経験過程理論入門

2021 年度前期 日本女子大学
2021 年度夏季集中 大阪府立大学

今野　良彦

平成 33 年 8 月 23 日

はじめに

経験過程の入門である.

2021 年 8 月 23 日更新

目 次

第 8 章　指数型確率不等式

1　チェビシェフの不等式

定理 8.1 確率空間 $(\Omega, \mathcal{F}, \mathbb{P})$ 上の実数値確率変数 X と増加函数 $\phi : \mathbb{R} \to [0, \infty)$ を考える．このとき, $\phi(a) > 0$ なるすべての $a \in \mathbb{R}$ に対して, 次の不等式が成り立つ:

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[\phi(X)]}{\phi(a)}.$$

ただし, $\mathbb{E}[\phi(X)] < \infty$ である.

証明　$A \in \mathcal{B}(\mathbb{R})$ に対し,

$$P(A) = \mathbb{P}(X \in A)$$

とおく．このとき,

$$\begin{aligned}
\mathbb{E}[\phi(X)] &= \int_{\mathbb{R}} \phi(x)\, dP(x) \\
&= \int_a^{\infty} \phi(x)\, dP(x) + \int_{-\infty}^{a} \phi(x)\, dP(x) \\
&\geq \int_a^{\infty} \phi(x)\, dP(x) \\
&\geq \phi(a) \int_a^{\infty} dP(x) \\
&= \phi(a)\mathbb{P}(X \geq a).
\end{aligned}$$

□

例 8.1 X は $\mathbb{R}$ 値確率変数とし,

$$\mu := \mathbb{E}[X], \qquad \sigma^2 := \mathbb{V}[X] \text{（有限）}$$

とする．$X_1, X_2, \ldots, X_n$ は独立同一に X と同じ分布に従うとする．

$$\overline{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$$

としたとき，チェビシェフの不等式（$\phi(x) = x^2\,(x > 0)$）より，任意の $a > 0$ に対して

$$\mathbb{P}(|\overline{X}_n - \mu| \geq a) \leq \frac{\mathbb{E}[(\overline{X}_n - \mu)^2]}{a^2} = \frac{\sigma^2}{na^2} \to 0 \qquad (n \to \infty).$$

したがって，

$$\overline{X}_n \xrightarrow{P} \mu \qquad (n \to \infty).$$

特に，

$$a = \sigma\sqrt{\frac{t}{n}} \qquad (t > 0)$$

とおけば，

$$\mathbb{P}\left(|\overline{X}_n - \mu| \geq \sigma\sqrt{\frac{t}{n}}\right) \leq \frac{1}{t}.$$

よって，

$$\overline{X}_n = \mu + O_P\left(\frac{1}{\sqrt{n}}\right).$$

例 8.2 $\mathcal{X}$ を距離空間とする．確率空間 $(\Omega, \mathcal{F}, \mathbb{P})$ 上の i.i.d. $\mathcal{X}$ 値確率要素列を $X_1, X_2, \ldots, X_n$ とする．$N \in \mathbb{N}$ とし，$\{g_1, g_2, \ldots, g_N\}$ を $\mathcal{X}$ 上の実数値可測函数[1] の集合とし，ある固定した $\sigma^2 > 0$ に対して，

$$\mathbb{V}[g_j(X_1)] < \sigma^2 \qquad (j = 1, 2, \ldots, N)$$

とする．このとき，任意の $a > 0$ に対して，

$$\mathbb{P}(\max_{1\leq j\leq N} |(P_n - P)g_j| \geq a) \leq \frac{N}{n}\frac{\sigma^2}{a^2}. \tag{8.1}$$

[1] $\mathcal{X}$ の距離函数によって誘導される開集合の族を含む最小の σ 集合族 $\mathcal{B}(\mathcal{X})$ と $\mathbb{R}$ 上のボレル集合族 $\mathcal{B}(\mathbb{R})$ の間の可測函数である．

これは,

$$\begin{aligned}
\mathbb{P}(\max_{1\le j\le N}|(P_n-P)g_j|\ge a) &\le \sum_{j=1}^{N}\mathbb{P}(|(P_n-P)g_j|\ge a)\\
&= \sum_{j=1}^{N}\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}g_j(X_i)-Pg_j\right|\ge a\right)\\
&\le \sum_{j=1}^{N}\frac{1}{na^2}\mathbb{E}[\{g_j(X_1)-\mathbb{E}[g_j(X_1)]\}^2]\\
&< \sum_{j=1}^{N}\frac{\sigma^2}{na^2}=\frac{N}{n}\frac{\sigma^2}{a^2}
\end{aligned}$$

からわかる. ここで, (8.1) において,

$$a=\sigma\sqrt{\frac{Nt}{n}}\qquad (t>0)$$

とおけば,

$$\mathbb{P}\left\{\max_{1\le j\le N}\left|(P_n-P)g_j\right|\ge\sigma\sqrt{\frac{Nt}{n}}\right\}\le\frac{1}{t}.$$

したがって,

$$\max_{1\le j\le N}\left|(P_n-P)g_j\right|=O_P\left(\sqrt{\frac{N}{n}}\right).$$

右辺をみると函数族 $\mathcal{G}$ の元の個数がおおくなると収束のスピードがおそくなることがわかる.

例 8.3 X は標準正規分布 $N(0,1)$ に従う確率変数とする. すると, 任意の $\lambda\in\mathbb{R}$ に対して,

$$\mathbb{E}[e^{\lambda X}]=\exp\left(\frac{\lambda^2}{2}\right).$$

よって, 定理 8.1 において, $\phi(x)=e^{\lambda x}$ とする. $t\in\mathbb{R}$ に対して,

$$\mathbb{P}(X\ge t)\le\frac{\mathbb{E}[e^{\lambda X}]}{e^{\lambda t}}=\frac{e^{\lambda^2/2}}{e^{\lambda t}}=\exp\left\{\frac{\lambda^2}{2}-\lambda t\right\}.$$

ここで, $\lambda = t\,(t > 0)$ とおけば,

$$\mathbb{P}(X \geq t) \leq \exp\left\{-\frac{t^2}{2}\right\}.$$

系 8.1 $X_1, X_2, \ldots, X_n$ は独立とし, 各 $X_i\,(i = 1, 2, \ldots, n)$ は正規分布 $N(0, \sigma_i^2)$ に従うとする. ただし, $\sigma_i > 0\,(i = 1, 2, \ldots, n)$ である.

$$b^2 := \sum_{i=1}^{n} \sigma_i^2$$

とおく. このとき, 任意の $a > 0$ に対して,

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq a\right) \leq \exp\left[-\frac{a^2}{2b^2}\right]$$

となる. さらに, $t > 0$ に対して,

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq b\sqrt{2t}\right) \leq e^{-t}$$

となる.

証明　正規分布の性質より, $\frac{1}{b}\sum_{i=1}^{n} X_i$ は $N(0, 1)$ に従うので, 例 8.3 の結果より

$$\mathbb{P}\left(\frac{1}{b}\sum_{i=1}^{n} X_i \geq a\right) \leq \exp\left(-\frac{a^2}{2}\right).$$

よって,

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq ab\right) \quad \leq \quad \exp\left(-\frac{a^2}{2}\right)$$

である. $\tilde{a} = ab$ とおけば,

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq \tilde{a}\right) \quad \leq \quad \exp\left(-\frac{\tilde{a}^2}{2b^2}\right).$$

さらに, $\tilde{a} = b\sqrt{2t}$ とおけば,

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq b\sqrt{2t}\right) \leq e^{-t}.$$

□

系 8.2 $X_1, X_2, \ldots, X_n$ は実数値確率変数列（独立でなくともよい）とし，各 $X_i\,(i = 1, 2, \ldots, n)$ は正規分布 $N(0, \sigma_i^2)$ に従うとする．

$$\sigma^2 = \max_{1 \le i \le n} \sigma_i^2$$

とおく．このとき，任意の $t > 0$ に対して

$$\mathbb{P}\left(\max_{1 \le i \le n} |X_i| \ge \sigma\sqrt{2(t + \log n)}\right) \le 2n \exp\bigl\{-(t + \log n)\bigr\} = 2e^{-t}.$$

証明　任意の $a > 0$ に対して，

$$\mathbb{P}\left(\max_{1 \le i \le n} |X_i| \ge a\right) \le 2n \exp\left(-\frac{a^2}{2\sigma^2}\right). \tag{8.2}$$

これは，

$$\begin{aligned}
\mathbb{P}\left(\max_{1 \le i \le n} |X_i| \ge a\right) &\le \sum_{i=1}^{n} \mathbb{P}\{|X_i| \ge a\} \\
&= 2\sum_{i=1}^{n} \mathbb{P}(X_i \ge a) \\
&= 2\sum_{i=1}^{n} \mathbb{P}\left(\frac{X_i}{\sigma_i} \ge \frac{a}{\sigma_i}\right) \\
&\le 2\sum_{i=1}^{n} \exp\left(-\frac{a^2}{2\sigma_i^2}\right) \qquad \text{（例 8.3）} \\
&\le 2\sum_{i=1}^{n} \exp\left(-\frac{a^2}{2\sigma^2}\right) \\
&\le 2n \exp\left(-\frac{a^2}{2\sigma^2}\right)
\end{aligned}$$

からわかる．ここで，(8.2) において，$a = \sigma\sqrt{2(t + \log n)}$ とおけば，

$$\mathbb{P}\left(\max_{1 \le i \le n} |X_i| \ge \sigma\sqrt{2(t + \log n)}\right) \le 2n \exp\left(-(t + \log n)\right) = 2e^{-t}$$

を得る． □

2　ヘフディンの不等式

$X_1, X_2, \ldots, X_n$ は確率空間 $(\Omega, \mathcal{F}, \mathbb{P})$ 上の実数値確率変数列とする．i.i.d. の場合には, これらは確率変数 X の複製と考える．.

定義 8.1 各 $i = 1, 2, \ldots, n$ に対して, $\mathbb{E}[X_i] = 0$ とし, ある定数 $c_i > 0$ が存在して,

$$|X_i| < c_i$$

とする．これらをヘフディン条件という．

補題 8.1 ヘフディン条件を仮定し, $b^2 = \sum_{i=1}^n c_i^2$ とする．このとき, 任意の $\lambda > 0$ に対して,

$$\mathbb{E}\left[\exp\left\{\lambda \sum_{i=1}^n X_i\right\}\right] \leq \exp\left\{\frac{\lambda^2 b^2}{2}\right\}.$$

証明　$\lambda > 0$ とする．函数

$$\mathbb{R} \ni x \mapsto e^{\lambda x} \in (0, \infty)$$

の凸性より, $0 \leq \alpha \leq 1$ と $x, y \in \mathbb{R}$ に対して,

$$\exp\{\alpha\lambda x + (1-\alpha)\lambda y\} \leq \alpha \exp\{\lambda x\} + (1-\alpha)\exp\{\lambda y\}.$$

いま,

$$\alpha_i = \frac{c_i - X_i}{2c_i} \qquad (i = 1, 2, \ldots, n)$$

と定義する．このとき,

$$X_i = \alpha_i(-c_i) + (1-\alpha_i)c_i$$

なので,

$$\exp\{\lambda X_i\} \leq \alpha_i \exp\{-\lambda c_i\} + (1-\alpha_i)\exp\{\lambda c_i\}.$$

しかし, $\mathbb{E}[\alpha_i] = \dfrac{1}{2}$ なので,

$$\begin{aligned}\mathbb{E}[\exp\{\lambda X_i\}] &\leq \mathbb{E}[\alpha_i]\exp\{-\lambda c_i\} + (1-\mathbb{E}[\alpha_i])\exp\{\lambda c_i\} \\ &= \frac{1}{2}\exp\{-\lambda c_i\} + \frac{1}{2}\exp\{\lambda c_i\}.\end{aligned}$$

いま, すべての x に対して,

$$e^{-x}+e^{x}=2\sum_{k=0}^{\infty}\frac{x^{2k}}{(2k)!}$$

である. 一方,

$$\exp\left\{\frac{x^2}{2}\right\}=\sum_{k=0}^{\infty}\frac{x^{2k}}{2^k k!}$$

である.

$$(2k)!\geq 2^k k!$$

なので,

$$e^{-x}+e^{x}\leq 2\exp\left\{\frac{x^2}{2}\right\}.$$

これより

$$\mathbb{E}[\exp\{\lambda X_i\}]\leq\frac{1}{2}\{e^{-\lambda c_i}+e^{\lambda c_i}\}\leq\exp\left\{\frac{\lambda^2 c_i^2}{2}\right\}.$$

したがって,

$$\mathbb{E}\left[\exp\left\{\lambda\sum_{i=1}^{n}X_i\right\}\right]\leq\exp\left\{\lambda^2\sum_{i=1}^{n}\frac{c_i^2}{2}\right\}.$$

□

定理 8.2 ヘフディン条件を仮定し, $b^2:=\sum_{i=1}^{n}c_i^2$ とする. このとき, 任意の $a>0$ に対し,

$$\mathbb{P}\left(\sum_{i=1}^{n}X_i\geq a\right)\leq\exp\left\{-\frac{a^2}{2b^2}\right\}.$$

さらに, 任意の $t>0$ に対して,

$$\mathbb{P}\left(\sum_{i=1}^{n}X_i\geq b\sqrt{2t}\right)\leq e^{-t}.$$

証明 $\lambda>0$ とする. $\phi(x)=\exp\{\lambda x\}$ として, チェビシェフの不等式 (定理 8.1) を用いて, 補題 8.1 を用いる. 任意の $a>0$ に対して,

$$\mathbb{P}\left(\sum_{i=1}^{n}X_i\geq a\right)\leq\exp\left(\frac{\lambda^2 b^2}{2}-\lambda a\right). \tag{8.3}$$

これは,

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq a\right) \leq \frac{\mathbb{E}\left[\exp\left\{\lambda \sum_{i=1}^{n} X_i\right\}\right]}{\exp\{\lambda a\}}$$
$$\leq \frac{\exp\left\{\frac{\lambda^2 b^2}{2}\right\}}{\exp\{\lambda a\}} = \exp\left(\frac{\lambda^2 b^2}{2} - \lambda a\right)$$

からわかる．ここで, (8.3) において,

$$\lambda = \frac{a}{b^2}$$

とおく．

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq a\right) \leq \exp\left(-\frac{\lambda^2 b^2}{2}\right)$$

を得る．

最後に, $a = b\sqrt{2t}$ とおけば,

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq b\sqrt{2t}\right) \leq e^{-t}$$

を得る． □

系 8.3 $X_1, X_2, \ldots, X_n$ は独立同一に分布 P に従う $\mathcal{X}$ 値確率変数列とし, P_n をこれらに基づく経験確率測度とする． N 個の函数 $g_j : \mathcal{X} \to \mathbb{R}\,(j = 1, 2, \ldots, N)$ を考える．ある定数 $K > 0$ が存在して,

$$\mathbb{E}[g_j(X)] = 0,$$
$$\sup_{x \in \mathcal{X}} |g_j(x)| \leq K \qquad (j = 1, 2, \ldots, N)$$

とする．このとき, すべての j と $t > 0$ に対して,

$$\mathbb{P}\left(|P_n g_j| \leq K\sqrt{\frac{2t}{n}}\right) \leq 2e^{-t}.$$

さらに,

$$\mathbb{P}\left(\max_{1 \leq j \leq N} |P_n g_j| \geq K\sqrt{\frac{2(t + \log N)}{n}}\right) \leq 2e^{-t}.$$

証明　まず, $a > 0$ と $j = 1, 2, \ldots, N$ に対して,

$$\mathbb{P}\left(|P_n g_j| \geq a\right) \leq 2\exp\left(-\frac{na^2}{2K^2}\right). \tag{8.4}$$

これは

$$\begin{aligned}
\mathbb{P}\left(|P_n g_j| \geq a\right) &= \mathbb{P}(P_n g_j \geq a) + \mathbb{P}(-P_n g_j \leq -a) \\
&= 2\mathbb{P}(P_n g_j \geq a) \\
&= 2\mathbb{P}\left(\sum_{i=1}^{n} g_j(X_i) \geq na\right) \\
&\leq 2\exp\left(-\frac{n^2a^2}{2nK^2}\right) \qquad (\text{定理 } 8.2) \\
&= 2\exp\left(-\frac{na^2}{2K^2}\right)
\end{aligned}$$

からわかる. ここで,

$$t = \frac{na^2}{2K^2} \qquad (t > 0)$$

とおけば,

$$a = K\sqrt{\frac{2t}{n}}$$

となる. これを (8.4) に代入すれば,

$$\mathbb{P}\left(|P_n g_j| \geq K\sqrt{\frac{2t}{n}}\right) \leq 2e^{-t}$$

を得る.

