 確率変数の上限の裾確率や積率を上から評価するものである．上極限の最大不等式は 2 つの工夫を組み合わせることで導出される．すなわち, chaining argument と有限個の確率変数の最大不等式である．前者の chaining argument は上極限の各要素を小さな偏差 (各項は指数的に小さなもの) の和を評価することで上限を得る手法である．まず, 指数型不等式から始めて, この不等式を有限個の確率変数の上限に適用して, 求めたい最大不等式を導出しよう．

補題 6.53 (Bernstein の不等式)**.** P を $\mathbb{R}$ 上の確率測度とし, $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P}$ とする．g を任意の有界可測関数とする．このとき, すべての $x > 0$ に対して

$$\mathsf{Pr}\left(\left|\sqrt{n}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)g\right| > x\right) \le 2\exp\left(-\frac{1}{4}\frac{x^2}{\mathsf{P}g^2 + x\|g\|_\infty/\sqrt{n}}\right) \qquad (x > 0)$$

が成立する．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．

Proof. まず, $x > 0$ に対して

$$\mathsf{Pr}\big(\sqrt{n}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)g > x\big) \le \exp\left(-\frac{1}{4}\frac{x^2}{\mathsf{P}g^2 + x\|g\|_\infty/\sqrt{n}}\right) \tag{6.62}$$

を示めそう．すると g を $-g$ に置き換えると

$$\mathsf{Pr}\big(\sqrt{n}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)g < -x\big) \le \exp\left(-\frac{1}{4}\frac{x^2}{\mathsf{P}g^2 + x\|g\|_\infty/\sqrt{n}}\right) \tag{6.63}$$

[28] [29, p.194, Exercise 9] を参照のこと．

を得る．Union bound(補題 1.5(7)) を用いてから, (6.62) と (6.63) を代入すると

$$
\begin{aligned}
&\mathsf{Pr}\Bigl(\Bigl|\sqrt{n}\bigl(\widehat{\mathsf{P}}_n-\mathsf{P}\bigr)g\Bigr|>x\Bigr)\\
&\quad=\mathsf{Pr}\Bigl(\Bigl\{\sqrt{n}\bigl(\widehat{\mathsf{P}}_n-\mathsf{P}\bigr)g>x\Bigr\}\bigcup\Bigl\{\sqrt{n}\bigl(\widehat{\mathsf{P}}_n-\mathsf{P}\bigr)g<-x\Bigr\}\Bigr)\\
&\quad\le\mathsf{Pr}\Bigl(\sqrt{n}\bigl(\widehat{\mathsf{P}}_n-\mathsf{P}\bigr)g>x\Bigr)+\mathsf{Pr}\Bigl(\sqrt{n}\bigl(\widehat{\mathsf{P}}_n-\mathsf{P}\bigr)g<-x\Bigr)\\
&\quad\le 2\exp\Bigl(-\frac{1}{4}\frac{x^2}{\mathsf{P}g^2+x\|g\|_\infty/\sqrt{n}}\Bigr)
\end{aligned}
$$

がわかる．

① (6.62) の証明. Markov の不等式 (命題 1.33) から, 任意の $\lambda>0$ に対して

$$
\mathsf{Pr}\Bigl(\sqrt{n}\bigl(\widehat{\mathsf{P}}_n-\mathsf{P}\bigr)g>x\Bigr)\le e^{-\lambda x}\Bigl\{1+\frac{1}{n}\sum_{k=2}^{\infty}\frac{1}{k!}\frac{\lambda^k}{n^{k/2-1}}\mathsf{P}\bigl[\{g-\mathsf{P}g\}^k\bigr]\Bigr\}^n \tag{6.64}
$$

と評価できること[29] がわかる．$\lambda>0$ は任意だったので

$$
\lambda=\frac{1}{2}\frac{x}{\mathsf{P}g^2+x\|g\|_\infty/\sqrt{n}}\le\min\Bigl(\frac{x}{2\,\mathsf{P}g^2},\,\frac{\sqrt{n}}{2\,\|g\|_\infty}\Bigr)=:\min\bigl(\lambda_1,\,\lambda_2\bigr)
$$

[29]実際,

$$
\begin{aligned}
&\mathsf{Pr}\Bigl(\sqrt{n}\bigl(\widehat{\mathsf{P}}_n-\mathsf{P}\bigr)g>x\Bigr)\\
&\quad=\mathsf{Pr}\Bigl(\exp\{\lambda(\sqrt{n}(\widehat{\mathsf{P}}_n-\mathsf{P})g)\}>e^{\lambda x}\Bigr)\le e^{-\lambda x}\mathsf{E}\Bigl[\exp\{\lambda(\sqrt{n}(\widehat{\mathsf{P}}_n-\mathsf{P})g)\Bigr]\\
&\quad=e^{-\lambda x}\mathsf{E}\Bigl[\exp\Bigl\{\frac{\lambda}{\sqrt{n}}\sum_{j=1}^{n}\bigl(g(X_j)-\mathsf{P}g\bigr)\Bigr\}\Bigr]=e^{-\lambda x}\mathsf{E}\Bigl[\prod_{j=1}^{n}\exp\Bigl\{\frac{\lambda}{\sqrt{n}}\bigl(g(X_j)-\mathsf{P}g\bigr)\Bigr\}\Bigr]\\
&\quad=e^{-\lambda x}\prod_{j=1}^{n}\mathsf{E}\Bigl[\exp\Bigl\{\frac{\lambda}{\sqrt{n}}\bigl(g(X_j)-\mathsf{P}g\bigr)\Bigr\}\Bigr]\Bigr]=e^{-\lambda x}\Bigl\{\mathsf{E}\Bigl[\exp\Bigl\{\frac{\lambda}{\sqrt{n}}\bigl(g(X1)-\mathsf{P}g\bigr)\Bigr\}\Bigr]\Bigr\}^n\\
&\quad=e^{-\lambda x}\Bigl\{\mathsf{E}\Bigl[1+\sum_{k=1}^{\infty}\frac{1}{k!}\Bigl(\frac{\lambda}{\sqrt{n}}\Bigr)^k\{g(X_1)-\mathsf{P}g\}^k\Bigr]\Bigr\}^n=e^{-\lambda x}\Bigl\{\mathsf{E}\Bigl[\exp\Bigl\{\frac{\lambda}{\sqrt{n}}\bigl(g(X_1)-\mathsf{P}g\bigr)\Bigr\}\Bigr]\Bigr\}^n\\
&\quad=e^{-\lambda x}\Bigl\{1+\sum_{k=1}^{\infty}\frac{1}{k!}\Bigl(\frac{\lambda}{\sqrt{n}}\Bigr)^k\mathsf{E}\bigl[\{g(X_1)-\mathsf{P}g\}^k\bigr]\Bigr\}^n=e^{-\lambda x}\Bigl\{1+\sum_{k=1}^{\infty}\frac{1}{k!}\Bigl(\frac{\lambda}{\sqrt{n}}\Bigr)^k\mathsf{P}\bigl[\{g-\mathsf{P}g\}^k\bigr]\Bigr\}^n\\
&\quad=e^{-\lambda x}\Bigl\{1+\sum_{k=1}^{\infty}\frac{1}{k!}\Bigl(\frac{\lambda}{\sqrt{n}}\Bigr)^k\mathsf{E}\bigl[\{g(X_1)-\mathsf{P}g\}^k\bigr]\Bigr\}^n=e^{-\lambda x}\Bigl\{1+\sum_{k=2}^{\infty}\frac{1}{k!}\Bigl(\frac{\lambda}{\sqrt{n}}\Bigr)^k\mathsf{P}\bigl[\{g-\mathsf{P}g\}^k\bigr]\Bigr\}^n\\
&\quad=e^{-\lambda x}\Bigl\{1+\sum_{k=1}^{\infty}\frac{1}{k!}\Bigl(\frac{\lambda}{\sqrt{n}}\Bigr)^k\mathsf{E}\bigl[\{g(X_1)-\mathsf{P}g\}^k\bigr]\Bigr\}^n=e^{-\lambda x}\Bigl\{1+\frac{1}{n}\sum_{k=2}^{\infty}\frac{1}{k!}\frac{\lambda^k}{n^{k/2-1}}\mathsf{P}\bigl[\{g-\mathsf{P}g\}^k\bigr]\Bigr\}^n
\end{aligned}
$$

からわかる．

とおく. $k \geq 2$ に対して

$$\lambda^k \leq \lambda_1 \lambda_2^{k-2} \lambda$$

と

$$\mathsf{P}\big[\big\{g - \mathsf{P}g\big\}^k\big] \leq \mathsf{P}g^2 \big(2\|g\|_\infty\big)^{k-2}$$

に注意して, 上の不等式の最右辺の項を評価すると

$$\begin{aligned}
&\frac{1}{n}\sum_{k=2}^{\infty}\frac{1}{k!}\frac{\lambda^k}{n^{k/2-1}}\mathsf{P}\big[\{g - \mathsf{P}g\}^k\big] \\
&\quad \leq \frac{1}{n}\sum_{k=2}^{\infty}\frac{1}{k!}\frac{\lambda_1\lambda_2^{k-2}\lambda}{n^{k/2-1}}\mathsf{P}g^2\big(2\|g\|_\infty\big)^{k-2} \\
&\quad = \frac{1}{n}\sum_{k=2}^{\infty}\frac{1}{k!}\left(\frac{x}{2\,\mathsf{P}g^2}\right)\left(\frac{\sqrt{n}}{2\,\|g\|_\infty}\right)^{k-2}\frac{\lambda}{n^{k/2-1}}\mathsf{P}g^2\big(2\|g\|_\infty\big)^{k-2} \\
&\quad = \frac{1}{n}\sum_{k=2}^{\infty}\frac{1}{k!}\left(\frac{x}{2\,\mathsf{P}g^2}\right)\left(\frac{\sqrt{n}}{2\,\|g\|_\infty}\right)^{k-2}\frac{\lambda}{n^{k/2-1}}\mathsf{P}g^2\big(2\|g\|_\infty\big)^{k-2} \\
&\quad = \frac{1}{n}\sum_{k=2}^{\infty}\frac{1}{k!}\frac{1}{2}\lambda x \\
&\quad = \frac{\lambda x(e-2)}{2n} \\
&\quad \leq \frac{\lambda x}{2n}
\end{aligned} \tag{6.65}$$

を得る. 最後から 2 番目の等号は

$$e = \sum_{k=0}^{\infty}\frac{1}{k!} = 2 + \sum_{k=2}^{\infty}\frac{1}{k!}$$

からわかる. (6.65) を (6.64) に代入して, $a > 0$ に対して, 不等式 $(1 + a)^n \leq e^{an}$ が成り立つことに注意すると

$$\begin{aligned}
\Pr\Big(\sqrt{n}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)g > x\Big) &\leq e^{-\lambda x}\left(1 + \frac{\lambda x}{2n}\right)^n \leq e^{-\lambda x}e^{\lambda x/2} = e^{-\lambda x/2} \\
&= \exp\left(-\frac{1}{2}\frac{x^2}{\mathsf{P}g^2 + x\|g\|_\infty/\sqrt{n}}\right)
\end{aligned}$$

を得る. よって, (6.62) が示せた. □

補題 6.54. $\mathcal{G} \subset L_2(\mathsf{P})$ を有界な可測関数の有限集合とし, $\#(\mathcal{G})$ を有限集合の要素の数とする. ただし, 要素の個数は少なくとも 2 以上とする.

このとき

$$
\begin{aligned}
&\mathsf{E}\left[\max_{g\in\mathcal{G}}\left|\sqrt{n}\big(\widehat{\mathsf{P}}_n-\mathsf{P}\big)g\right|\right]\\
&\qquad\le 48\left\{\max_{g\in\mathcal{G}}\frac{\|g\|_\infty}{\sqrt{n}}\log\big(1+\#(\mathcal{G})\big)+\max_{g\in\mathcal{G}}\|g\|_{L_2(\mathsf{P})}\sqrt{\log\#(\mathcal{G})}\right\} &(6.66)\\
&\qquad\le 96\left\{\max_{g\in\mathcal{G}}\frac{\|g\|_\infty}{\sqrt{n}}\log\big(1+\#(\mathcal{G})\big)+\max_{g\in\mathcal{G}}\|g\|_{L_2(\mathsf{P})}\sqrt{\log\#(\mathcal{G})}\right\} &(6.67)
\end{aligned}
$$

が成立する.