　さらに, $a > 0$ に対して,

$$\mathbb{P}\left(\max_{1\leq j\leq N} |P_n g_j| \geq a\right) \leq 2\exp\left(-\frac{na^2}{2K^2} + \log N\right). \tag{8.5}$$

これは

$$
\begin{aligned}
\mathbb{P}\left(\max_{1\le j\le N}|P_n g_j|\ge a\right) &\le \sum_{j=1}^{N}\mathbb{P}\left(|P_n g_j|\ge a\right)\\
&\le N\mathbb{P}\left(|P_n g_j|\ge a\right)\\
&\le 2N\exp\left(-\frac{na^2}{2K^2}\right)\\
&= 2\exp\left(-\frac{na^2}{2K^2}+\log N\right)
\end{aligned}
$$

からわかる．ここで，

$$
t=\frac{1}{2}\left(\frac{na^2}{K^2}-2\log N\right)
$$

とおけば，

$$
a=K\sqrt{\frac{2(t+\log N)}{n}}
$$

となる．これを (8.5) に代入すれば，

$$
\mathbb{P}\left(\max_{1\le j\le N}|P_n g_j|\ge K\sqrt{\frac{2(t+\log N)}{n}}\right)\le e^{-t}
$$

を得る． □

補題 8.2 $X_1, X_2, \ldots, X_n$ は独立同一に分布 P に従う $\mathcal{X}$ 値確率変数列とし，P_n をこれらに基づく経験確率測度とする． N 個の函数 $g_j : \mathcal{X} \to \mathbb{R}\,(j = 1, 2, \ldots, N)$ を考える．ある定数 $K>0$ が存在して，

$$
\begin{aligned}
&\mathbb{E}[g_j(X)]=0,\\
&\sup_{x\in\mathcal{X}}|g_j(x)|\le K \qquad (j=1,2,\ldots,N)
\end{aligned}
$$

とする．このとき，

$$
\mathbb{E}\left[\max_{1\le j\le N}|P_n g_j|\right]\le\frac{K}{n}\sqrt{2\log(2N)}.
$$

証明 $x \in \mathbb{R}$ に対して,

$$e^{|x|} \leq e^{x} + e^{-x}$$

であることに注意して, 任意の $\lambda > 0$ に対して,

$$\begin{aligned}
\mathbb{E}\left[\exp\left(\lambda n |P_n g_j|\right)\right] &\leq \mathbb{E}\left[\exp\left(\lambda n P_n g_j\right)\right] + \mathbb{E}\left[\exp\left(-\lambda n P_n g_j\right)\right] \\
&\leq 2\exp\left(\frac{\lambda K^2}{2}\right) \qquad (\text{補題 } 8.1) \qquad (8.6)
\end{aligned}$$

となる. したがって,

$$\mathbb{E}\left[\max_{1\leq j\leq N} n|P_n g_j|\right] \leq \frac{\log(2N)}{\lambda} + \frac{\lambda K^2}{2} \qquad (8.7)$$

を得る. これは, (8.6) と定理 5.11 に用いれば,

$$\begin{aligned}
\mathbb{E}\left[\max_{1\leq j\leq N} n|P_n g_j|\right] &= \frac{1}{\lambda}\mathbb{E}\left[\log\exp\left(\lambda \max_{1\leq j\leq N} n|P_n g_j|\right)\right] \\
&\leq \frac{1}{\lambda}\log\left(\mathbb{E}\left[\exp\left(\lambda \max_{1\leq j\leq N} n|P_n g_j|\right)\right]\right) \\
&\qquad (\text{Jensen の不等式 (定理 } 5.11)) \\
&= \frac{1}{\lambda}\log\left(\mathbb{E}\left[\max_{1\leq j\leq N}\exp\left(\lambda n|P_n g_j|\right)\right]\right) \\
&\leq \frac{1}{\lambda}\log\left(\sum_{j=1}^{N}\mathbb{E}\left[\exp\left(\lambda n|P_n g_j|\right)\right]\right) \\
&\leq \frac{1}{\lambda}\log\left(2N\exp\left(\frac{\lambda K^2}{2}\right)\right) \\
&= \frac{\log(2N)}{\lambda} + \frac{\lambda K^2}{2}
\end{aligned}$$

からわかる. 関数

$$\mathbb{R} \ni \lambda \mapsto \frac{\log(2N)}{\lambda} + \frac{\lambda K^2}{2} \in \mathbb{R}$$

を最小にする λ の値は

$$\lambda = \sqrt{\frac{2\log(2N)}{K}}$$

となり, 最小値は

$$\frac{\log(2N)}{\lambda} + \frac{\lambda K^2}{2} = K\sqrt{2\log(2N)}$$

となる．このことと (8.7) より

$$\mathbb{E}\left[\max_{1\le j\le N}|P_n g_j|\right] \le \frac{K}{n}\sqrt{2\log(2N)}$$

を得る． □

3　ベルンシュタインの不等式

$X_1, X_2, \ldots, X_n$ を実数値確率変数列とする．

定義 8.2 (ベルンシュタイン条件) $K > 0$ と $\sigma_i\,(i = 1, 2, \ldots, n)$ と $m = 2, 3, \ldots$ とする．すべての $i = 1, 2, \ldots, n$ に対して,

$$\begin{aligned}
&\mathbb{E}[X_i] = 0,\\
&\mathbb{E}[|X_i|^m] \le \frac{m!}{2}K^{m-2}\sigma_i^2
\end{aligned}$$

をみたす．

補題 8.3 ベルンシュタイン条件をみたすとする．

$$b^2 := \sum_{i=1}^n \sigma_i^2$$

とする．このとき, $0 < \lambda < K^{-1}$ に対して,

$$\mathbb{E}\left[\exp\left\{\lambda\sum_{i=1}^n X_i\right\}\right] \le \exp\left\{\frac{\lambda^2 b^2}{2(1-\lambda K)}\right\}.$$

証明　$0 < \lambda < K^{-1}$ に対して,

$$\mathbb{E}[\exp\{\lambda X_i\}] \le \exp\left(\frac{\lambda\sigma_i^2}{2(1-\lambda K)}\right). \tag{8.8}$$

なぜならば, ベルンシュタイン条件より

$$\begin{aligned}\mathbb{E}[\exp\{\lambda X_i\}] &\leq 1+\sum_{k=2}^{\infty}\frac{1}{k!}\lambda^k\mathbb{E}[X_i^k]\\ &\leq 1+\sum_{k=2}^{\infty}\frac{1}{k!}(\lambda K)^{k-2}\sigma_i^2\\ &= 1+\frac{\lambda^2\sigma_i^2}{2(1-\lambda K)}\\ &\leq \exp\left(\frac{\lambda\sigma_i^2}{2(1-\lambda K)}\right)\end{aligned}$$

からわかる. (8.8) より

$$\begin{aligned}\mathbb{E}\left[\exp\left(\lambda\sum_{i=1}^{n}X_i\right)\right] &= \Pi_{i=1}^{n}\mathbb{E}\left[\exp\left(\lambda X_i\right)\right]\\ &\leq \Pi_{i=1}^{n}\exp\left(\frac{\lambda^2\sigma_i^2}{2(1-\lambda K)}\right)\\ &= \exp\left(\frac{\lambda^2 b^2}{2(1-\lambda K)}\right).\end{aligned}$$

□

定理 8.3 ベルンシュタイン条件を仮定し,

$$b^2=\sum_{i=1}^{n}\sigma_i^2$$

とする. このとき, $a>0$ に対して,

$$\mathbb{P}\left(\sum_{i=1}^{n}X_i\geq a\right)\leq\exp\left\{-\frac{a^2}{2(aK+b^2)}\right\}.$$

したがって, $t>0$ に対して,

$$\mathbb{P}\left(\sum_{i=1}^{n}X_i\geq b\sqrt{2t}+Kt\right)\leq e^{-t}.$$

証明　いま, チェビシェフの不等式を函数

$$\phi : \mathbb{R} \ni x \mapsto e^{\lambda x} \in (0, \infty)$$

に適用し, 補題 8.3 を用いると

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq a\right) \quad \leq \quad \exp\left(\frac{\lambda^2 b^2}{2(1-\lambda K)} - \lambda a\right). \tag{8.9}$$

これは

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq a\right) \quad &\leq \quad \frac{\mathbb{E}\left[\exp\left(\lambda \sum_{i=1}^{n} X_i\right)\right]}{\exp\{\lambda a\}} \\
&\leq \quad \frac{\exp\left(\frac{\lambda^2 b^2}{2(1-\lambda K)}\right)}{\exp\{\lambda a\}} \\
&= \quad \exp\left(\frac{\lambda^2 b^2}{2(1-\lambda K)} - \lambda a\right)
\end{aligned}$$

からわかる. ここで,

$$\lambda = \frac{a}{Ka + b^2}$$

とおけば, (8.9) の右辺の指数函数の冪は

$$\begin{aligned}
\frac{\lambda^2 b^2}{2(1-\lambda K)} - \lambda a \quad &= \quad \frac{\left(\frac{a}{Ka+b^2}\right)^2 b^2}{2\left(1 - \frac{aK}{Ka+b^2}\right)} - \frac{a^2}{Ka+b^2} \\
&= \quad = \frac{\frac{a^2 b^2}{(Ka+b^2)^2}}{\frac{2b^2}{Ka+b^2}} - \frac{a^2}{Ka+b^2} \\
&= \quad \frac{a^2}{2(Ka+b^2)} - \frac{a^2}{Ka+b^2} = -\frac{a^2}{2(Ka+b^2)}
\end{aligned}$$

となるので,

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq a\right) \leq \exp\left(-\frac{a^2}{2(Ka+b^2)}\right) \tag{8.10}$$

を得る.

$$
\begin{aligned}
\frac{a^2}{2(Ka+b^2)} = t \quad &\Longleftrightarrow \quad a^2 = 2Kta + 2b^2 \\
&\Longleftrightarrow \quad a^2 - 2Kta - 2b^2 t = 0 \\
&\Longleftrightarrow \quad a = Kt \pm \sqrt{Kt^2 + 2b^2 b}
\end{aligned}
$$

となり, これを (8.10) に代入すれば,

$$
\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq b\sqrt{2t} + Kt\right) \quad \leq \quad \mathbb{P}\left(\sum_{i=1}^{n} X_i \geq \sqrt{Kt^2 + 2b^2 b} + Kt\right) \leq e^{-t}
$$

を得る. □

系 8.4 N 個の函数 $g_j : \mathcal{X} \to \mathbb{R}\,(j = 1, 2, \ldots, N)$ は, 定数 σ^2 と $K > 0$ に対して, 以下をみたす:

$$
\begin{aligned}
&\mathbb{E}[g_j(X)] = 0, \\
&\mathbb{E}[g_j^2(X)] \leq \sigma^2, \\
&\sup_{x \in \mathcal{X}} |g_j(x)| \leq K.
\end{aligned}
$$

このとき, すてべの $j = 1, 2, \ldots, N$ と $t > 0$ に対して,

$$
\mathbb{P}\left(|P_n g_j| \geq \sigma\sqrt{\frac{2t}{n}} + \frac{Kt}{n}\right) \leq 2e^{-t}.
$$

したがって,

$$
\mathbb{P}\left(\max_{1 \leq j \leq N} |P_n g_j| \geq \sigma\sqrt{\frac{2(t + \log N)}{n}} + \frac{K(t + \log N)}{n}\right) \leq 2e^{-t}.
$$

証明　仮定より, $m = 2, 3, \ldots$ に対して,

$$
\mathbb{E}[|g_j(X)|^m] \leq K^{m-2}\mathbb{E}[g_j^2(X)] = K^{m-2}\sigma^2
$$

となるので, ベルンシュタイン条件をみたすので, 定理 8.3 を用いる: 任意の $t > 0$ に対して,

$$
\mathbb{P}\left(|P_n g_j| \geq a\right) \leq 2\exp\left(-\frac{n^2 a^2}{2(naK + n\sigma^2)}\right). \tag{8.11}
$$

これは

$$\begin{aligned}\mathbb{P}\left(|P_n g_j| \geq a\right) &= \mathbb{P}\left(n|P_n g_j| \geq na\right)\\ &\leq \mathbb{P}\left(nP_n g_j \geq na\right) + \mathbb{P}\left(nP_n g_j \leq -na\right)\\ &= \mathbb{P}\left(\sum_{i=1}^{n} g_j(X_i) \geq na\right) + \mathbb{P}\left(\sum_{i=1}^{n}\{-g_j(X_i)\} \geq na\right)\\ &\leq 2\exp\left(-\frac{n^2a^2}{2(naK + n\sigma^2)}\right).\end{aligned}$$

からわかる．ここで,

$$t = \frac{n^2a^2}{2(naK + n\sigma^2)}$$

とおけば,

$$\begin{aligned}&n^2a^2 = 2nKta + 2n\sigma^2 t\\ &\Longleftrightarrow n^2a^2 - 2nKta - 2n\sigma^2 t = 0\\ &\Longleftarrow a = \frac{nKt \pm \sqrt{(nKt)^2 + 2n^3\sigma^2 t}}{n^2} = \frac{Kt}{n} \pm \sqrt{\frac{Kt^2}{n^2} + \frac{2t\sigma^2}{n}}\end{aligned}$$

となる．これを (8.11) に代入すれば,

$$\mathbb{P}\left(|P_n g_j| \geq \frac{Kt}{n} + \sigma\sqrt{\frac{2t}{n}}\right) \leq \mathbb{P}\left(|P_n g_j| \geq \frac{Kt}{n} + \sqrt{\frac{Kt^2}{n^2} + \frac{2t\sigma^2}{n}}\right) \leq e^{-t}$$

を得る．

さらに,

$$\mathbb{P}\left(\max_{1\leq j\leq N} |P_n g_j| \geq a\right) \leq 2\exp\left(-\frac{n^2a^2}{2(naK + n\sigma^2)} - \log N\right). \tag{8.12}$$

これは

$$\begin{aligned}\mathbb{P}\left(\max_{1\leq j\leq N} |P_n g_j| \geq a\right) &\leq \sum_{j=1}^{N} \mathbb{P}\left(|P_n g_j| \geq a\right)\\ &\leq N\mathbb{P}(|P_n g_1| \geq a)\\ &\leq 2N\exp\left(-\frac{n^2a^2}{2(naK + n\sigma^2)}\right)\\ &= 2\exp\left(-\frac{n^2a^2}{2(naK + n\sigma^2)} - \log N\right).\end{aligned}$$