Proof. 関数 $g\in\mathcal{G}$ を固定する. 関数 g を乗数倍しても (6.66) には影響しないので, 一般性を失うことなく, $\sqrt{n\mathsf{P}g^2}\ge 1$ と仮定してよい. $a:=24\|g\|_\infty/\sqrt{n}$ と $b:=24\mathsf{P}g^2$ とおく. すると

$$
c:=\frac{b}{a}=\frac{\sqrt{n}\mathsf{P}g^2}{\|g\|_\infty}\ge\frac{\sqrt{n}\mathsf{P}g^2}{\mathsf{P}|g|}\ge\frac{\sqrt{n}\mathsf{P}g^2}{\sqrt{\mathsf{P}g^2}}=\sqrt{n\mathsf{P}g^2}\ge 1
$$

となる. すると

$$
x>\frac{b}{a}\ \Rightarrow\ \frac{1}{4}\frac{x^2}{\mathsf{P}g^2+x\|g\|_\infty/\sqrt{n}}\ge\frac{3x}{a} \tag{6.68}
$$

となる. 実際

$$
\begin{aligned}
x>\frac{b}{a}=\frac{\mathsf{P}g^2}{\|g\|_\infty/\sqrt{n}}\ \Rightarrow&\frac{1}{4}\frac{x^2}{\mathsf{P}g^2+x\|g\|_\infty/\sqrt{n}}=\frac{1}{4}\frac{x^2}{\frac{24\{\mathsf{P}g^2+x\|g\|_\infty/\sqrt{n}\}}{24}}\\
&=6\frac{x^2}{a(b/a+1)}\ge 6\frac{x(b/a)}{a(b/a+1)}\ge\frac{3x}{a}
\end{aligned}
$$

からわかる. 最後の不等式は

$$
\frac{b/a}{b/a+1}=\frac{c}{1+c}\ge\frac{1}{2}\Leftrightarrow c-1\ge 0
$$

よりわかる. また, $c\ge 1$ から $a+b\le 2b$ となるので

$$
x\le\frac{b}{a}\ \Rightarrow\ \frac{1}{4}\frac{x^2}{\frac{24\mathsf{P}g^2+24x\|g\|_\infty/\sqrt{n}}{24}}=6\frac{x^2}{a+b}\ge 3\frac{x^2}{b} \tag{6.69}
$$

がわかる. $\nu_n=\sqrt{n}\big(\widehat{\mathsf{P}}_n-\mathsf{P}\big)$ とおき, $\nu_n\quad g$ の打ち切り変数

$$
\begin{aligned}
A_g&:=\nu_n g\mathbb{1}\{|\nu_n g|>b/a\},\\
B_g&:=\nu_n g\mathbb{1}\{|\nu_n g|\le b/a\}
\end{aligned}
$$

とおく. Bernstein の不等式を用いて, (6.68) を用いると, $x > 0$ に対して

$$\begin{aligned}\mathsf{Pr}\Big(|A_g| > x\Big) &= \mathsf{Pr}\Big(|\nu_n g| > x \text{ かつ } x > b/a\Big)\\ &\le 2\exp\Big(-\frac{3x}{a}\Big) \end{aligned}\tag{6.70}$$

を得る. 同様に, Bernstein の不等式を適用してから, (6.69) を用いる. すると, $x > 0$ に対して

$$\begin{aligned}\mathsf{Pr}\Big(|B_g| > x\Big) &= \mathsf{Pr}\Big(|\nu_n g| > x \text{ かつ } x \le b/a\Big)\\ &\le 2\exp\Big(-\frac{3x^2}{b}\Big) \end{aligned}\tag{6.71}$$

を得る. ここで, $p = 1, 2$ に対して, $\Psi_p(x) = e^{x^p} - 1\,(x \ge 0)$ とおく. 関数 Ψ_p は凸であることに注意する. Funibi の定理と (6.70) を用いると

$$\begin{aligned}\mathsf{E}\Big[\Psi_1\Big(\frac{|A_g|}{a}\Big)\Big] &= \mathsf{E}\Big[\int_0^{|A_g|/a} e^x\,\mathrm{d}x\Big]\\ &= \int_0^\infty \mathsf{Pr}\big(|A_g| > xa\big)e^x\,\mathrm{d}x\\ &\le 2\int_0^\infty e^{-3x}e^x\,\mathrm{d}x = 1 \end{aligned}\tag{6.72}$$

となる. Ψ_1 は凸関数なので Jensen の不等式を用い, さらに (6.72) を使うと

$$\begin{aligned}\Psi_1\Big(\mathsf{E}\Big[\max_{g\in\mathcal{G}}\frac{|A_g|}{a}\Big]\Big) &= \mathsf{E}\Big[\Psi_1\Big(\max_{g\in\mathcal{G}}\frac{|A_g|}{a}\Big)\Big]\\ &\le \mathsf{E}\Big[\sum_{g\in\mathcal{G}}\Psi_1\Big(\frac{|A_g|}{a}\Big)\Big]\\ &\le \#(\mathcal{G}) \end{aligned}$$

を得る. 次に, Ψ_1 の逆関数は $\Psi_1^{-1}(u) = \log(1+u)\,(u > -1)$ であり, Ψ_1^{-1} は単調増加関数であることに注意すると注意すると

$$\begin{aligned}\frac{\mathsf{E}\big[\max_{g\in\mathcal{G}}|A_g|\big]}{24\max\limits_{g\in\mathcal{G}}\|g\|_\infty/\sqrt{n}} &\le \mathsf{E}\Big[\max_{g\in\mathcal{G}}\frac{|A_g|}{a}\Big]\\ &= \Psi_1^{-1}\Big(\Psi_1\Big(\mathsf{E}\Big[\max_{g\in\mathcal{G}}\frac{|A_g|}{a}\Big]\Big)\Big)\\ &\le \Psi_1^{-1}\big(\#(\mathcal{G})\big) \end{aligned}$$

がわかる. したがって

$$\mathsf{E}\Big[\max_{g\in\mathcal{G}}|A_g|\Big] \le 24\max_{g\in\mathcal{G}}\frac{\|g\|_\infty}{\sqrt{n}}\log(1+\#(\mathcal{G})) \tag{6.73}$$