からわかる．ここで

$$t = -\frac{n^2 a^2}{2(naK + n\sigma^2)} - \log N$$

とおけば, a について解けば,

$$a = \frac{K(t + \log N)}{n} \pm \sqrt{\frac{K^2(t + \log N)^2}{n^2} + \frac{2\sigma^2(t + \log N)}{n}} \tag{8.13}$$

を得る．これは,

$$\begin{aligned}
&2nKta + 2nt\sigma^2 = n^2 a^2 - 2(nK \log N)a - 2n(\log N)\sigma^2 \\
&\iff n^2 a^2 - 2nK(t + \log N)a - 2n\sigma^2(t + \log N) = 0 \\
&\iff a = \frac{nK(t + \log N) \pm \sqrt{n^2 K^2 (t + \log N)^2 + 2n^3\sigma^2(t + \log N)}}{n^2} \\
&\iff a = \frac{K(t + \log N)}{n} \pm \sqrt{\frac{K^2(t + \log N)^2}{n^2} + \frac{2\sigma^2(t + \log N)}{n}}
\end{aligned}$$

からわかる．(8.12) と合わせれば,

$$\mathbb{P}\left(\max_{1 \leq j \leq N} |P_n g_j| \geq a\right) \leq 2e^{-t}. \tag{8.14}$$

ただし, a は (8.13) の大きい方である．(8.13) より

$$a \geq \sigma\sqrt{\frac{2(t + \log N)}{n}} + \frac{K(t + \log N)}{n}$$

なるので, (8.14) より

$$\begin{aligned}
&\mathbb{P}\left(\max_{1 \leq j \leq N} |P_n g_j| \geq \sigma\sqrt{\frac{2(t + \log N)}{n}} + \frac{K(t + \log N)}{n}\right) \\
&\leq \mathbb{P}\left(\max_{1 \leq j \leq N} |P_n g_j| \geq \sqrt{\frac{K^2(t + \log N)^2}{n^2} + \frac{2\sigma^2(t + \log N)}{n}} + \frac{K(t + \log N)}{n}\right) \\
&\leq 2e^{-t}
\end{aligned}$$

を得る．

□

第 9 章　ブラケット計量エントロピーに基づく大数の一様強法則

1　古典的グリベンコ・カンテリの定理

$(\Omega, \mathcal{F}, \mathbb{P})$ を確率空間とし, $X_1, X_2, \ldots$ を独立同一分布に従う確率変数列で

$$F(x) := \mathbb{P}(X_1 \le x) \qquad (x \in \mathbb{R})$$

とする. $\widehat{F}_n$ を $X_1, X_2, \ldots, X_n$ に基づく経験分布函数とする:

$$\begin{aligned} \widehat{F}_n(x) &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,\, x]}(X_i), \\ \mathbb{1}_{(-\infty,\, x]}(X_i) &= \begin{cases} 1 & (X_i \le x) \\ 0 & (X_i > x). \end{cases} \end{aligned}$$

定理 9.1

$$\sup_{x\in\mathbb{R}} |\widehat{F}_n(x) - F(x)| \xrightarrow{P} 0 \qquad (n \to \infty).$$

証明　F が連続の場合の証明を与える: 任意の $x \in \mathbb{R}$ に対して,

$$|\mathbb{1}_{(-\infty,\, x]}(X_1) - F(x)| \le 1 \qquad (t \in \mathbb{R}).$$

さらに,

$$\mathbb{E}[\mathbb{1}_{(-\infty,\, x]}(X_1) - F(x)] = \mathbb{P}(X_1 \le x) - F(x) = 0$$

である. これらに注意して, ヘフディンの不等式（系 8.3）を用いると

$$\mathbb{P}\left(|\widehat{F}_n(x) - F(x)| \ge \sqrt{\frac{2t}{n}}\right) \le 2e^{-t} \qquad (t > 0).$$

いま, $0 < \delta < 1$ を任意にとり, $a_0 < a_1 < \cdots < a_N$ を

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

をみたすように取る．すると
$$N \leq 1 + \frac{1}{\delta}$$
となる．$x \in (a_{j-1},\, a_j]\,(j = 1,\, 2,\, ,\ldots,\, N)$ に対して，

$$(-\infty,\, a_{j-1}] \subset (-\infty,\, x] \subset (-\infty, a_j]$$

なので，$t \in \mathbb{R}$ に対して，

$$\mathbb{1}_{(-\infty,\, a_{j-1}]}(t) \leq \mathbb{1}_{(-\infty,\, x]}(t) \leq \mathbb{1}_{(-\infty, a_j]}(t)$$

となる．これより，$x \in (a_{j-1},\, a_j]$ に対して，

$$\sup_{x \in \mathbb{R}} |\widehat{F}_n(x) - F(x)| \leq \max_{1 \leq j \leq N} |\widehat{F}_n(a_j) - F(a_j)| + \delta. \tag{9.1}$$

となる．これは，

$$\begin{aligned} \widehat{F}_n(x) - F(x) &\leq \widehat{F}_n(a_j) - F(x) \\ &\leq \widehat{F}_n(a_j) - F(a_{j-1}) \\ &= \widehat{F}_n(a_j) - F(a_j) + F(a_j) - F(a_{j-1}) \\ &\leq \widehat{F}_n(a_j) - F(a_j) + \delta \end{aligned}$$

と

$$\begin{aligned} \widehat{F}_n(x) - F(x) &\geq \widehat{F}_n(a_{j-1}) - F(x) \\ &\geq \widehat{F}_n(a_{j-1}) - F(a_j) \\ &= \widehat{F}_n(a_{j-1}) - F(a_{j-1}) - \{F(a_j) - F(a_{j-1})\} \\ &\geq \widehat{F}_n(a_{j-1}) - F(a_{j-1}) - \delta \end{aligned}$$

からわかる．ヘフディンの不等式（系 8.4）より

$$\mathbb{P}\left(\max_{1 \leq j \leq N} |\widehat{F}_n(a_j) - F(a_j)| \geq \sqrt{\frac{2(t + \log N)}{n}} \right) \leq 2e^{-t} \qquad (t > 0).$$

ここで
$$t = \frac{n\delta^2}{4}$$

とし, n を十分大きくとり,

$$n \geq \frac{4}{\delta^2} \log N$$

とする. このとき,

$$\sqrt{\frac{2(t+\log N)}{n}} \leq \sqrt{\frac{2\left(\frac{n\delta^2}{4}+\frac{n\delta^2}{4}\right)}{n}} = \delta$$

となるので,

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\leq j\leq N} |\widehat{F}_n(a_j) - F(a_j)| \geq \delta\right) \\
&\quad \leq \mathbb{P}\left(\max_{1\leq j\leq N} |\widehat{F}_n(a_j) - F(a_j)| \geq \sqrt{\frac{2(t+\log N)}{n}}\right) \leq 2\exp\left(-\frac{n\delta^2}{4}.\right)
\end{aligned}$$

したがって,

$$n \geq \frac{4}{\delta^2}\log\left(1+\frac{1}{\delta}\right) \geq \frac{4}{\delta^2}\log N$$

に対して,

$$\mathbb{P}\left(\max_{1\leq j\leq N} |\widehat{F}_n(a_j) - F(a_j)| \geq \delta\right) \leq 2\exp\left(-\frac{n\delta^2}{4}\right). \tag{9.2}$$

(9.1) と (9.2) より

$$n \geq \frac{4}{\delta^2}\log\left(1+\frac{1}{\delta}\right) \geq \frac{4}{\delta^2}\log N$$

のとき,

$$\mathbb{P}\left(\sup_{x\in\mathbb{R}} |\widehat{F}_n(x) - F(x)| \geq 2\delta\right) \leq 2\exp\left(-\frac{n\delta^2}{4}\right).$$

これは,

$$X = \max_{1\leq j\leq N} |\widehat{F}_n(a_j) - F(a_j)|, \qquad Y = \sup_{x\in\mathbb{R}} |\widehat{F}_n(x) - F(x)|$$

とおけば, (9.1) から $Y \leq X + \delta$. したがって, $Y \geq 2\delta$ ならば, $X \geq \delta$ なので,

$$\mathbb{P}(Y \geq 2\delta) \leq \mathbb{P}(X \leq \delta)$$

となることからわかる. よって, 定理は示された. □

注意 9.1 リバース・マルチンゲールを用いた議論により, ある定数 c が存在して,

$$\sup_{x\in\mathbb{R}} |\widehat{F}_n(x) - F(x)| \xrightarrow{a.s.} c \qquad (n \to \infty).$$

であることがわかる. Pollard(2002, p.157) を参照. 定理 9.1 より, $c = 0$ となるので,

$$\sup_{x\in\mathbb{R}} |\widehat{F}_n(x) - F(x)| \xrightarrow{a.s.} 0 \qquad (n \to \infty)$$

がわかる. したがって, 確率収束を示すことができれば, 自動的に概収束であることがわかる.

2　計量エントロピー

定義 9.1 $(\mathcal{M}, d_M)$ をする. 距離空間 M の空でない部分集合 $\mathcal{S}$ を考える. 任意の $\delta > 0$ に対して, $N(\delta, \mathcal{S}, \|\cdot\|_{d_M})$ を $\mathcal{S}$ を覆うために必要な δ 球の最小個数とする. ただし, $\|x\|_d = \sqrt{d_M(x, x)}\,(x \in \mathcal{M})$ である. さらに,

$$H(\delta, \mathcal{S}, \|\cdot\|_{d_M}) := \log N(\delta, \mathcal{S}, \|\cdot\|_{d_M})$$

を $\mathcal{S}$ の計量エントロピーとよぶ.

注意 9.2 ときどき, 開近傍の中心が $\mathcal{S}$ に含まれることを要求することを定義に含めている場合がある.

例 9.1 任意の $\boldsymbol{x} = (x_1, x_2, \ldots, x_p), \boldsymbol{y} = (y_1, y_2, \ldots, y_p) \in \mathbb{R}^p$ に対し,

$$d(\boldsymbol{x}, \boldsymbol{y}) = \max_{1\le j\le p} |x_j - y_j| =: |\boldsymbol{x} - \boldsymbol{y}|_\infty$$

とし, $\mathcal{S} = [-1, 1]^p \subset \mathbb{R}^p$ とする. すると　$\|\cdot\|_d = |\cdot|_\infty$ である. 任意の $a > 0$ に対して,

$$\lceil a \rceil := \min\{m \in \mathbb{N};\, m \ge a\}$$

とおく. このとき,

$$N(\delta, \mathcal{S}, |\cdot|_\infty) \le \{\lceil 1/\delta \rceil\}^p \le \left(1 + \frac{1}{\delta}\right)^p.$$

よって,

$$H(\delta, \mathcal{S}, |\cdot|_\infty) \le p \log\left(1 + \frac{1}{\delta}\right).$$

3　ブラケット計量エントロピー

$(\Omega, \mathcal{F}, \mathbb{P})$ を確率空間とし, $\mathcal{X}$ を距離空間とする. $X : \Omega \to \mathcal{X}$ を $\mathcal{X}$ 値確率変数とし,

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

とする.

$p \geq 1$ とし,

$$L^p(P) := \{g : \mathcal{X} \to \mathbb{R};\, P|g|^p < \infty\}$$

とし, この空間上のノルムを

$$\|g\|_{L^p(P)} := \{P|g|^p\}^{1/p}, \qquad (g \in L^p(P))$$

で定める. また,

$$\begin{aligned} L^\infty &:= \{g;\, \|g\|_\infty < \infty\}, \\ \|g\|_\infty &:= \sup_{x \in \mathcal{X}} |g(x)| \end{aligned}$$

で定める. 例 9.1 の Euclid ノルムと同じ記号を用いる乱用をしていることに注意せよ.

定義 9.2 函数族 $\mathcal{G} \subset L^p(P)$ を考える. N を自然数とし, 任意の $\delta > 0$ に対して,

$$\{[g_j^L, g_j^U]\}_{j=1}^N \subset L^p(P)$$

を次の条件をみたすものとする:

(i) $j = 1, 2, \ldots, N$ に対して,

$$g_j^L(x) \leq g_j^U(x)(x \in \mathcal{X}) \quad \text{かつ} \quad \|g_j^U - g_j^L\|_{L^p(P)} \leq \delta.$$

(ii) 任意の $g \in \mathcal{G}$ に対して, ある $j \in \{1, 2, \ldots, N\}$ が存在して,

$$g_j^L(x) \leq g(x) \leq g_j^U(x) \qquad (x \in \mathcal{X}).$$

$\{[g_j^L, g_j^U]\}_{j=1}^N$ を函数族 $\mathcal{G}$ に対する δ ブラケット集合とよぶ. ブラケット付き被覆数を

$$N_B(\delta, \mathcal{G}, \|\cdot\|_{L^p(P)}) := \min\{N;\, \{[g_j^L, g_j^U]\}_{j=1}^N \text{ は函数族 } \mathcal{G} \text{ の } \delta \text{ ブラケット集合}\}$$

とする．ブラケット付き計量エントロピーを

$$H_B(\delta, \mathcal{G}, \|\cdot\|_{L^p(P)}) = \log N_B(\delta, \mathcal{G}, \|\cdot\|_{L^p(P)})$$

で定める．

注意 9.3 $H(\cdot, \mathcal{G}, \|\cdot\|_\infty)$ をノルム $\|\cdot\|_\infty$ から誘導される函数族 $\mathcal{G}$ の計量エントロピーとする．このとき，

$$H_B(\delta, \mathcal{G}, \|\cdot\|_{L^p(P)}) \le H(\delta/2, \mathcal{G}, \|\cdot\|_\infty)$$

である．これは，以下の議論から確認ができる．

$\forall\delta > 0$ に対して，

$$r_0 = -\infty < r_1 < r_2 < \cdots < r_N < r_{N+1} = \infty$$

と定める．函数 $\mathbb{R} \ni x \mapsto F(x) \in [0, 1]$ は連続非減少で，$0 \le F(x) \le 1$ なので，

$$\max_{j=1,2,\ldots,N+1} |F(r_j) - F(r_{j-1})|^{1/p} \le \delta$$

とできる．これは，

$$1 + N \le 1 + \frac{1}{\delta^p} \le \left(1 + \frac{1}{\delta}\right)^p$$

とし，

$$F(r_j) = \frac{j}{N+1} \qquad (j = 0, 1, \ldots, N+1)$$

ととればよい．さらに，

$$\begin{aligned} \|\mathbb{1}_{(-\infty, r_j]}(x) - \mathbb{1}_{(-\infty, r_{j-1}]}(x)\|_{L^p(P)} &= \left(\int_{r_{j-1}}^{r_j} dF(x)\right)^{1/p} \\ &= \{F(r_j) - F(r_{j-1})\}^{1/p} \le \delta \end{aligned}$$

となる．したがって，$\forall r \in \mathbb{R}$ に対して，ある $j \in \{1, 2, \ldots, N\}$ が存在して，

$$\mathbb{1}_{(-\infty, r]}(x) \in \{\mathbb{1}_{(-\infty, s]}(x)\,(s \in \mathbb{R});\ \|\mathbb{1}_{(-\infty, s]}(x) - \mathbb{1}_{(-\infty, r_j]}(x)\|_{L^p(P)} \le \delta\}$$

となる．よって，

$$N \ge N(\delta, \mathcal{G}, \|\cdot\|_{L^p(P)})$$

である．ゆえに，

$$H(\delta, \mathcal{G}, \|\cdot\|_{L^p(P)}) \leq \log\left(1+\frac{1}{\delta}\right)^{1/p} = p\log\left(1+\frac{1}{\delta}\right).$$

例 9.2 $L > 0$ を定数とる．区間 $\mathcal{I} \subset \mathbb{R}$ 上の実数値函数 $g : \mathcal{I} \to \mathbb{R}$ がリプシッツ係数 L のリプシッツ函数とする．任意の $x, x' \in \mathcal{I}$ に対し，

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

をみたすことである．

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

$$\mathcal{G} := \{g : (0, 1] \to [0, 1];\ g \text{ リプシッツ係数 } 1 \text{ のリプシッツ函数 }\}$$

とする．

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

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

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，

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

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，

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

なので，

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

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

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

を得る.

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

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

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

となる. 一度, $\lfloor\frac{g(a_j)}{\delta}\rfloor$ を選ぶと, $\lfloor\frac{g(a_{j+1})}{\delta}\rfloor$ 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, 函数 $\tilde{g}$ の個数の最大は

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

よって, $0<\delta<1$ に対して,

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

例 9.3 $\mathcal{X}=\mathbb{R}$ と $\mathcal{G}:=\{\mathbb{1}_{(-\infty, r]}; r\in\mathbb{R}\}$ とする. $\mathcal{X}$ 値確率変数 X の分布函数

$$F(x)=\mathbb{P}(X\le x), \qquad (x\in\mathbb{R})$$

は連続とする. また, X の分布を P と書く. $p\ge 1$ と $0<\delta\le 1$ に対して,

$$H(\delta, \mathcal{G}, \|\cdot\|_{L^p(P)})\le p\log\left(1+\frac{1}{\delta}\right)$$

となる.

以下の記法を思う出そう:

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

補題 9.1 $0 < \delta \le 1$ に対して,

$$H_B(\delta, \mathcal{G}, \|\cdot\|_{L^1(P)}) < \infty$$

とする．このとき,

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

証明 $\delta > 0$ は任意とし, $\{[g_j^L, g_j^U]\}_{j=1}^N \subset L^1(P)$ は $N = N_B(\delta, \mathcal{G}, \|\cdot\|_{L^1(P)})$ のブラケットによる最小の δ 被覆集合とする．これは有限集合なので,

$$\max\left\{\left(\max_{1\le j\le N} |(P_n - P)g_j^L|\right), \left(\max_{1\le j\le N} |(P_n - P)g_j^U|\right)\right\} \xrightarrow{P} 0 \qquad (n \to \infty)$$

となる．ただし, $\max\{a, b\}$ $(a, b \in \mathbb{R})$ は a, b の小さくない方である．$g_j^L(x) \le g(x) \le g_j^U(x)$ と $P(g_j^U - g_j^L) \le \delta$ のとき,

$$|(P_n - P)g| \le \max\{|(P_n - P)g_j^L|, |(P_n - P)g_j^U|\} + \delta. \tag{9.3}$$

これは

$$\begin{aligned}
(P_n - P)g &\le P_n g_j^U - P g_j^L \\
&= (P_n - P)g_j^U + P(g_j^U - g_j^L) \\
&\le (P_n - P)g_j^U + \delta, \\
(P_n - P)g &\ge P_n g_j^L - P g_j^U \\
&= (P_n - P)g_j^L - P(g_j^U - g_j^L) \\
&\le (P_n - P)g_j^L - \delta
\end{aligned}$$

からわかる．よって, (9.3) より

$$\sup_{g\in\mathcal{G}} |(P_n - P)g| \le \max\left\{\left(\max_{1\le j\le N} |(P_n - P)g_j^L|\right), \left(\max_{1\le j\le N} |(P_n - P)g_j^U|\right)\right\} + \delta.$$

ヘフディンの不等式 (系 8.3) より

$$\mathbb{P}\left(\max\left\{\left(\max_{1\le j\le N} |(P_n - P)g_j^L|\right), \left(\max_{1\le j\le N} |(P_n - P)g_j^U|\right)\right\} \ge K\sqrt{\frac{2(t + \log(2N))}{n}}\right)$$
$$\le 2e^{-t} \qquad (t > 0)$$

となる. ただし,

$$K = \max_{1 \le j \le N} \left\{ \sup_{x \in \mathcal{X}} |g_j^U(x) - Pg_j^U|, \ \sup_{x \in \mathcal{X}} |g_j^L(x) - Pg_j^L| \right\}$$

である. ここで,

$$t = \frac{n\delta^2}{4}$$

とし, n を十分大きく取り

$$n \ge \frac{4K^2}{\delta^2} \log(2N)$$

とする. このとき,

$$\sqrt{\frac{2(t + \log(2N))}{n}} \le \sqrt{2\left(\frac{n\delta^2}{4} + \frac{n\delta^2}{4}\right)} = \delta$$

となるので,

$$\begin{aligned}
&\mathbb{P}\left[\max\left\{\left(\max_{1 \le j \le N} |(P_n - P)g_j^L\right), \left(\max_{1 \le j \le N} |(P_n - P)g_j^U\right)\right\} \ge \delta\right] \\
&\quad \le \mathbb{P}\Bigg[\max\left\{\left(\max_{1 \le j \le N} |(P_n - P)g_j^L\right), \left(\max_{1 \le j \le N} |(P_n - P)g_j^U\right)\right\} \\
&\qquad\qquad \ge K\sqrt{\frac{2(t + \log(2N))}{n}}\Bigg] \\
&\quad \le 2\exp\left(-\frac{n\delta^2}{4}\right).
\end{aligned}$$

したがって,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{g \in \mathcal{G}} |(P_n - P)g| \ge 2\delta\right) \\
&\quad \le \mathbb{P}\left[\max\left\{\left(\max_{1 \le j \le N} |(P_n - P)g_j^L\right), \left(\max_{1 \le j \le N} |(P_n - P)g_j^U\right)\right\} \ge \delta\right] \\
&\quad \le 2\exp\left(-\frac{n\delta^2}{4}\right) \longrightarrow 0 \qquad (n \to \infty).
\end{aligned}$$

□

4　包絡函数

定義 9.3 函数族 $\mathcal{G}$ の包絡函数 G とは,

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

である. $\mathcal{G}$ が一様有界とは,

$$\|G\|_\infty = \sup_{x \in \mathcal{X}} |G(x)| < \infty$$

のときをいう.

注意 9.4 $p \geq 1$ と $0 < \delta \leq 1$ に対して,

$$H_B(\delta, \mathcal{G}, \|\cdot\|_{L^p(P)}) < \infty \Longrightarrow G \in L^p(P).$$

同様に,

$$H(\delta, \mathcal{G}, \|\cdot\|_\infty) < \infty \Longrightarrow \|G\|_\infty < \infty$$

である.

5　ブラケット計量エントロピーと母数空間のコンパクト性

補題 9.2 (Θ, d) をコンパクトな距離空間とし, $\mathcal{G} := \{g_\theta; \theta \in \Theta\}$ とする. 以下のを仮定する:

- すべての $x \in \mathcal{X}$ に対して, 写像

$$\Theta \ni \theta \mapsto g_\theta(x) \in \mathcal{G} \subset L^p(P)$$

  は連続. ただし, $\mathcal{X}$ は距離空間とし, $\mathcal{G}$ のノルムは $L^p(P)$ とする.

- $G(x) := \sup_{\theta \in \Theta} |g_\theta(x)| \in L^p(P)$.

このとき, $0 < \delta \leq 1$ に対して,

$$H_B(\delta, \mathcal{G}, \|\cdot\|_{L^p(P)}) < \infty.$$

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

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

と定める. すべての x に対して, 函数

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

は連続なので,

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

$G \in L^p(P)$ なので, 優収束定理 (定理 3.6) より

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

任意の $\delta > 0$ に対して, $\rho_\theta > 0$ をうまく取り,

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

とできる.

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

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

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

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

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

とおけば,

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

よって,

$$H_B(2\delta,\,\mathcal{G},\,\|\cdot\|_{L^p(P)}) \leq \log N < \infty.$$

□

6 混合型モデルの最尤推定量の一致性

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

Y は実数値確率変数で, 分布函数 F_0 を持つ. $Y = y\,(y \in \mathbb{R})$ を与えたときに, 実数値確率変数 X 確率密度函数 $k(x|\,y)$ を持つ既知の分布に従うものとする. このとき, X は混合密度函数

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

を持つ.

いま, 観測 $X_1, X_2, \ldots, X_n$ は X の i.i.d. 複製とする. $\mathcal{F}$ はすべての分布函数の集まりとする. F_0 の最尤推定量（存在するならば）は

$$\begin{aligned} p_{\widehat{F}_n}(x) &:= \operatorname*{argmax}_{F \in \mathcal{F}} \widehat{F}_n \log p_F, \\ \widehat{F}_n(x) &:= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,\,x]}(X_i) \end{aligned}$$

である. Hellinger 距離は

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

であった.

補題 9.3

$$h(p_{\widehat{F}_n},\,p_{F_0}) \xrightarrow{P} 0 \qquad (n \to \infty).$$

証明

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

とする. $\mu,\,\mu_n \in \mathcal{H}^*\,(n = 1,\,2,\,\ldots)$ に対して, 任意の有界連続函数 f に対して,

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

が成立するとき, μ_n は μ に弱収束（漠収束）するといい,

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

と書く．すると, 漠位相によって定義される $\mathcal{H}^*$ 上の距離函数 d が存在して,

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

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

もし, ほとんどすべての $x \in \mathbb{R}$ に対して,

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

は連続で,

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

であれば, 写像

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

は d に関して連続である．ただし, $\mathcal{H}$ はすべての分布函数の集まりであり, d は

$$d(\widehat{F}_n,\, F_0) \longrightarrow 0\,(n \to \infty) \quad \Longleftrightarrow \quad \widehat{F}_n \xrightarrow{w} F_0\,(n \to \infty)$$

をみたす $\mathcal{H}$ 上の距離関数である．なぜならば,

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

は有界連続函数であるので,

$$\widehat{F}_n \xrightarrow{w} F_0 \quad \Longrightarrow \quad \int_{\mathbb{R}} k(x|\, y) d\widehat{F}_n(y) \to \int_{\mathbb{R}} k(x|\, y) dF_0(y) \qquad \text{a.e. } x.$$

したがって, 優収束定理（定理 3.6）より

$$\lim_{n \to \infty} h(p_{\widehat{F}_n},\, p_{F_0}) = \left\{ \frac{1}{2} \int_{\mathbb{R}} \{ \sqrt{p_{\widehat{F}_n}} - \sqrt{p_{F_0}} \}^2 \, d\mu(x) \right\}^{1/2} = 0$$

よりわかる．すると, 写像

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

も d に関して連続となる．

いま, つぎの函数族を考える:

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

この函数族の包絡函数 G は

$$G(x) \le 2$$

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

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

は連続なので, 函数族 $\mathcal{G}$ はコンパクトになる. 補題 9.2 より $\mathcal{G}$ は GC 族である. よって,

$$1 - P\left(\frac{2p_{\widehat{F}_n}}{p_{\widehat{F}_n} + p_{F_0}}\right) \le (P_n - P)\left(\frac{2p_{\widehat{F}_n}}{p_{\widehat{F}_n} + p_{F_0}}\right). \tag{9.4}$$

これは

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

からわかる. しかし,

$$h^2(p, p_0) \le 1 - P\left(\frac{2p}{p + p_0}\right) \tag{9.5}$$

である. これは,

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

からわかる. (9.2) と (9.3) より

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

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

$$
h(p_{\widehat{F}_n}, p_{F_0}) \xrightarrow{P} 0 \qquad (n\to\infty)
$$

を得る. □

第 10 章　経験過程の対称化

1　準備

$(\Omega, \mathcal{F}, \mathbb{P})$ は確率空間とし, $\mathcal{X}$ を距離空間とする. X は $\mathcal{X}$ 値確率変数とする. すなわち, $\mathcal{B}(\mathcal{X})$ を $\mathcal{X}$ の開集合族を含む最小の σ 加法族としたとき,

$$X^{-1}(B) \in \mathcal{F} \qquad (B \in \mathcal{B}(\mathcal{X}))$$

である. また, X の分布 P を

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

で定める.

$X_1, X_2, \ldots, X_n$ を X の独立な複製とし, $X_1', X_2', \ldots, X_n'$ も X の独立な複製で $X_1, X_2, \ldots, X_n$ とは独立とする.

$$\boldsymbol{X} = (X_1, X_2, \ldots, X_n), \qquad \boldsymbol{X}' = (X_1', X_2', \ldots, X_n'),$$

と書く.

$\mathcal{G}$ を $\mathcal{X}$ 上の実数値可測函数全体のなす集合とし, $\mathcal{G}$ で添字付けられた経験過程を考える: $B \in \mathcal{B}(\mathcal{X})$ に対して,

$$\begin{aligned} P_n(B) &:= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_B(X_i), \\ P_n'(B) &:= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_B(X_i'). \end{aligned}$$

ただし,

$$\mathbb{1}_B(x) = \begin{cases} 1 & (x \in B) \\ 0 & (x \notin B). \end{cases}$$

いま, $\sup_{g\in\mathcal{G}} P|g| < \infty$ とし,

$$\begin{aligned}\|P_n - P\|_{\mathcal{G}} &:= \sup_{g\in\mathcal{G}} |(P_n - P)g|,\\ \|P'_n - P\|_{\mathcal{G}} &:= \sup_{g\in\mathcal{G}} |(P'_n - P)g|,\\ \|P_n - P'_n\|_{\mathcal{G}} &:= \sup_{g\in\mathcal{G}} |(P_n - P'_n)g|,\end{aligned}$$

と定義する.

次に, 期待値の性質について述べる. $\mathcal{T}$ を空でない集合とし, 各 $t\in\mathcal{T}$ に対して, 函数 $f_t : \mathcal{X} \to \mathbb{R}$ を考える. このとき,

$$\mathbb{E}\left[\sup_{t\in\mathcal{T}} |f_t(X)|\right] \geq \sup_{t\in\mathcal{T}} \mathbb{E}[|f_t(X)|] \geq \sup_{t\in\mathcal{T}} |\mathbb{E}[f_t(X)]|$$

である.

X と Y は実数値確率変数としたとき,

$$\mathbb{E}[\mathbb{E}[Y|\, X]] = \mathbb{E}[Y]$$

である. ただし, Y は $\mathbb{P}$ 可積分である. また, 函数 $f : \mathbb{R} \to \mathbb{R}$ と $g : \mathbb{R}^2 \to \mathbb{R}$ に対して,

$$\mathbb{E}[f(X)g(X, Y)|\, X] = f(X)\mathbb{E}[g(X, Y)|\, X]$$

である. ただし, $f(X)$ と $g(X, Y)$ は $\mathbb{P}$ 可積分である. さらに, $A\in\mathcal{B}(\mathbb{R})$ と $B\in\mathcal{B}(\mathbb{R}^2$ に対して,

$$\mathbb{P}(Y\in A|\, X) = \mathbb{E}[\mathbb{1}_A(Y)|\, X]$$

かつ

$$\begin{aligned}\mathbb{P}(X\in A, (X, Y)\in B|\, X) &= \mathbb{E}[\mathbb{1}_A(X)\mathbb{1}_B(X, Y)|\, X] = \mathbb{1}_A(X)\mathbb{E}[\mathbb{1}_B(X, Y)|\, X]\\ &= \mathbb{1}_A(X)\mathbb{P}((X, Y)\in B|\, X]\end{aligned}$$

である. ここで, Y は $\mathbb{P}$ 可積分.