を得る. 同様に, Funibi の定理と (6.72) を用いると

$$\begin{aligned}\mathsf{E}\bigg[\Psi_2\bigg(\frac{|B_g|}{\sqrt{b}}\bigg)\bigg] &= \mathsf{E}\bigg[\int_0^{|B_g|/\sqrt{b}} e^{x^2}\,\mathrm{d}x\bigg] \\ &= \int_0^\infty \mathsf{Pr}\big(|B_g| > x\sqrt{b}\big)e^{x^2}\,\mathrm{d}x \\ &\le 2\int_0^\infty e^{-3x^2}e^{x^2}\,\mathrm{d}x = \sqrt{\frac{\pi}{2}} \le 2\end{aligned}$$

となる. Ψ_2 の逆関数は $\Psi_2^{-1}(u) = \sqrt{\log(1+u)}\,(u > -1)$ であり, Ψ_2^{-1} も単調増加関数であることに注意すると

$$\begin{aligned}\frac{\mathsf{E}\big[\max_{g\in\mathcal{G}}|B_g|\big]}{24\max_{g\in\mathcal{G}}\sqrt{\mathsf{P}g^2}} &\le \mathsf{E}\bigg[\max_{g\in\mathcal{G}}\frac{|B_g|}{\sqrt{b}}\bigg] \\ &= \Psi_2^{-1}\bigg(\Psi_2\bigg(\mathsf{E}\bigg[\max_{g\in\mathcal{G}}\frac{|B_g|}{\sqrt{b}}\bigg]\bigg)\bigg) \\ &\le \Psi_2^{-1}\big(2\#(\mathcal{G})\big)\end{aligned}$$

がわかる. したがって

$$\begin{aligned}\mathsf{E}\Big[\max_{g\in\mathcal{G}}|B_g|\Big] &\le 24\max_{g\in\mathcal{G}}\|g\|_{L_2(\mathsf{P})}\sqrt{\log(1+2\#(\mathcal{G}))} \\ &\le 24\sqrt{2}\max_{g\in\mathcal{G}}\|g\|_{L_2(\mathsf{P})}\sqrt{\log(1+\#(\mathcal{G}))} \qquad (6.74)\end{aligned}$$

を得る. 最後に (6.73) と (6.74) を合わせると

$$\begin{aligned}&\mathsf{E}\Big[\max_{g\in\mathcal{G}}|\nu_n|\Big] \\ &\quad\le \mathsf{E}\Big[\max_{g\in\mathcal{G}}|A_g|\Big] + \mathsf{E}\Big[\max_{g\in\mathcal{G}}|B_g|\Big] \\ &\quad\le 24\max_{g\in\mathcal{G}}\frac{\|g\|_\infty}{\sqrt{n}}\log(1+\#(\mathcal{G})) + 24\sqrt{2}\max_{g\in\mathcal{G}}\|g\|_{L_2(\mathsf{P})}\sqrt{\log(1+\#(\mathcal{G}))} \\ &\quad\le 48\bigg\{\max_{g\in\mathcal{G}}\frac{\|g\|_\infty}{\sqrt{n}}\log(1+\#(\mathcal{G})) + \max_{g\in\mathcal{G}}\|g\|_{L_2(\mathsf{P})}\sqrt{\log(1+\#(\mathcal{G}))}\bigg\}\end{aligned}$$

を得る.

最後に, $x \ge 2$ に対して

$$x^2 - (1+x) = \bigg(x - \frac{1}{2}\bigg)^2 - \frac{4}{5} \ge \frac{9}{4} - \frac{5}{4} = 1$$

となるので

$$\log(1+x) \le 2\log x$$

である. また

$$\sqrt{\log(1+x)} \le \frac{\log 3}{\log 2}\sqrt{\log x} \le 2\sqrt{\log x}$$

に注意すればよい. □

補題 6.55. P を $\mathbb{X} \subset \mathbb{R}$ 上の確率測度とし, $X_1, X_2, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P}$ とする. $\delta > 0$ とし, $\mathcal{G} = \{g : \mathbb{X} \longrightarrow \mathbb{R};\, \mathsf{P}g^2 < \delta^2\}$ とする.

$$\begin{aligned} G(x) &:= \sup_{g\in\mathcal{G}} |g(x)| \qquad (x \in \mathbb{X}), \\ a(\delta) &:= \frac{\delta}{\sqrt{\max\bigl\{1,\, \log N_B\bigl(\delta,\, \mathcal{G},\, \|\cdot\|_{L_2(\mathsf{P})}\bigr)\bigr\}}} \end{aligned}$$

とおいたとき

$$\begin{aligned} \mathsf{E}\Biggl[\sup_{g\in\mathcal{G}}\Bigl|\sqrt{n}\bigl(\widehat{\mathsf{P}}_n - \mathsf{P}\bigr)g\Bigr|\Biggr] \le 382\int_0^\delta \sqrt{\log N_B\bigl(\epsilon,\, \mathcal{G},\, \|\cdot\|_{L_2(\mathsf{P})}\bigr)}\,\mathrm{d}\epsilon \\ + 2\sqrt{n}\mathsf{P}G\mathbb{1}\{G > \sqrt{n}a(\delta)\} \end{aligned} \tag{6.75}$$

が成り立つ.

Proof. $\nu_n = \sqrt{n}\bigl(\widehat{\mathsf{P}}_n - \mathsf{P}\bigr)$ とおく. 関数 $G : \mathbb{X} \longrightarrow \mathbb{R}$ を関数族 $\mathcal{G}$ の封筒関数とする. すると, 任意の $g \in \mathcal{G}$ に対して, $|g(x)| \le G(x)\,(x \in \mathbb{X})$ となる. よって

$$\bigl|\nu_n g\bigr| \le \bigl|\sqrt{n}\bigl(\widehat{\mathsf{P}}_n - \mathsf{P}\bigr)g\bigr| \le \sqrt{n}\bigl(\widehat{\mathsf{P}}_n + \mathsf{P}\bigr)G$$

が成立することがわかる. このことから

$$\begin{aligned} \mathsf{E}\Biggl[\sup_{g\in\mathcal{G}}\bigl|\nu_n g\mathbb{1}\{G > \sqrt{n}a(\delta)\}\bigr|\Biggr] &\le \mathsf{E}\Bigl[\sqrt{n}\bigl(\widehat{\mathsf{P}}_n + \mathsf{P}\bigr)G\mathbb{1}\{G > \sqrt{n}a(\delta)\}\Bigr] \\ &= 2\sqrt{n}\mathsf{P}G\mathbb{1}\{G > \sqrt{n}a(\delta)\} \end{aligned}$$