2 対称化と期待値の評価

補題 10.1 $\sup_{g\in\mathcal{G}} P|g| < \infty$ とする.

$$\mathbb{E}\big[\|P_n - P\|_{\mathcal{G}}\big] \leq \mathbb{E}\big[\|P_n - P'_n\|_{\mathcal{G}}\big].$$

証明　明らかに

$$
\begin{aligned}
\mathbb{E}[P_n g|\,\boldsymbol{X}] &= P_n g\\
\mathbb{E}[P'_n g|\,\boldsymbol{X}] &= P g.
\end{aligned}
$$

これらより

$$(P_n - P)g = \mathbb{E}[(P_n - P'_n)g|\,\boldsymbol{X}].$$

したがって,

$$
\begin{aligned}
\|P_n - P\|_{\mathcal{G}} &= \sup_{g\in\mathcal{G}} |(P_n - P)g|\\
&= \sup_{g\in\mathcal{G}} |\mathbb{E}[(P_n - P'_n)g|\,\boldsymbol{X}]|.
\end{aligned}
$$

前節の評価不等式より

$$
\begin{aligned}
\sup_{g\in\mathcal{G}} |\mathbb{E}[(P_n - P'_n)g|\,\boldsymbol{X}]| &\le \mathbb{E}[\sup_{g\in\mathcal{G}} |(P_n - P'_n)g||\,\boldsymbol{X}]]\\
&= \mathbb{E}[\|P_n - P'_n\|_{\mathcal{G}}|\,\boldsymbol{X}].
\end{aligned}
$$

よって,

$$\|P_n - P\|_{\mathcal{G}} \le \mathbb{E}[\|P_n - P'_n\|_{\mathcal{G}}\,\boldsymbol{X}].$$

最後に,

$$
\begin{aligned}
\mathbb{E}[\|P_n - P'_n\|_{\mathcal{G}}] &\le \mathbb{E}[\mathbb{E}[\|P_n - P'_n\|_{\mathcal{G}}|\,\boldsymbol{X}]]\\
&= \mathbb{E}[\|P_n - P'_n\|_{\mathcal{G}}].
\end{aligned}
$$

□

定義 10.1 ラーディマッハル（Rademacher）列 $\{\sigma_i\}_{i=1}^n$ とは, 独立な確率変数列で,

$$\mathbb{P}(\sigma_i = 1) = \mathbb{P}(\sigma_i = -1) = \frac{1}{2}\,(i = 1,\,2,\,\ldots,\,n)$$

をみたすものである.

$\{\sigma_i\}_{i=1}^n$ はラーデマッハル列で, $\boldsymbol{X}$ と $\boldsymbol{X}'$ は独立とする．対称化経験測度を

$$\begin{aligned} P_n^{\underline{\sigma}} g &:= \frac{1}{n}\sum_{i=1}^n \sigma_i g(X_i), \qquad (g \in \mathcal{G}) \\ \underline{\sigma} &:= (\sigma_1, \sigma_2, \ldots, \sigma_n) \end{aligned}$$

で定義する．さらに,

$$\|P_n^{\underline{\sigma}}\|_{\mathcal{G}} := \sup_{g \in \mathcal{G}} |P_n^{\underline{\sigma}} g|$$

とする.

補題 10.2 $\sup_{g\in\mathcal{G}} P|g| < \infty$ とする.

$$\mathbb{E}\big[\|P_n - P\|_{\mathcal{G}}\big] \le 2\,\mathbb{E}\big[\|P_n^{\underline{\sigma}}\|_{\mathcal{G}}\big].$$

証明 $\boldsymbol{X}'$ の対称化経験測度 $P_n'^{\underline{\sigma}}$ を

$$P_n'^{\underline{\sigma}} := \frac{1}{n}\sum_{i=1}^n \sigma_i g(X_i')$$

で定義する．このとき, $X_i - X_i'$ と $-(X_i - X_i')$ は同じ分布で, $X_i - X_i'$ と $\sigma_i(X_i - X_i')$ は同じ分布なので,

$$\|P_n - P_n'\|_{\mathcal{G}}$$

と

$$\|P_n^{\underline{\sigma}} - P_n'^{\underline{\sigma}}\|_{\mathcal{G}}$$

は同じ分布を持つ．したがって,

$$\begin{aligned} \mathbb{E}[\|P_n - P_n'\|_{\mathcal{G}}] &= \mathbb{E}[\|P_n^{\underline{\sigma}} - P_n'^{\underline{\sigma}}\|_{\mathcal{G}}] \\ &\le \mathbb{E}[\|P_n^{\underline{\sigma}}\|_{\mathcal{G}}] + \mathbb{E}[\|P_n'^{\underline{\sigma}}\|_{\mathcal{G}}] \\ &= 2\mathbb{E}[\|P_n^{\underline{\sigma}}\|_{\mathcal{G}}] \end{aligned}$$

□

3　対称化と確率の評価

補題 10.3 $0 < \delta \leq 1$ とする．このとき, すべての $g \in \mathcal{G}$ に対して,

$$\mathbb{P}\left(|(P_n - P)g| > \frac{\delta}{2}\right) \leq \frac{1}{2}$$

とする．このとき,

$$\mathbb{P}\big(\|P_n - P\|_{\mathcal{G}} > \delta\big) \leq 2\mathbb{P}\left(\|P_n - P_n'\|_{\mathcal{G}} > \frac{\delta}{2}\right).$$

証明　$\mathbb{P}_{\boldsymbol{X}}$ は $\boldsymbol{X}$ を与えたときの条件付き確率とする．$\|P_n - P\|_{\mathcal{G}} > \delta$ のとき, $\boldsymbol{X}$ に依存するある確率変数 $g_* = g_*(\boldsymbol{X})$ が存在して,

$$|(P_n - P)g_*| > \delta.$$

$\boldsymbol{X}'$ は $\boldsymbol{X}$ と独立なので, 仮定より

$$\mathbb{P}_{\boldsymbol{X}}\left(|(P_n' - P)g_*| > \frac{\delta}{2}\right) \leq \frac{1}{2} \Longleftrightarrow \mathbb{P}_{\boldsymbol{X}}\left(|(P_n' - P)g_*| \leq \frac{\delta}{2}\right) \geq \frac{1}{2}$$

となる．したがって,

$$\mathbb{P}\left(|(P_n - P)g_*| > \delta \text{ かつ } |(P_n' - P)g_*| \leq \frac{\delta}{2}\right) \geq \frac{1}{2}\mathbb{P}\left(|(P_n - P)g_*| > \delta\right). \tag{10.1}$$

これは,

$$\begin{aligned}
&\mathbb{P}\left(|(P_n - P)g_*| > \delta \text{ かつ } |(P_n' - P)g_*| \leq \frac{\delta}{2}\right) \\
&\quad = \mathbb{E}\left[\mathbb{P}_{\boldsymbol{X}}\left(|(P_n - P)g_*| > \delta \text{ かつ } |(P_n' - P)g_*| \leq \frac{\delta}{2}\right)\middle|\boldsymbol{X}\right] \\
&\quad = \mathbb{E}\left[\mathbb{P}_{\boldsymbol{X}}\left(|(P_n' - P)g_*| \leq \frac{\delta}{2}\right) \times \mathbb{1}\{|(P_n - P)g_*| > \delta\}\middle|\boldsymbol{X}\right] \\
&\quad \geq \frac{1}{2}\mathbb{E}\left[\mathbb{1}\{|(P_n - P)g_*| > \delta\}\right] \\
&\quad = \frac{1}{2}\mathbb{P}\left(|(P_n - P)g_*| > \delta\right)
\end{aligned}$$

からわかる. (10.1) より

$$\begin{aligned}\mathbb{P}\left(\|P_n - P\|_{\mathcal{G}} > \delta\right) &\leq \mathbb{P}\left(|(P_n - P)g_*| > \delta\right)\\ &\leq 2\mathbb{P}\left(\|P_n - P_n'\|_{\mathcal{G}} > \frac{\delta}{2}\right). \end{aligned} \tag{10.2}$$

これは

$$\begin{aligned}\mathbb{P}\left(\|P_n - P\|_{\mathcal{G}} > \delta\right) &\leq \mathbb{P}\left(|(P_n - P)g_*| > \delta\right)\\ &\leq 2\mathbb{P}\left(|(P_n - P)g_*| > \delta \text{ かつ } |(P_n' - P)g_*| \leq \frac{\delta}{2}\right)\\ &\leq 2\mathbb{P}\left(|(P_n - P_n')g_*| > \frac{\delta}{2}\right)\\ &\leq 2\mathbb{P}\left(\|P_n - P_n'\|_{\mathcal{G}} > \frac{\delta}{2}\right)\end{aligned}$$

からわかる. 一番目の不等号は

$$\|P_n - P\|_{\mathcal{G}} > \delta \quad \text{ならば} \quad |(P_n - P)g_*| > \delta$$

なので,

$$\{\|P_n - P\|_{\mathcal{G}} > \delta\} \subset \{|(P_n - P)g_*| > \delta\}$$

であるので,

$$\mathbb{P}\left(\|P_n - P\|_{\mathcal{G}} > \delta\right) \leq \mathbb{P}\left(|(P_n - P)g_*| > \delta\right)$$

がわかる. 二番目の不等号は

$$|(P_n - P)g_*| > \delta \text{ かつ } |(P_n' - P)g_*| \leq \frac{\delta}{2}$$

ならば,

$$|(P_n - P_n')g_*| \geq |(P_n - P)g_*| - |(P_n' - P)g_*| > \delta - \frac{\delta}{2} = \frac{\delta}{2}$$

よりわかる. 最後の不等号は (10.1) よりわかる. □

系 10.1 $0 < \delta \leq 1$ とする. このとき, すべての $g \in \mathcal{G}$ に対して,

$$\mathbb{P}\left(|(P_n - P)g| > \frac{\delta}{2}\right) \leq \frac{1}{2}$$

とする．このとき，

$$\mathbb{P}\big(\|P_n - P\|_{\mathcal{G}} > \delta\big) \le 4\,\mathbb{P}\left(\|P_n^{\sigma}\|_{\mathcal{G}} > \frac{\delta}{4}\right).$$

証明　補題 10.3 より

$$\mathbb{P}\left(\|P_n - P\|_{\mathcal{G}} > \delta\right) \le 4\mathbb{P}\left(\|P_n^{\sigma}\|_{\mathcal{G}} > \frac{\delta}{4}\right). \tag{10.3}$$

これは

$$\begin{aligned}
\mathbb{P}\left(\|P_n - P\|_{\mathcal{G}} > \delta\right) &\le 2\mathbb{P}\left(\|P_n - P_n'\|_{\mathcal{G}} > \frac{\delta}{2}\right) \\
&= 2\mathbb{P}\left(\|P_n^{\sigma} - P_n'^{\,\sigma}\|_{\mathcal{G}} > \frac{\delta}{2}\right) \\
&\le 2\left\{\mathbb{P}\left(\|P_n^{\sigma}\|_{\mathcal{G}} > \frac{\delta}{4}\right) + \mathbb{P}\left(\|P_n'^{\,\sigma}\|_{\mathcal{G}} > \frac{\delta}{4}\right)\right\} \\
&= 4\mathbb{P}\left(\|P_n^{\sigma}\|_{\mathcal{G}} > \frac{\delta}{4}\right)
\end{aligned}$$

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，

$$|U - V| \le |U| + |V| \le \frac{\delta}{4}$$

となる．したがって，

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

これの対偶をとれば，

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

なので，

$$\mathbb{P}\left(|U - V| > \frac{\delta}{4}\right) \le \mathbb{P}\left(|U| > \frac{\delta}{4}\right) + \mathbb{P}\left(|V| > \frac{\delta}{4}.\right)$$

よりわかる．　□

第 11 章　対称化された経験過程に基づく大数の一様強法則

1　函数族

観測に基づく経験測度に関する函数 g の平均値に対する大数の一様法則を, 函数 g が函数族 $\mathcal{G}$ を動いたときの場合の一様なものを証明する. まず, 函数族 $\mathcal{G}$ は有限集合の場合を考える. ヘフェディンの不等式が適用でききように対称化を行い, 標本平均値と平均偏差の絶対差に対する指数型の確率評価式を求める.

函数族 $\mathcal{G}$ が有限集合でない場合には, 有限集合による函数族 $\mathcal{G}$ の近似を考える. $\delta > 0$ に対して, 有限集合による δ 近似を δ 被覆という. 最小の δ 被覆の要素の個数を δ 被覆数とよぶ. VC（Vapnik-Chervonenkis）族を導入する. この族に対する被覆数を求める.

$\mathcal{X}$ を距離空間, X を $\mathcal{X}$ 値確率変数とし, その分布を P とする. $X_1, X_2, \ldots, X_n$ は X と独立で同一分布に従う複製とする. 問題は

$$\|P_n - P\|_{\mathcal{G}} \xrightarrow{P} 0, \qquad (n \to \infty) \tag{11.1}$$

を示すことである. (11.1) が成立するような函数族 $\mathcal{G}$ をグリベンコ・カンテリ（Glivenko-Cantelli）族という.

<u>記法の定義</u>　距離空間 $\mathcal{X}$ 上の実数値函数 g に対して, 一様評価ノルムー（sup norm）を

$$\|g\|_\infty := \sup_{x \in \mathcal{X}} |g(x)|$$

で定める.

$N \in \mathbb{N}$ とし, $\{A_k\}_{k=1}^N$ を有限個の事象の集まりとしたとき,

$$\mathbb{P}\left(\bigcup_{k=1}^N A_k\right) \le \sum_{k=1}^N \mathbb{P}(A_k) \le N \max_{1 \le k \le N} \mathbb{P}(A_k). \tag{11.2}$$

補題 11.1 函数族 $\mathcal{G}$ 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,

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

と仮定する. このとき, $t > 0$ に対して,

$$\mathbb{P}\left(\|P_n^\sigma\|_{\mathcal{G}} > K\sqrt{\frac{2(\log N + t)}{n}} \right) \le e^{-t}.$$

さらに, $\log N + t \ge 1$ なる $t > 0$ に対して,

$$\mathbb{P}\left(\|P_n - P\|_{\mathcal{G}} > 4K\sqrt{\frac{2(\log N + t)}{n}} \right) \le 8\, e^{-t}.$$

証明 $g \in \mathcal{G}$ に対して,

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

だったので,

$$\begin{aligned} \mathbb{E}\left[\sigma_i g(X_i)\right] &= \mathbb{E}[\mathbb{E}[\sigma_i g(X_i) |\, X_i]] \\ &= \mathbb{E}\left[\frac{1}{2}g(X_i) - \frac{1}{2}g(X_i)\right] = 0. \end{aligned}$$

また

$$|\sigma_i g(x)| \le K \qquad (x \in \mathcal{X}).$$

ヘフェディングの不等式(系 8.1)より

$$\mathbb{P}\left(|P_n^\sigma g| \ge K\sqrt{\frac{2t}{n}} \right) \le 2e^{-t} \qquad (t > 0)$$

である. さらに, $\#\mathcal{G} = N$ なので, 系 8.3 より

$$\mathbb{P}\left(\|P_n^\sigma\|_{\mathcal{G}} \ge K\sqrt{\frac{2(t + \log N)}{n}} \right) \le 2e^{-t} \qquad (t > 0)$$

となる. 各 $g \in \mathcal{G}$ と $K^2/(n\delta^2) \le 1/2$ なる $\delta > 0$ に対して, チェビシェフの不等式(定理 8.1)より

$$\mathbb{P}\left(|(P_n - P)g| \ge \delta\right) \le \frac{\mathbb{V}[g(X)]}{n\delta^2} \le \frac{K^2}{n\delta^2} \le \frac{1}{2}$$

となる. $\log N + t \geq 1$ のとき, 確率の対称化定理（系 10.1）より

$$\mathbb{P}\left(\|P_n - P\|_{\mathcal{G}} > K\sqrt{\frac{2(t+\log N)}{n}}\right) \leq 4\mathbb{P}\left(\|P_n^{\sigma}\|_{\mathcal{G}} > K\sqrt{\frac{2(t+\log N)}{n}}\right) \leq 8e^{-t}$$

を得る. □

定義 11.1 $(\mathcal{M}, d)$ を距離空間とし, $\mathcal{S} \subset \mathcal{M}$ を空でない部分集合とする. $\delta > 0$ に対して, $\mathcal{S}$ の δ 被覆数 $N(\delta, \mathcal{S}, \|\cdot\|_d) =: N$ を, $\mathcal{S}$ を覆うために必要な δ 開球の最小の個数である. ただし, $\|x\|_d = \sqrt{d(x, x)}\,(x \in M)$ である. すなわち, $N \in \mathbb{N}$ と $s_1, s_2, \ldots, s_N \in \mathcal{M}$ が存在して,

$$\min_{j=1, 2, \ldots, N} d(s, s_j) \leq \delta, \qquad (\forall s \in \mathcal{S})$$

である. このような点 $s_1, s_2, \ldots, s_N$ が存在しないとき, $N(\delta, \mathcal{S}, \|\cdot\|_d) = \infty$ とする. 集合 $s_1, s_2, \ldots, s_N$ を $\mathcal{S}$ の最小 δ 被覆という.

$$H(\delta, \mathcal{S}, \|\cdot\|_d) = \log N(\delta, \mathcal{S}, \|\cdot\|_d)$$

を $\mathcal{S}$ の計量エントロピーという.

記号　$p \geq 1$ とする. $N(\delta, \mathcal{G}, \|\cdot\|_{L^p(P_n)})$ をノルム

$$\left(P_n|g|^p\right)^{1/p}, \qquad (g \in \mathcal{G})$$

に対応した半ノルム[1] に関する δ 被覆数とし,

$$H(\delta, \mathcal{G}, \|\cdot\|_{L^p(P_n)}) := \log N(\delta, \mathcal{G}, \|\cdot\|_{L^p(P_n)})$$

とする. ここで, $L^p(P_n)$ は半ノルム空間であることに注意. また, $N(\delta, \mathcal{G}, \|\cdot\|_{L^p(P_n)})$ と記すべきところであるが, 上記のように簡単に書いている.

[1] $\|\cdot\|$ がノルムであるとは, $\|\boldsymbol{x}\| \geq 0\,(\boldsymbol{x} \in M)$ で

(i)$\|\boldsymbol{x}\| = 0 \Longleftrightarrow \boldsymbol{x} = \boldsymbol{0}$,　(ii)$\|\alpha\boldsymbol{x}\| = |\alpha|\|\boldsymbol{x}\|\,(\alpha \in \mathbb{R})$,　(iii)$\|\boldsymbol{x}+\boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|\,(\boldsymbol{x}, \boldsymbol{y} \in M)$

をみたすものである. (i) をみたさないものを半ノルム (semi-norm) という. (iii) をみたさないものを準ノルム（quasi-norom）という.

定理 11.1 $K > 0$ とする．任意の $g \in \mathcal{G}$ に対して，

$$\|g\|_\infty < K$$

とする．さらに，$\delta > 0$ に対して，

$$\frac{1}{n} H(\delta, \mathcal{G}, \|\cdot\|_{L^p(P_n)}) \xrightarrow{P} 0, \qquad (n \to \infty)$$

とする．このとき，

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

証明　$\delta > 0$, $N = N(\delta, \mathcal{G}, \|\cdot\|_{L^p(P_n)})$ とし，$g_1, g_2, \ldots, g_N$ を函数族 $\mathcal{G}$ の δ 被覆数とする．

$g \in \mathcal{G}$ を任意に取る．ある $j \in \{1, 2, \ldots, N\}$ に対して，

$$P|g_j - g| = \int_{\mathcal{X}} |g_j(x) - g(x)| \, dP(x) < \delta$$

のとき，

$$\begin{aligned} |P_n^\sigma g| &\leq |P_n^\sigma g_j| + |P_n^\sigma (g - g_j)| \\ &\leq |P_n^\sigma g_j| + P_n^\sigma (|g - g_j|) \\ &\leq |P_n^\sigma g_j| + \delta. \end{aligned}$$

したがって，

$$\|P_n^\sigma\|_{\mathcal{G}} \leq \max_{j=1,2,\ldots,N} |P_n^\sigma g_j| + \delta. \tag{11.3}$$

$\mathbb{P}_{\boldsymbol{X}}$ は $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ を与えたときの条件付き確率とする．ヘフェディングの不等式（補題 11.1）より $t > 0$ に対して，

$$\mathbb{P}_{\boldsymbol{X}} \left(\max_{j=1,2,\ldots,N} |P_n^\sigma g_j| > K \sqrt{\frac{2(t + \log N)}{n}} \right) \leq 2e^{-t}. \tag{11.4}$$

(11.4) より

$$\begin{aligned} &\mathbb{P}_{\boldsymbol{X}} \left(\|P_n^\sigma\|_{\mathcal{G}} > \delta + K \sqrt{\frac{2(t + \log N)}{n}} \right) \\ &\quad \leq \mathbb{P}_{\boldsymbol{X}} \left(\max_{j=1,2,\ldots,N} |P_n^\sigma g_j| > K \sqrt{\frac{2(t + \log N)}{n}} \right) \leq 2e^{-t}. \end{aligned} \tag{11.5}$$

条件付き期待値の性質と 11.5) を用いると

$$\mathbb{P}\left(\|P_n^{\sigma}\|_{\mathcal{G}} >> 2\delta + K\sqrt{\frac{2t}{n}}\right) \leq 2e^{-t} + \mathbb{P}\left(K\sqrt{\frac{2H_1(\delta, \mathcal{G}, \|\cdot\|_{L^p(P_n)})}{n}} > \delta\right) \tag{11.6}$$

となる．これは,

$$\begin{aligned}
&\mathbb{P}\left(\|P_n^{\sigma}\|_{\mathcal{G}} > 2\delta + K\sqrt{\frac{2t}{n}}\right) \\
&\quad \leq \mathbb{P}\left(\|P_n^{\sigma}\|_{\mathcal{G}} > 2\delta + K\sqrt{\frac{2t}{n}} \text{ かつ } K\sqrt{\frac{2H_1(\delta, \mathcal{G}, \|\cdot\|_{L^p(P_n)})}{n}} \leq \delta\right) \\
&\qquad +\mathbb{P}\left(\|P_n^{\sigma}\|_{\mathcal{G}} > 2\delta + K\sqrt{\frac{2t}{n}} \text{ かつ } K\sqrt{\frac{2H_1(\delta, \mathcal{G}, \|\cdot\|_{L^p(P_n)})}{n}} > \delta\right) \\
&\quad \leq \mathbb{P}\left(\|P_n^{\sigma}\|_{\mathcal{G}} > \delta + K\sqrt{\frac{2\log N}{n}} + K\sqrt{\frac{2t}{n}}\right) \\
&\qquad +\mathbb{P}\left(K\sqrt{\frac{2H_1(\delta, \mathcal{G}, \|\cdot\|_{L^p(P_n)})}{n}} > \delta\right) \\
&\quad \leq \mathbb{P}_{\boldsymbol{X}}\left(\|P_n^{\sigma}\|_{\mathcal{G}} > \delta + K\sqrt{\frac{2(t+\log N)}{n}}\right) \\
&\qquad +\mathbb{P}\left(K\sqrt{\frac{2H_1(\delta, \mathcal{G}, \|\cdot\|_{L^p(P_n)})}{n}} > \delta\right) \\
&\quad = \mathbb{E}\left[\mathbb{P}\left(\|P_n^{\sigma}\|_{\mathcal{G}} > \delta + K\sqrt{\frac{2(t+\log N)}{n}}\right) \middle| \boldsymbol{X}\right] \\
&\qquad +\mathbb{P}\left(K\sqrt{\frac{2H_1(\delta, \mathcal{G}, \|\cdot\|_{L^p(P_n)})}{n}} > \delta\right) \\
&\quad \leq 2e^{-t} + \mathbb{P}\left(K\sqrt{\frac{2H_1(\delta, \mathcal{G}, \|\cdot\|_{L^p(P_n)})}{n}} > \delta\right).
\end{aligned}$$

からわかる．いま,

$$\mathbb{P}\left(|(P_n - P)g| > \frac{\delta}{2}\right) \quad \leq \quad \frac{4}{\delta^2}\mathbb{E}\left[|(P_n - P)g|\right] \leq \frac{8K}{n\delta^2} \leq \frac{1}{2}$$

より
$$n \geq \frac{16K}{\delta^2}$$
のとき, 系 10.1 と (11.6) を用いれば,
$$\begin{aligned}\mathbb{P}\left(\|P_n - P\|_{\mathcal{G}} > 8\delta + 4K\sqrt{\frac{2t}{n}}\right) &\leq 4\mathbb{P}\left(\|P_n^{\sigma}\|_{\mathcal{G}} > 2\delta + K\sqrt{\frac{2t}{n}}\right) \\ &\leq 8e^{-t} + 4\mathbb{P}\left(K\sqrt{\frac{2H_1(\delta, \mathcal{G}, \|\cdot\|_{L^p(P_n)})}{n}} > \delta\right)\end{aligned}$$
を得る. $\epsilon > 0$ に対して
$$8e^{-t} \leq \frac{\epsilon}{2} \text{ かつ } 4\mathbb{P}\left(K\sqrt{\frac{2H_1(\delta, \mathcal{G}, \|\cdot\|_{L^p(P_n)})}{n}} > \delta\right) \leq \frac{\epsilon}{2}$$
になるように t と n を取り, さらに,
$$4K\sqrt{\frac{2t}{n}} \leq 2\delta$$
になるように n を大きくとり直せば,
$$\mathbb{P}\left(\|P_n - P\|_{\mathcal{G}} > 10\delta\right) \leq \epsilon$$
とできるので,
$$\|P_n - P\|_{\mathcal{G}} \xrightarrow{P} 0 \qquad (n \to \infty)$$
がわかる. □

例 11.1 $\mathcal{X} = \mathbb{R}$ とし,
$$\mathcal{G} := \{g : \mathbb{R} \to \mathbb{R};\ 0 \leq g(x) \leq 1 \text{ で } g \text{ は単調増加 }\}$$
とする. $N(\,\cdot\,, \mathcal{G}, \|\cdot\|_{L^\infty(P_n)})$ を半ノルム
$$\max_{1 \leq j \leq n} |g(X_j)|$$
により誘導される擬距離に関する被覆数とする. $\delta > 0$ とする. 函数 $g \in \mathcal{G}$ を
$$\tilde{g} = \left\lceil \frac{g(x)}{\delta} \right\rceil \qquad (x \in \mathbb{R})$$

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

$$\left(\begin{array}{c} m+n-1 \\ m \end{array} \right)$$

となる．したがって,

$$N(\delta, \mathcal{G}, \|\cdot\|_{L^\infty(P_n)}) \quad \leq \quad \left(\begin{array}{c} m+n-1 \\ m \end{array} \right)$$

となる． したがって,

$$\begin{aligned} H(\delta, \mathcal{G}, \|\cdot\|_{L^1(P_n)}) \quad &\leq \quad \log N_\infty(\delta, \mathcal{G}, \|\cdot\|_{L^\infty(P_n)}) \\ &\leq \quad m \log(m+n-1) \\ &\leq \quad \left(1 + \frac{1}{\delta}\right) \log\left(n + \frac{1}{\delta}\right) \end{aligned}$$

より

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

なので, 定理 11.1 より, $\mathcal{G}$ は GC 族である．

定理 11.2 G を関数族 $\mathcal{G}$ の包絡関数とし,

$$G \in L^1(P)$$

で, $\delta > 0$ に対して,

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

と仮定する．このとき,

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

証明　すべての $g \in \mathcal{G}$ と任意の $K > 0$ に対して,

$$|(P_n - P)g| \quad \leq \quad |(P_n - P)g\mathbb{1}\{G \leq K\}| + |(P_n - P)g\mathbb{1}\{G > K\}|.$$

(i) $\mathcal{G}_K := \{g\mathbb{1}\{G \le K\} : g \in \mathcal{G}\}$ とする. すると, 任意の $g \in \mathcal{G}_K$ に対して,

$$\sup_{xin\mathcal{X}} |g(x)| \le K.$$

さらに,

$$H(\delta, \mathcal{G}_K, \|\cdot\|_{L^1(P_n)}) \le H(\delta, \mathcal{G}, \|\cdot\|_{L^1(P_n)})$$

なので, 定理 11.1 より

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

(ii)

$$\begin{aligned}\sup_{g\in\mathcal{G}} |(P_n - P)g\mathbb{1}\{G > K\}| &\le (P_n + P)G\mathbb{1}\{G > K\} \\ &= (P_n - P)G\mathbb{1}\{G > K\} + 2P\big(G\mathbb{1}\{G > K\}\big).\end{aligned}$$

任意の K に対して, 大数の法則より

$$(P_n - P)G\mathbb{1}\{G > K\} \xrightarrow{P} 0 \qquad (n \to \infty).$$

また, $G \in L^1(P)$ なので,

$$\lim_{K\to\infty} P\big(G\mathbb{1}\{G > K\}\big) = 0.$$

よって,

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

がわかる. □

2 VC 集合族

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

定義 11.2

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

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

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

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

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

と書く．

例 11.2 $\mathcal{X} = \mathbb{R}$ 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 11.3 次の 2 つは同値である:

(i)

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

(ii)

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

証明 (i) $\Longrightarrow$ (ii) の証明:

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

とおき, 定理 11.2 を用いる．

まず, 任意の $g \in \mathcal{G}$ に対して,

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

$\delta > 0$ に対して, $N(\delta, \mathcal{G}, L^\infty(P_n))$ を疑ノルム

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

から誘導された半ノルムに関する δ 被覆数とする．このとき,

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

であることが

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

からすぐにわかる．しかし, $0 < \delta < 1$ に対して,

$$N(\delta, \{\mathbb{1}_D;\, D \in \mathcal{D}), \|\cdot|_{L^\infty(P_n)}) = \Delta^{\mathcal{D}}(X_1, X_2, \ldots, X_n)$$

である．もし,

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

ならば,

$$\frac{1}{n}H(\delta, \{\mathbb{1}_D;\, D \in \mathcal{D}\}, \|\cdot\|_{L^1(P_n)}) \le \frac{1}{n}\log \Delta^{\mathcal{D}}(X_1, X_2, \ldots, X_n)$$

より

$$\frac{1}{n}H(\delta, \{\mathbb{1}_D;\, D \in \mathcal{D}\}, \|\cdot\|_{L^1(P_n)}) \xrightarrow{P} 0 \qquad (n \to \infty)$$

がわかる． □

定義 11.3 $n \in \mathbb{N}$ に対して,

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

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 11.4 $\mathcal{D}$ が VC 族ならば, GC 族.