となる. 上の不等式の右辺は補題の不等式の上限の第 2 項目の 2 倍である. よって

$$\mathsf{E}\Biggl[\sup_{g\in\mathcal{G}}\bigl|\nu_n g\mathbb{1}\{G \le \sqrt{n}a(\delta)\}\bigr|\Biggr] \tag{6.76}$$

を評価すればよい.

$$\widetilde{\mathcal{G}} := \left\{ g\mathbb{1}\{G \le \sqrt{n}a(\delta)\};\, g \in \mathcal{G} \right\}$$

とおいたとき, $\epsilon > 0$ に対して

$$N_B\big(\epsilon, \widehat{\mathcal{G}}, \|\cdot\|_{L_2(\mathsf{P})}\big) \le N_B\big(\epsilon, \mathcal{G}, \|\cdot\|_{L_2(\mathsf{P})}\big)$$

となるので, 一般性を失うことなく

$$|g| \le \sqrt{n}a(\delta) \qquad (g \in \mathcal{G}) \tag{6.77}$$

をみたいしていると仮定してよい.

I. chaining argument のための分割の構成. 固定した $\delta > 0$ に対して, 整数 q_0 を

$$4\delta \le 2^{-q_0} \le 8\delta \tag{6.78}$$

をみたすように取る. 整数 $q\,(\ge q_0)$ で添え字つけられた関数族 $\mathcal{G}$ の入れ子上の分割の列 $\{\mathcal{G}_{qj}\}_{j=1}^{N_q}\,(N_q \in \mathbb{N})$ と可測関数 $\Delta_{qj} : \mathbb{X} \longrightarrow \mathbb{R}$ で $\Delta_{qj} \le 2G$ で, 以下の条件をみたすものを取る.

- $\mathcal{G} = \bigcup_{j=1}^{N_q} \mathcal{G}_{qj}$.
- $\sum_{q \ge q_0} \dfrac{1}{2^q}\sqrt{\max\{1, \log Nq\}} \lesssim \displaystyle\int_0^\delta \sqrt{\max\big\{1, \log N_B\big(\epsilon, \mathcal{G}, \|\cdot\|_{L_2(\mathsf{P})}\big)\big\}}\,\mathrm{d}\epsilon$.
- $\displaystyle\sup_{g_1, g_2 \in \mathcal{G}_{qj}} |g_1 - g_2| \le \Delta_{qj}\,(j = 1, 2, \ldots, N_q;\, q \ge q_0)$.
- $\mathsf{P}\Delta_{qj}^2 \le \dfrac{1}{2^{2q}}\,(j = 1, 2, \ldots, N_q;\, q \ge q_0)$.

実際, 上記のような入れ子状の分割を次のステップを経ることで作ることができる. まず, 半径 2^{-q} の $L_2(\mathsf{P})$-brackets の最小数で関数族 $\mathcal{G}$ を覆う. この bracket から前の段階の bracket を引いたもので, この brackets を置き換える. このことで, 上の関数と下の関数の差が Δ_{qj} に等しく, $Delta_{jq}$ の条件をみたした分割を与えることになる. 仮にこの分割の列が順次細分化した分割を構成できない場合には, $\bigcap_{p=q_0}^{q} \mathcal{G}_{p, j_p}$ の集合で q 段階目の分割を置き換えればよい. これらの手順から $\overline{N}_q = N_{q_0} \cdots N_q$ 個の集合に分割できる. 不等式

$$\left(\log \prod_{p=q_0}^{q} N_p\right)^{1/2} \le \sum_{p=q_0}^{q} \big(\log N_p\big)^{1/2}$$

を用いると 2 番目と 3 番目の条件はみたされることがわかる.

II. chain の構成. 各 $q \geq q_0$ に対して, 分割の集合 $\mathcal{G}_{qj}$ から一つの関数 g_{qj} を固定し

$$\Pi_q g = g_{qj}, \quad \Delta_q g = \Delta_{qj} (g \in \mathcal{G}_{qj} \text{ の場合})$$

と定める. 関数 g が関数族 $\mathcal{G}$ を動くとき, $\Pi_q g$ と $\Delta_q g$ は N_q 個の関数の集合を動く. 固定した n と $q \geq q_0$ に対して, 以下の指示関数を定義する.

$$\begin{aligned}
a_q &= \frac{1}{2^q} \Big/ \sqrt{\max\{1, \log N_{q+1}\}}, \\
A_{q-1} g &= \mathbb{1}\{\Delta_{q_0} g \leq \sqrt{n} a_{q_0}, \dots, \Delta_{q-1} g \leq \sqrt{n} a_{q-1}\}, \\
B_q g &= \mathbb{1}\{\Delta_{q_0} g \leq \sqrt{n} a_{q_0}, \dots, \Delta_{q-q} g \leq \sqrt{n} a_{q-1}, \Delta_q g > \sqrt{n} a_q\}.
\end{aligned}$$

分割は入れ子状になっているので, q 段階において, 各分割の集合 $\mathcal{G}_{qj}$ 上の関数 g に対して, $A_q g$ と $\Delta_q g$ は定数となる. 分割の作成の仕方と q_0 の取り方から

$$\begin{aligned}
2a(\delta) &= \frac{2\delta}{\sqrt{\max\{1, \log N_B(\delta, \mathcal{G}, \|\cdot\|_{L_2(\mathsf{P})})\}}} \\
&\leq \frac{2^{-(q_0+1)}}{\sqrt{\max\{1, \log N_{q_0+1}\}}} \\
&\leq \frac{2^{-q_0}}{\sqrt{\max\{1, \log N_{q_0}\}}} = a_{q_0}
\end{aligned}$$

となっている. したがって

$$\Delta_{q_0} g = \Delta_{qj} \leq 2G \leq 2\sqrt{n} a(\delta) \leq \sqrt{n} a_{q_0}$$

となるので

$$A_{q_0} g = \mathbb{1}\{\Delta_{q_0} \leq \sqrt{n} a_{q_0}\} = 1$$

となる. よって, $B_{q_0} g = \mathbb{1}\{\Delta_{q_0} \sqrt{n} a_{q_0}\} = 0$ となる. もし, $q_1 > q_0$ があって, $B_{q_1} g = 1$ となると $\Delta_{q_1} g > \sqrt{n} a_{q_1}$ となる. $q > q_1$ に対して, $B_q g = 1$ となるためには, $\Delta_{q_1} g \leq \sqrt{n} a_{q_1}$ となければならないが, $B_{q_1} g = 1$ のときはそうならないので, $B_q g = 0$ となる. 以上の議論から, 以下のふたつの場合だけが起こる.