証明

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

とおく. すると, $\delta > 0$ に対して,

$$\begin{aligned}\frac{1}{n}H(\delta,\, \mathcal{G},\, \|\cdot\|_{L^1(P_n)}) &\leq \frac{1}{n}\log\Delta^{\mathcal{D}}(X_1,\, X_2,\, \ldots,\, X_n) \leq \log \mathrm{VC}^{\mathcal{D}}(n) \\ &\leq \frac{1}{n}\log\{C\,n^V\} = V\frac{\log n}{n} + \frac{C}{n} \to 0, \qquad (n \to \infty).\end{aligned}$$

□

例 11.3 (a) $\mathcal{X} = \mathbb{R}$, $\mathcal{D} = \{(-\infty, r];\ r \in \mathbb{R}\}$ は VC 族である. なぜならば, $\mathrm{VC}^{\mathcal{D}}(n) \leq n + 1$ よりわかる.
(b) $p \geq 2$ を自然数とする. $\mathcal{X} = \mathbb{R}^p$, $\mathcal{D} = \{\mathbb{1}_{(-\infty,\, \boldsymbol{r}]}(x);\ \boldsymbol{r} \in \mathbb{R}^p\}$ は VC 族である. なぜならば, $\mathrm{VC}^{\mathcal{D}}(n) \leq (n+1)^p$ よりわかる. ただし, $\boldsymbol{r} = (r_1,\, r_2,\, \ldots,\, r_p)$ に対して,

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

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

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

は VC 族である. なぜならば,

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

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

補題 11.2 $\mathcal{D}$, $\mathcal{D}_1$, $\mathcal{D}_2$ は VC 族とする. このとき, 以下の族も VC 族である.

(i) $\mathcal{D}^c := \{D^c;\, D \in \mathcal{D}\}$.
(ii) $\mathcal{D}_1 \cap \mathcal{D}_2 := \{D_1 \cap D_2;\, D_1 \in \mathcal{D}_1,\, D_2 \in \mathcal{D}_2\}$.
(iii) $\mathcal{D}_1 \cup \mathcal{D}_2 := \{D_1 \cup D_2;\, D_1 \in \mathcal{D}_1,\, D_2 \in \mathcal{D}_2\}$.
(iv) p を自然数とする．$\mathbb{R}^p$ の開球

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

証明　証明は略. □

定義 11.4 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 11.3 次の 2 つは同値である:
(i) $\mathcal{D}$ は VC 族.
(ii) $\mathrm{VC}^{\mathcal{D}} < \infty$.

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

$$\mathrm{VC}^{\mathcal{D}} \leq \sum_{k=0}^{V} \left(\begin{array}{c} n \\ k \end{array} \right).$$

□

3　VC 函数族

定義 11.5 函数 $g;\, \mathcal{X} \to \mathbb{R}$ のサブグラフとは

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

である．函数族 $\mathcal{G}$ 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 11.4 $\mathcal{D}$ が VC 族ならば, $\mathcal{G} = \{\mathbb{1}_D;\, D \in \mathcal{D}\}$ も VC 族.

注意 11.1 自然数 p に対して, $\varphi_k : \mathcal{X} \to \mathbb{R}\,(k = 1, 2, \ldots, p)$ を固定した函数とする. このとき,

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

は VC 族となる. 実際,

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

としたとき,

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

である. Pollard (1984) を参照.

定義 11.6 $\mathcal{S}$ は距離空間 $(\mathcal{M}, d)$ の空でない部分集合とする. $\delta > 0$ に対して, $\mathcal{S}$ の δ パッキング数 $D(\delta, \mathcal{S}, \|\cdot\|_d)$ は, 次をみたす最大の自然数 N である. $\mathcal{S}$ の元 $s_1, s_2, \ldots, s_N$ が存在して,

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

補題 11.4 任意の $\delta > 0$ に対して以下が成り立つ.
(i)

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

(ii)

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

証明 (i) は明らか. (ii) を示すために,

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

とする. 明らかに

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

しかし,

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

□

定理 11.5 Q を $\mathcal{X}$ 上の確率測度, $\mathcal{G}$ を $\mathcal{X}$ 上の実数値函数の族, G を函数族 $\mathcal{G}$ の包絡函数, $N_1(\delta, \mathcal{G}, \|\cdot\|_{L^1(Q)})\,(\delta > 0)$ を $L^1(Q)$ ノルムから誘導される距離に関する δ 被覆数とする．族 $\mathcal{G}$ の VC 次元 $\mathrm{VC}^{\mathcal{G}} =: V$ が $V < \infty$ のとき, V に依存する定数 $A > 0$ が存在して,

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

となる．

証明　一般性を失うことなく, $G > 0$ で

$$QG = \int_{\mathcal{X}} G(x)\,dQ(x) = 1$$

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

$$Q_S(A) = \mathbb{P}(S \in A) = \int_A G(x)\,dQ(x) \qquad (A \in \mathcal{B}(\mathcal{X}))$$

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 $g_j : \mathcal{X} \longrightarrow \mathbb{R}\,(j = 1, 2, \ldots, N)$ 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

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

上に T が落ちる確率は

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

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

$$\begin{aligned}\int_{\mathcal{X}} \frac{|g_j(s) - g_k(s)|}{2G(s)}\,dQ_S(x) &= \int_{\mathcal{X}} \frac{|g_j(s) - g_k(s)|}{2}\,dQ(x) \\ &= \frac{Q|g_j - g_k|}{2} > \frac{\delta}{2}\end{aligned}$$

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る確率は, 大きくとも

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

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

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

となる. n を大きくして

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

とすれば,

$$\left(\begin{array}{c} N \\ 2 \end{array}\right)\left(1-\frac{\delta}{2}\right)^n \leq \frac{1}{2} < 1.$$

そのような n に対して, 任意の $j \neq k\,(j,\, k = 1,\, 2,\, \ldots,\, N)$ 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なので,

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

ただし, C は定義 11.3 の定数であり,

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

である.

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

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

よって,

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

より

$$N \leq c^2(16V)^{2V} \delta^{-2V}$$

がわかるので,

$$N(\delta, \mathcal{G}, L^1(Q)) \leq D(\delta, \mathcal{G}, L^1(Q)) \leq \max\{A\delta^{-2V}, e^{\delta/4}\}$$

がわかる. $QG = 1$ に規準化していたので,

$$N(\delta QG, \mathcal{S}, \|\cdot\|_{L^1(P)}) \leq \max\{A\delta^{-2V}, e^{\delta/4}\}$$

がわかる. □

系 11.1 $\mathcal{G}$ を函数族とし, G を函数族 $\mathcal{G}$ の包絡函数とする. 函数族 $\mathcal{G}$ が VC 族で, $PG < \infty$ ならば, $\mathcal{G}$ は GC 族である.

証明 $PG < \infty$ より, $\mathcal{G} \subset L^1(P)$. 函数族 $\mathcal{G}$ の VC 次元を $V < \infty$ としたとき, 定理 11.5 より, $\delta > 0$ に対して,

$$N(\delta PG, \mathcal{G}, \|\cdot\|_{L^1(P)}) \leq \max\{A\delta^{-2\delta}, e^{\delta/4}\}.$$

したがって,

$$\frac{1}{n}\log N(\delta P_n G, \mathcal{G}, \|\cdot\|_{L^1(P_n)}) \leq \frac{1}{n}\log\left(\max\{A\delta^{-2\delta}, e^{\delta/4}\}\right) \longrightarrow 0 \qquad (n \to \infty).$$

さらに,

$$|P_n G - PG| \xrightarrow{P} 0 \qquad (n \to \infty)$$

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

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

よって, 定理 11.1 より, 主張は証明された. □

第 13 章　経験過程

1　$\mathbb{R}^p$ 値確率変数列に対する中心極限定理

$\mathcal{X} = \mathbb{R}$ とする.

定理 13.1 $X_1, X_2, \ldots$ は独立同一分布に従う確率変数列で, $\mathbb{E}[X_1] = \mu$, $\mathbb{V}[X_1] = \sigma^2$ は存在する. このとき,

$$\mathbb{P}\left\{ \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \leq z \right\} \longrightarrow \Phi(z) \qquad (n \to \infty), \qquad (\forall z \in \mathbb{R})$$

が成立する. ただし, Φ は標準正規分布の分布函数, すなわち

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$$

である.

証明　証明は略. □

記号

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,\, 1)$$

または

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} N(0,\, \sigma^2)$$

などと記す.

次に, $\mathcal{X} = \mathbb{R}^p$ $(p \geq 2,\, p \in \mathbb{N})$ とする.

$X_1, X_2, \ldots$ は独立同一分布に従う確率変数列で,

$$\begin{aligned} \boldsymbol{\mu} &= \mathbb{E}[X_1], \\ \Sigma &= \mathbb{E}[X_1 X_1^\top] - \boldsymbol{\mu}\boldsymbol{\mu}^\top \end{aligned}$$

とする. ただし, $\mu^\top$ は縦ベクトル μ の転置である.

定理 13.2

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} N_p(\mathbf{0}, \Sigma),$$

すなわち, 任意の縦ベクトル $\boldsymbol{a} \in \mathbb{R}^p$ に対して,

$$\sqrt{n}[\boldsymbol{a}^\top(\overline{X}_n - \mu)] \xrightarrow{d} N(0, \boldsymbol{a}^\top \Sigma \boldsymbol{a}),$$

.

証明　証明は略. □

2　ドンスカー型中心極限定理

$\mathcal{X} = \mathbb{R}$ とし, 確率変数 X は分布函数 F を持つ, すなわち,

$$F(x) = \mathbb{P}(X \leq x) \qquad (\forall x \in \mathbb{R}).$$

$X_1, X_2, \ldots, X_n$ は X の独立複製としたとき, 経験分布函数を

$$\begin{aligned} \widehat{F}_n(x) \quad &:= \quad \frac{1}{n}\#\{X_i \leq x, i = 1, 2, \ldots, n\} \\ &= \quad \frac{1}{n}\sum_{i=1} \mathbb{1}_{(-\infty, x]}(X_i) \qquad (x \in \mathbb{R}) \end{aligned}$$

で定義する.

これらを用いて

$$W_n(x) := \sqrt{n}(\widehat{F}_n(x) - F(x)), \qquad (x \in \mathbb{R})$$

と定める.

中心極限定理（定理 13.1）より, 各 x に対して,

$$W_n(x) \xrightarrow{d} N(0, F(x)(1 - F(x))).$$

また, 定理 13.2 より, $x, y \in \mathbb{R}\,(x < y)$ に対して,

$$\begin{bmatrix} W_n(x) \\ W_n(y) \end{bmatrix} \xrightarrow{d} N_2(0, \Sigma(x, y))$$

$$\Sigma(x, y) = \begin{bmatrix} F(x)(1 - F(x)) & F(x)(1 - F(y)) \\ F(x)(1 - F(y)) & F(y)(1 - F(y)) \end{bmatrix}.$$

確率過程

$$\mathbb{W}_n = \{W_n(x);\, x \in \mathbb{R}\}$$

を定義する.

定義 13.1 $\mathcal{K}_0$ を $[0,\,1]$ 上の有界な実数値函数の集まりとする. 確率過程 $\mathbb{B} \in \mathcal{K}_0$ は標準ブラウン橋であるとは, 次の条件をみたすものである.
(i) $\mathbb{B}(0) = \mathbb{B}(1) = 0$.
(ii) 任意の有限個の点 $t_1,\, t_2,\, \ldots,\, t_p \in (0,\,1)$ において,

$$\begin{pmatrix} \mathbb{B}(t_1) \\ \mathbb{B}(t_2) \\ \vdots \\ \mathbb{B}(t_p) \end{pmatrix} \sim N_p(\mathbf{0},\, \Sigma)$$

であり,
(iii) $0 \le s \le t \le 1$ に対して,

$$\mathbb{COV}[\mathbb{B}(s),\, \mathbb{B}(t)] = s(1-t).$$

(iv) $\mathbb{B}$ の見本路はほとんど確実に連続.

次に, 確率過程 $\mathbb{W}_F$ を

$$\mathbb{W}_F(t) := \mathbb{B}(F(t)), \qquad (t \in \mathbb{R})$$

で定義する.
したがって,

$$\mathbb{W}_F = \mathbb{B} \circ F$$

である.

定理 13.3 $\mathcal{K} = \{f : \mathbb{R} \longrightarrow \mathbb{R};\ f$ は有界 $\}$ とする. $\mathcal{K}$ 上の確率過程

$$\begin{aligned} \mathbb{W}_n(t) &= \sqrt{n}(\widehat{F}_n(t) - F(t)), \\ \mathbb{W}_F(t) &= \mathbb{B}(F(t)), \qquad (t \in \mathbb{R}) \end{aligned}$$

を考える．このとき, 以下が成立する: 任意の有界連続函数 $h : \mathcal{K} \longrightarrow \mathbb{R}$ に対して,

$$\mathbb{E}^*[h(\mathbb{W}_n)] \longrightarrow \mathbb{E}[h(\mathbb{W}_F)], \qquad (n \to \infty)$$

が成立する．

証明　証明は後ほどする． □

注意 13.1 F は連続函数とする．このとき, $\mathbb{B}$ はほとんど確実に連続なので, $\mathbb{W}_F = \mathbb{B} \circ T$ もほとんど確実に連続である．したがって, ある意味で, $\mathbb{W}_n$ は連続的に近似されなけらばならない．実際, 任意の $t \in \mathbb{R}$ と t に収束する点列 $\{t_n\}_{n=1}^{\infty}$ に対して,

$$|\mathbb{W}_n(t_n) - \mathbb{W}_F(t)| \xrightarrow{P} 0, \qquad (n \to \infty)$$

である．これを漸近連続という．

3　ドンスカー族

$\mathcal{X}$ を距離空間または準距離空間とし, $X : \Omega \longrightarrow \mathcal{X}$ は分布 P をもつ確率要素とし, $X_1, X_2, \ldots, X_n$ を X の独立複製とする．函数 $g : \mathcal{X} \longrightarrow \mathbb{R}$ の族 $\mathcal{G}$ を考える．

函数 g の期待値は

$$Pg := \mathbb{E}[g(X)]\,(P|g| < \infty)$$

であり, $X_1, X_2, \ldots, X_n$ に基づく, g の（経験的）期待値は

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

である．P_n を経験測度という．

定義 13.2 函数族 $\mathcal{G}$ で添え字付られた経験過程を

$$\nu_n(g) \quad := \quad \sqrt{n}(P_n - P)g = \sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n} g(X_i) - \mathbb{E}[g(X)]\right\} \qquad (g \in \mathcal{G})$$

で定義する．

まず, 固定した g に対する中心極限定理を思い出そう. $g(X)$ の分散を

$$\sigma^2(g) := \mathbb{V}[g(X)] = Pg^2 - (Pg)^2$$

で記す.