① すべての $q \geq q_0$ に対して, $B_q g = 0$.

② ある $q_1 > q_0$ が存在して, $B_{q_1} g = 1$ でその他は $B_q g = 0$.

① の場合, すべての $q \geq q_0$ に対して, $A_q g = 1$ となる. ② の場合には, $\Delta_{q_1} g > \sqrt{n} a_{q_1}$ となるので,

$$A_q g = \begin{cases} 1 & (q < q_1) \\ 0 & (q \geq q_1) \end{cases}$$

となる. いま, 各 $x \in \mathbb{X}$ に対して, $q_1 = q_1(g, x)$ は $\Pi_1 g - \Pi_{p-1} g$ の上限 $\Delta_p g$ が一様に $\sqrt{n} a(\delta)$ より小さくなる最小の数である. $q_1 = \infty$ の場合もある. すると

$$g - \Pi_{q_0} g = \sum_{p=q_0+1}^{\infty} \bigl(g - \Pi_p g\bigr) B_p g + \sum_{p=q_0+1}^{\infty} \bigl(\Pi_p g - \Pi_{p-1} g\bigr) A_{p-1} g$$

と分解できる. 実際

$$\begin{aligned}
&\sum_{p=q_0+1}^{\infty} \bigl(g - \Pi_p g\bigr) B_p g + \sum_{p=q_0+1}^{\infty} \bigl(\Pi_p g - \Pi_{p-1} g\bigr) A_{p-1} g \\
&\quad = \sum_{p=q_0+1}^{q_1-1} \bigl(g - \Pi_p g\bigr) \underbrace{B_p g}_{=0} + \bigl(g - \Pi_{q_1} g\bigr) \underbrace{B_{q_1} g}_{=1} + \sum_{p=q_1+1}^{\infty} \bigl(g - \Pi_p g\bigr) \underbrace{B_p g}_{=0} \\
&\qquad \sum_{p=q_0+1}^{q_1} \bigl(\Pi_p g - \Pi_{p-1} g\bigr) \underbrace{A_{p-1} g}_{=1} + \sum_{p=q_1+1}^{\infty} \bigl(\Pi_p g - \Pi_{p-1} g\bigr) \underbrace{A_{p-1} g}_{=0} \\
&\quad = \bigl(g - \Pi_{q_1} g\bigr) + \sum_{p=q_0+1}^{q_1} \bigl(\Pi_p g - \Pi_{p-1} g\bigr) \\
&\quad = g - \Pi_{q_0} g
\end{aligned}$$

からわかる.

III. $\nu_n\bigl(\sum_{p=q_0+1}^{\infty} \bigl(g - \Pi_p g\bigr) B_p g\bigr)$ **の評価**. $g_1, g_1 \in \mathcal{G}_{qj} | g_1 - g_2 | \leq \Delta_{jq}$ だったので

$$\bigl| g - \Pi_p g \bigr| \leq \Delta_{qj} = \Delta p g$$

である. さらに, 分割は入れ子状だったので, $B_q g = 1$ ならば, $B_q g$ の定義から, $\Delta_{q-1} g \leq \sqrt{n} a_{q-1}$ となる. よって

$$\Delta_p g B_p g \leq \Delta_{p-1} g B_q g \leq \sqrt{n} a_{p-1} \tag{6.79}$$

となる. また, $B_q g = 0$ ならば, 明らかに上記の不等式は成立する. $\mathsf{P}\Delta_{qj}^2 \leq \dfrac{1}{2^{2q}}$ だったので

$$\mathsf{P}\bigl(\Delta_q g\bigr)^2 B_q g \leq \frac{1}{2^{2q}} \tag{6.80}$$

である. 任意の $|g| \le h$ なる組に対して

$$\big|\nu_n g\big| \le \nu_n h + 2\sqrt{n}\mathsf{P}h \tag{6.81}$$

となる. なぜならば

$$\begin{aligned}\big|\nu_n g\big| &= \big|\sqrt{n}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)g\big| \\ &\le \sqrt{n}\big|\widehat{\mathsf{P}}_n g\big| + \sqrt{n}\big|\mathsf{P}g\big| \\ &\le \sqrt{n}\widehat{\mathsf{P}}_n|g| + \sqrt{n}\mathsf{P}|g| \\ &\le \sqrt{n}\widehat{\mathsf{P}}_n h + \sqrt{n}\mathsf{P}h \\ &\le \sqrt{n}\big(\widehat{\mathsf{P}} - \mathsf{P}\big)h + 2\sqrt{n}\mathsf{P}h\end{aligned}$$

からわかる. $\big|(g - \Pi_q g)B_q g\big| \le \Delta_q g B_q g$ と (6.81) から

$$\begin{aligned}\big|\nu_n(g - \Pi_q g)B_q g\big| &\le \nu_n \Delta_q g B_q g + 2\sqrt{n}\mathsf{P}\Delta_q g B_q G \\ &\le \big|\nu_n \Delta_q g B_q g\big| + 2\sqrt{n}\big|\mathsf{P}\Delta_q g B_q\big| \end{aligned} \tag{6.82}$$

となる. (6.82) と補題 6.54 から

$$\begin{aligned}
&\mathsf{E}\left[\sup_{g\in\mathcal{G}}\left|\sum_{q=q_0+1}^{\infty}\nu_n(g - \Pi_q g)B_q g\right|\right] \\
&\quad\le \mathsf{E}\left[\sum_{q=q_0+1}^{\infty}\sup_{g\in\mathcal{G}}\left|\nu_n(g - \Pi_q g)B_q g\right|\right] \\
&\quad= \sum_{q=q_0+1}^{\infty}\mathsf{E}\left[\sup_{g\in\mathcal{G}}\left|\nu_n(g - \Pi_q g)B_q g\right|\right] \\
&\quad\le \sum_{q=q_0+1}^{\infty}\mathsf{E}\left[\sup_{g\in\mathcal{G}}\left\{\big|\nu_n\Delta_q g B_q g\big| + 2\sqrt{n}\big|\mathsf{P}\Delta_q g B_q\big|\right\}\right] \qquad \big(\because (6.82)\big) \\
&\quad\le \sum_{q=q_0+1}^{\infty}\mathsf{E}\left[\sup_{g\in\mathcal{G}}\big|\nu_n\Delta_q g B_q g\big| + 2\sqrt{n}\sup_{g\in\mathcal{G}}\big|\mathsf{P}\Delta_q g B_q\big|\right] \\
&\quad\le \sum_{q=q_0+1}^{\infty}\mathsf{E}\Bigg[96\max\frac{\|\Delta_q g B_q g\|_\infty}{\sqrt{n}}\log N_q \\
&\qquad\qquad + 96\max\|\Delta_q g B_q g\|_{L_2(\mathsf{P})}\sqrt{\log N_q} + 2\sqrt{n}\sup_{g\in\mathcal{G}}\big|\mathsf{P}\Delta_q g B_q g\big|\Bigg] \\
&\qquad\qquad \big(\because \text{補題 } 6.54\big) \\
&\quad\le \sum_{q=q_0+1}^{\infty}\left\{96 a_{q-1}\log N_q + \frac{96}{2^q}\sqrt{\log N_q} + \frac{4}{a_q}2^{-2q}\right\} \qquad \big(\because (6.79), (6.82)\big) \\
&\quad\le 194\sum_{q=q_0+1}^{\infty}\frac{\sqrt{\log(1+N_q)}}{2^q}
\end{aligned}$$

を得る．

IV. $\nu_n\big(\sum_{p=q_0+1}^{\infty}\big(\Pi_p g-\Pi_{p-1}g\big)A_{p-1}g\big)$ **の評価**．分割は入れ子状なので，N_q 個の $\{\Pi_q g-\Pi_{q-1}g;\ g\in\mathcal{G}\}$ は最大でも N_q 個の要素を持つ．また，$\{A_{q-1}g;\ g\in\mathcal{G}\}$ は N_{q-1} 個の要素をもつ．さらに，$A_{q-1}g$ の作り方から

$$A_{q-1}g=1\ \Rightarrow\ \Delta_{q-1}g\le\sqrt{n}a_{q-1}$$

なので

$$\big|\Pi_q g-\Pi_{q-1}g\big|A_{q-1}g\le\Delta_{q-1}A_{q-1}g\le\sqrt{n}a_{q-1}$$

となる．また，$\sup_{g_1,\,g_2\in\mathcal{G}_{qj}}|g_1-g_2|\le\Delta_{qj}$ かつ $\mathsf{P}\Delta_{qj}^2\le\frac{1}{2^{2q}}$ なので

$$\sup_{g\in\mathcal{G}}\big|\Pi_q g-\Pi_{q-1}g\big|\le\sup_{g_1,\,g_2\in\mathcal{G}_{qj}}|g_1-g_2|\le\frac{1}{2^{2q}}$$

となる．よって

$$\big\|\Pi_q g-\Pi_{q-1}g\big\|_{L_2(\mathsf{P})}\le\frac{1}{2^{2q}}\le\frac{1}{2^{q-1}}$$

がわかる．補題 6.54 を用いると

$$\begin{aligned}
&\mathsf{E}\Big[\sup_{g\in\mathcal{G}}\Big|\nu_n\big(\Pi_q g-\Pi_{q-1}g\big)A_{q-1}g\Big|\Big]\\
&\quad\le 96\max_g\frac{\|\Pi_q g-\Pi_{q-1}g\|_\infty}{\sqrt{n}}\log N_q\\
&\qquad+96\max_g\big\|\Pi_q g-\Pi_{q-1}g\big\|_{L_2(\mathsf{P})}\sqrt{\log N_q}\\
&\quad\le 96a_{q-1}\log N_q+\frac{96}{2^q}\sqrt{\log N_q}\\
&\quad\le 192\frac{\sqrt{\log N_q}}{2^q}
\end{aligned}$$

を得る．よって

$$\begin{aligned}
&\mathsf{E}\Big[\sup_{g\in\mathcal{G}}\Big|\nu_n\Big(\sum_{q=q_0+1}^{\infty}\big(\Pi_q g-\Pi_{q-1}g\big)A_{q-1}g\Big)\Big|\Big]\\
&\quad\le\mathsf{E}\Big[\sum_{q=q_0+1}^{\infty}\sup_{g\in\mathcal{G}}\Big|\nu_n\big(\Pi_q g-\Pi_{q-1}g\big)A_{q-1}g\big)\Big|\Big]\\
&\quad\le\sum_{q=q_0+1}^{\infty}\mathsf{E}\Big[\sup_{g\in\mathcal{G}}\Big|\nu_n\big(\Pi_q g-\Pi_{q-1}g\big)A_{q-1}g\big)\Big|\Big]\\
&\quad\le 192\sum_{q=q_0+1}^{\infty}\frac{\sqrt{\log N_q}}{2^q}
\end{aligned}$$

を得る.

<u>**V.** $\Pi_{q_0} g$ **の評価**</u>. (6.77) から

$$\left|\Pi_{q_0} g\right| \le G \le a(\delta)\sqrt{n} \le \sqrt{n} a_{q_0}$$

である. さらに, 補題の仮定 $\mathsf{P}g^2 < \delta^2$ $(\forall g \in \mathcal{G})$ から

$$\|\Pi_{q_0} g\|_{L_2(\mathsf{P})} \le \delta$$

となることがわかる. これらのことに注意をして, 補題 6.54 を用いると

$$\begin{aligned}
\mathsf{E}\left[\sup_{g\in\mathcal{G}}\left|\nu_n \Pi_{q_0} g\right|\right] &\le 96 \max_{g\in\mathcal{G}} \frac{\|\Pi_{q_0} g\|_\infty}{\sqrt{n}} \log N_{q_0} + 96 \max_{g\in\mathcal{G}} \|\Pi_{q_0} g\|_{L_2(\mathsf{P})} \sqrt{\log N_{q_0}} \\
&\le 96 a_{q_0} \log N_{q_0} + 96\delta \sqrt{\log N_{q_0)}} \\
&\le 96 \frac{\sqrt{\log N_{q_0}}}{2^{q_0}} + 96 \frac{\sqrt{\log N_{q_0}}}{2^{q_0+2}} \\
&\le 130 \frac{\sqrt{\log N_{q_0}}}{2^{q_0+2}} \\
&\le 75 \sum_{q=q_0+1}^{\infty} \frac{1}{2^q} \sqrt{\log N_{q_0}} \\
&\le 75 \sum_{q=q_0+1}^{\infty} \frac{\sqrt{\log N_q}}{2^q}
\end{aligned}$$

がわかる.

III. $\sim$ V の結果を合わせると

$$\begin{aligned}
\mathsf{E}\left[\sup_{g\in\mathcal{G}}\left|\nu_n g \mathbb{1}\{G \le \sqrt{n} a(\delta)\}\right|\right] &\le 382 \sum_{q=q_0+1}^{\infty} \frac{\sqrt{\log N_q}}{2^q} \\
&\le 382 \int_0^\delta \sqrt{\log N_B\big(\epsilon, \mathcal{G}, \|\cdot\|_{L_2(\mathsf{P})}\big)}\, \mathrm{d}\epsilon
\end{aligned}$$

を得る. 最後に I の結果と合わせると

$$\begin{aligned}
&\mathsf{E}\left[\sup_{g\in\mathcal{G}}\left|\nu_n g\right|\right] \\
&\quad = \mathsf{E}\left[\sup_{g\in\mathcal{G}}\left|\nu_n g \mathbb{1}\{G \le \sqrt{n} a(\delta)\}\right|\right] + \mathsf{E}\left[\sup_{g\in\mathcal{G}}\left|\nu_n g \mathbb{1}\{G > \sqrt{n} a(\delta)\}\right|\right] \\
&\quad \le 382 \int_0^\delta \sqrt{\log N_B\big(\epsilon, \mathcal{G}, \|\cdot\|_{L_2(\mathsf{P})}\big)}\, \mathrm{d}\epsilon + 2\sqrt{n}\mathsf{P}G\mathbb{1}\{G > \sqrt{n} a(\delta)\}
\end{aligned}$$

がわかる. □

系 6.56. P を $\mathbb{X} \subset \mathbb{R}$ 上の確率測度とし, $X_1, X_2, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathsf{P}$ とする. $\mathcal{G} = \{g : \mathbb{X} \longrightarrow \mathbb{R}; \mathsf{P}g^2 < \infty\}$ とし, G を関数族 $\mathcal{G}$ の封筒関数とする. このとき

$$\mathsf{E}\left[\sup_{g \in \mathcal{G}} \left|\sqrt{n}\big(\widehat{\mathsf{P}}_n - \mathsf{P}\big)g\right|\right] \leq 2\int_0^{2\|G\|_{L_2(\mathsf{P})}} \sqrt{\log N_B\big(\epsilon, \mathcal{G}, \|\cdot\|_{L_2(\mathsf{P})}\big)}\, \mathrm{d}\epsilon \tag{6.83}$$

が成り立つ.

Proof. 関数族 $\mathcal{G}$ は一つの bracket $[-G, G]$ で覆われるので, $\delta = 2\|G\|_{L_2(\mathsf{P})}$ とおくと

$$N_B\big(\delta, \mathcal{G}, \|\cdot\|_{L_2(\mathsf{P})}\big) = 1$$

となる. 補題 6.55 で定義された $a(\delta)$ は

$$a(\delta) = \frac{\delta}{\sqrt{\max\big\{1, \log N_B\big(\delta, \mathcal{G}, \|\cdot\|_{L_2(\mathsf{P})}\big)\big\}}} = \delta = 2\|G\|_{L_2(\mathsf{P})}$$

となる. Cauchy-Schwarz の不等式と Markov の不等式を用いると

$$\begin{aligned}
\sqrt{n}\mathsf{P}G\mathbb{1}\{G > \sqrt{n}a(\delta)\} &\leq \sqrt{n}\sqrt{\mathsf{P}G^2}\sqrt{\mathsf{P}\mathbb{1}\{G > \sqrt{n}a(\delta)\}} \\
&\leq \sqrt{n}\|G\|_{L_2(\mathsf{P})}\mathsf{P}\mathbb{1}\{G > \sqrt{n}a(\delta)\} \\
&\qquad \big(\because \sqrt{x} \leq x\ (0 \leq x \leq 1)\big) \\
&\leq \sqrt{n}\|G\|_{L_2(\mathsf{P})}\frac{\mathsf{P}G}{\sqrt{n}a(\delta)} \\
&= \|G\|_{L_2(\mathsf{P})}\frac{\mathsf{P}G}{a(\delta)} \\
&= \frac{1}{2}\mathsf{P}G \\
&\leq \frac{1}{2}\sqrt{\mathsf{P}G^2} \\
&= \frac{1}{2}\|G\|_{L_2(\mathsf{P})}
\end{aligned}$$

がわかる. 補題 6.55 から

$$
\begin{aligned}
\mathsf{E}\left[\sup_{g\in\mathcal{G}}|\nu_n g|\right] &\le \int_0^\delta \sqrt{N_B\big(\epsilon,\mathcal{G},\ \|\cdot\|_{L_2(\mathsf{P})}\big)}\,\mathrm{d}\epsilon + \sqrt{n}\mathsf{P}G\mathbb{1}\{G>\sqrt{n}a(\delta)\} \\
&\le \int_0^\delta \sqrt{N_B\big(\epsilon,\mathcal{G},\ \|\cdot\|_{L_2(\mathsf{P})}\big)}\,\mathrm{d}\epsilon + \frac{\delta}{4} \\
&\le \int_0^\delta \sqrt{N_B\big(\epsilon,\mathcal{G},\ \|\cdot\|_{L_2(\mathsf{P})}\big)}\,\mathrm{d}\epsilon + \frac{1}{4}\int_0^\delta \mathrm{d}\epsilon \\
&\le \int_0^\delta \sqrt{N_B\big(\epsilon,\mathcal{G},\ \|\cdot\|_{L_2(\mathsf{P})}\big)}\,\mathrm{d}\epsilon + \frac{1}{4}\int_0^\delta \sqrt{N_B\big(\epsilon,\mathcal{G},\ \|\cdot\|_{L_2(\mathsf{P})}\big)}\,\mathrm{d}\epsilon \\
&\qquad \left(\because 1 \le \sqrt{N_B\big(\epsilon,\mathcal{G},\ \|\cdot\|_{L_2(\mathsf{P})}\big)}\right)
\end{aligned}
$$

からわかる. □