$\sigma^2(g) < \infty$ のとき,

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

となる. 中心極限定理は有限個の g に対しても同時に成立する. $r \in \mathbb{N}$ に対し, $g_k, g_\ell \in \mathcal{G}\,(k, \ell = 1, 2, \ldots, r, k \neq \ell)$ は異なる函数とし, $g_k(X)$ と $g_\ell(X)$ の共分散を

$$\sigma(g_k, g_\ell) := \mathbb{C}[g_k(X), g_\ell(X)] = P(g_k g_\ell) - (Pg_k)(Pg_\ell)$$

と記す. $k = 1, 2, \ldots, r$ に対して, $\sigma^2(g_k) < \infty$ のとき,

$$\begin{pmatrix} \nu_n(g_1) \\ \nu_n(g_2) \\ \vdots \\ \nu_n(g_r) \end{pmatrix} \xrightarrow{d} N_r(\mathbf{0}, \Sigma_{g_1, g_2, \ldots, g_r}), \qquad (n \to \infty).$$

ただし, $\Sigma_{g_1, g_2, \ldots, g_r}$ は対称行列で

$$\Sigma_{g_1, g_2, \ldots, g_r} = \begin{pmatrix} \sigma^2(g_1) & \sigma(g_1, g_2) & \cdots & \sigma(g_1, g_r) \\ \sigma(g_2, g_1) & \sigma^2(g_2) & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sigma(g_r, g_1) & \cdots & \cdots & \sigma^2(g_r) \end{pmatrix} \tag{13.1}$$

定義 13.3 ν を $\mathcal{G}$ で添え字付られたガウス過程とする. 各 $r \in \mathbb{N}$ と各有限限の部分集合 $\{g_1, g_2, \ldots, g_r\} \subset \mathcal{G}$ に対して, r 次元ベクトル

$$\begin{pmatrix} \nu(g_1) \\ \nu(g_2) \\ \vdots \\ \nu(g_r) \end{pmatrix}$$

が多変量正規分布 $N_r(\mathbf{0}, \Sigma_{g_1, g_2, \ldots, g_r})$ を持つとする．ただし，$\Sigma_{g_1, g_2, \ldots, g_r}$ は (13.1) で与えられる分散共分散行列である．このとき, ν を $\mathcal{G}$ で添字付けれた P ブラウン橋という．

定義 13.4 ν_n と ν は $\mathcal{G}$ 上の有界函数とする．すなわち, $\nu_n, \nu \in \mathcal{K} = \{g : \mathcal{X} \longrightarrow \mathbb{R};\ g$ は有界$\}$ である. $\mathcal{G}$ はドンスカー族であるとは,

$$\nu_n \xrightarrow{d} \nu, \qquad (n \to \infty)$$

をみたすときをいう．すなわち, $\mathcal{K}$ 上の任意の有界連続実数値函数[1] f に対して,

$$\mathbb{E}^*[f(\nu_n)] \longrightarrow \mathbb{E}[f(\nu)], \qquad (n \to \infty)$$

をみたすことである.

定義 13.5 $\mathcal{G}$ 上の ν_n が $g_0 \in \mathcal{G}$ において，漸近連続であるとは，任意の列 $\{g_n\}_{n=1}^{\infty} \subset \mathcal{G}$ で $\sigma(g_n - g) \longrightarrow 0\,(n \to \infty)$ なるものに対して,

$$|\nu_n(g_n) - \nu(g_0)| \xrightarrow{P} 0, \qquad (n \to \infty)$$

が成り立つときをいう．これがすべての $g_0 \in \mathcal{G}$ に対して成り立つとき, ν_n は $\mathcal{G}$ 上で漸近連続であるという.

記法: $g \in L^2(P)$ に対して

$$\|g\|^2_{L^2(P)} := Pg^2$$

と記す．すなわち, $\|\cdot\|_{L^2(P)}$ は $L^2(P)$ ノルムである.

注意 13.2 $g \in L^2(P)$ に対して, $\sigma(g) = \sqrt{Pg^2 - (Pg)^2} \le \|g\|_{L^2(P)}$ である.

定義 13.6 族 $\mathcal{G}$ は全有界であるとは, $\|\cdot\|_{L^2(P)}$ によって誘導される距離に関して全有界であるときをいう．すなわち, $\mathcal{G}$ の計量エントロピー $H_2(\,\cdot\,, \mathcal{G}, \|\cdot\|_{L^2(P)})$ は有界のときである.

[1] 任意の $\epsilon > 0$ に対して, $\delta > 0$ が存在して, $\mathcal{K}$ の任意の要素 $\tilde{k}$ が

$$\sup_{x \in \mathcal{X}} |\tilde{k}(x) - k(x)| < \delta$$

ならば,

$$|f(\tilde{k}) - f(k)| < \epsilon$$

をみたすときをいう.

定理 13.4 $\mathcal{X}$ を距離空間とし, $\mathcal{X}$ 上の実数値函数のなす族 $\mathcal{G}$ は全有界とする. このとき, $\mathcal{G}$ がドンスカー族であるための必要十分条件は, $\mathcal{G}$ 上の過程として ν_n は漸近連続であることである.

証明 証明は後ほどする. $\square$

第 14 章　一様中心極限定理たち

1　連鎖

$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ に基づく対称化経験過程と $\boldsymbol{X}$ の条件付き分布を考える.

$\mathbb{P}_{\boldsymbol{X}}$ を $\boldsymbol{X}$ を与えたときの条件付き確率とする. 経験ノルムを

$$\|g\|_{L^2(P_n)} := \sqrt{P_n g^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} g^2(X_i)} \qquad (g \in \mathcal{G})$$

で定め, 経験半径 (ランダムなもの) を

$$R_n := \sup_{g\in\mathcal{G}} \|g\|_{L^2(P_n)} = \sup_{g\in\mathcal{G}} \sqrt{\frac{1}{n}\sum_{1=1}^{n} g^2(X_i)}$$

で定める.

記法を簡便化するために, 函数族 $\mathcal{G}$ は母数空間 $\theta \in \Theta$ (ここでは, 無限次元の場合もゆるしている) で添字付けられているとする:

$$\mathcal{G} := \{g_\theta;\ \theta \in \Theta\}.$$

$s = 0, 1, 2, \ldots$ に対して $\{g_j^s\}_{j=1}^{N_s}$ を半ノルム空間 $(\mathcal{G}, \|\cdot\|_{L^2(P_n)})$ の最小 $2^{-s}R_n$ 被覆とする. したがって,

$$N_s := N(2^{-s}R_n, \mathcal{G}, \|\cdot\|_{L^2(P_n)}) \qquad s = 0, 1, 2, \ldots$$

であり, 各 θ に対して, $g_\theta^s \in \{g_1^s, g_2^s, \ldots, g_{N_s}^s\}$ が存在して,

$$\|g_\theta - g_\theta^s\|_{L^2(P_n)} \le 2^{-s}R_n$$

とできる. 本来ならば,

θ を用いて, 特定の g を近似する集合の中の函数列を構築できることを示そう.

まず, $g_\theta^0 = 0$ とする. なぜならば, $\|g_\theta^0\|_{L^2(P_n)} \le R_n$ だから, いま, すべての $s = 0, 1, 2, \ldots$ に対して,

$$H_s := \log N_s$$

とおく.

$S \in \mathbb{N}$ とし,

$$g_\theta^{S+1} \in \{g_1^{S+1}, g_2^{S+1}, \ldots, g_{N_{S+1}}^{S+1}\}$$

に対して,

$$g_\theta^S : \operatorname{argmin}_{k=1, 2, \ldots, N_S}\{\|g_\theta^{S+1} - g_k^S\|_{L^2(P_n)}; g_k^S \in \{g_1^S, g_2^S, \ldots,, g_{N_S}^S\}\}$$

とし, $s \in \{0, 1, \ldots, S-1\}$ に対し,

$$g_\theta^s := \operatorname{argmin}_{k=1, 2, \ldots, N_s}\{\|g_\theta^{s+1} - g_k^s\|_{L^2(P_n)}; g_k^s \in \{g_1^s, g_2^s, \ldots,, g_{N_s}^s\}\}$$

とおく. さらに,

$$\mathcal{G}_S := \bigcup_{s=1}^{S}\{g_1^s, g_2^s, \ldots,, g_{N_s}^s\}$$

とおく. このとき,

$$g_\theta := \sum_{s=0}^{S}(g_\theta^{s+1} - g_\theta^s) + (g_\theta - g_\theta^{S+1})$$

と書ける. すわわち, g_θ と $S \in \mathbb{N}$ に対して,

$$\begin{aligned}
&\|g_\theta - g_\theta^{S+1}\|_{L^2(P_n)} \le 2^{-(S+1)}R_n, \\
&\|g_\theta^{s+1} - g_\theta^s\|_{L^2(P_n)} \le 2^{-s}R_n, \qquad (s = 0, 1, \ldots, S) \\
&g_\theta^s \in \{g_1^s, g_2^s, \ldots, g_{N_s}^s\} \qquad (s = 0, 1, \ldots, S+1)
\end{aligned}$$

をみたす列 $\{g_\theta^s\}_{s=0}^{S+1}$ を選ぶことができる.

2 対称過程と増分

$n = 1, 2, \ldots$ と $S \in \mathbb{N}$ に対して,

$$J_n := \sum_{s=0}^{S} 2^{-s} R_n \sqrt{2H_{s+1}} \tag{14.1}$$

とする. ただし,

$$\begin{aligned} R_n &= \sup_{g\in\mathcal{G}} \|g\|_{L^2(P_n)}, \\ H_2 &= \log N_s = \log N(2^{-s}R_n, \mathcal{G}, \|\cdot\|_{L^2(P_n)}) \end{aligned}$$

であった.

注意 14.1 以下の評価を得る.

$$J_n \le 2\int_{2^{-(S+2)}R_n}^{R_n} \sqrt{2H(u, \mathcal{G}, L^2(P_n))}\, du.$$

右辺を Dudley エントロピー積分と呼ばれている.

定理 14.1 $t > 0$ に対して,

$$\mathbb{P}\left\{\max\left|\sum_{s=1}^{S} P_n^{\sigma}(g_\theta^{s+1} - g_\theta^s)\right| \ge \frac{J_n}{\sqrt{n}} + 4R_n\sqrt{\frac{1+t}{n}}\right\} \le 2e^{-t}.$$

ただし, max は $\{g_1^s, g_2^s, \ldots, g_{N_s}^s \ (s = 1, 2, \ldots, S+1)\}$ の $\Pi_{1=1}^{S+1} N_s$ 個の函数に関して取ったものである.

証明　ヘフディンの不等式 (系 8.3) を用いる: $\underline{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_n)$ に関して期待値を取れば,

$$\begin{aligned} \mathbb{E}[g_\theta^{s+1} - g_\theta^s] &= 0 \\ \sup_{x\in\mathcal{X}} |g_\theta^{s+1}(x) - g_\theta^s(x)| &\le 2^{-s}R_n \end{aligned}$$

となる. したがって, $t > 0$ に対して,

$$\mathbb{P}_{\boldsymbol{X}}\left(|P_n^{\sigma}(g_\theta^{s+1} - g_\theta^s)| \ge 2^{-s}R_n\sqrt{\frac{2t}{n}}\right) \le 2e^{-t}$$

を得る. さらに, 同じ系 8.3 の後半部分より

$$\mathbb{P}_{\boldsymbol{X}}\left(|P_n^{\sigma}(g_\theta^{s+1} - g_\theta^s)| \ge 2^{-s}R_n\sqrt{\frac{2(H_{s+1}+t)}{n}}\right) \le 2e^{-t}$$

を得る.

固定した t と $s=0, 1, \ldots, S$ に対して,

$$\alpha_s := 2^{-s}R_n\{\sqrt{H_{s+1}}++\sqrt{2(1+s)(1+t)}\}$$

とする. このとき,

$$\sum_{s=0}^{S}\alpha_s \leq J_n + 4R_n\sqrt{1+t}. \tag{14.2}$$

これは,

$$\begin{aligned}
\sum_{s=0}^{S}\alpha_s &= J_n + \sum_{s=0}^{S}2^{-s}R_n\sqrt{2(1+s)(1+t)} \\
&\leq J_n + R_n\sqrt{2(1+t)}\int_0^2\sqrt{1+s}\,ds \\
&= J_n + R_n\sqrt{2(1+t)}\left[\frac{2}{3}(1+s)^{2/3}\right]_0^2 \\
&= J_n + R_n\sqrt{2(1+t)}\frac{2}{3}\{3^{3/2}-1\} \\
&= J_n + R_n\sqrt{2(1+t)}\frac{2}{3}\{3\sqrt{3}-1\} \\
&\qquad \left(\sqrt{2}\frac{2}{3}\{3\sqrt{3}-1\} \leq 4\right) \\
&\leq J_n + 4R_n\sqrt{1+t}
\end{aligned}$$

からわかる. したがって, (14.2) より

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{X}}\left(\max_{g_\theta^s\in\{g_1^s,\, g_2^s,\, \ldots,\, g_{N_{s+1}}\};\, s=0,\, 1,\, \ldots,\, S}\left|\sum_{s=0}^{S}P_n^{\sigma}(g_\theta^{s+1}-g_\theta^s)\right|\right. \\
&\qquad\qquad \left. \geq \frac{J_n+4R_n\sqrt{1+t}}{\sqrt{n}}\right) \leq 2e^{-t}.
\end{aligned} \tag{14.3}$$

を得る．これは, (14.2) より

$$
\begin{aligned}
&\mathbb{P}_{\boldsymbol{X}}\left(\max_{g_\theta^s \in \{g_1^s, g_2^s, \ldots, g_{N_{s+1}}\};\, s=0,1,\ldots,S}\left|\sum_{s=0}^{S} P_n^{\sigma}(g_\theta^{s+1} - g_\theta^s)\right| \geq \frac{J_n + 4R_n\sqrt{1+t}}{\sqrt{n}}\right)\\
&\leq \mathbb{P}_{\boldsymbol{X}}\left(\max_{g_\theta^s \in \{g_1^s, g_2^s, \ldots, g_{N_{s+1}}\};\, s=0,1,\ldots,S}\left|\sum_{s=0}^{S} P_n^{\sigma}(g_\theta^{s+1} - g_\theta^s)\right| \geq \sum_{s=0}^{S}\frac{\alpha_s}{\sqrt{n}}\right)\\
&\leq \sum_{s=0}^{S}\mathbb{P}_{\boldsymbol{X}}\left(\max_{g_\theta^s \in \{g_1^s, g_2^s, \ldots, g_{N_{s+1}}\};\, s=0,1,\ldots,S}\left|\sum_{s=0}^{S} P_n^{\sigma}(g_\theta^{s+1} - g_\theta^s)\right| \geq \frac{\alpha_s}{\sqrt{n}}\right)\\
&= \sum_{s=0}^{S}\mathbb{P}_{\boldsymbol{X}}\left\{\max_{g_\theta^s \in \{g_1^s, g_2^s, \ldots, g_{N_{s+1}}\};\, s=0,1,\ldots,S}\left|\sum_{s=0}^{S} P_n^{\sigma}(g_\theta^{s+1} - g_\theta^s)\right|\right.\\
&\qquad\qquad\qquad \left.\geq 2^{-s}R_n\left(\frac{\sqrt{2H_{s+1}} + \sqrt{(1+s)(1+t)}}{\sqrt{n}}\right)\right\}\\
&\leq 2\sum_{s=0}^{S}\exp\left(-(1+s)(1+t)\right)\\
&= 2\frac{e^{-(1+t)}\{1 - e^{-(1+s)(1+t)}\}}{1 - e^{-(1+t)}}\\
&\leq 2e^{-t}
\end{aligned}
$$

からわかる．よって, (14.3) より,

$$
\begin{aligned}
&\mathbb{P}\left(\max\left|\sum_{s=1}^{S} P_n^{\sigma}(g_\theta^{s+1} - g_\theta^s)\right| \geq \frac{J_n}{\sqrt{n}} + 4R_n\sqrt{\frac{1+t}{n}}\right)\\
&\quad = \mathbb{E}\left[\mathbb{P}_{\boldsymbol{X}}\left(\max\left|\sum_{s=1}^{S} P_n^{\sigma}(g_\theta^{s+1} - g_\theta^s)\right| \geq \frac{J_n}{\sqrt{n}} + 4R_n\sqrt{\frac{1+t}{n}}\right)\middle|\boldsymbol{X}\right]\\
&\quad \leq 2e^{-t}
\end{aligned}
$$

がわかる． □

3　脱対称化

$L^2(P)$ ノルム

$$
\|g\|_{L^2(P)} = \sqrt{Pg^2} \qquad (g \in L^2(P))
$$

であった. いま,

$$R := \sup_{g\in\mathcal{G}} \|g\|_{L^2(P_n)}$$

を函数族 $\mathcal{G}$ の半径とする.

任意の確率測度 Q に対して, $H(\,\cdot\,,\mathcal{G}, \|\cdot\|_{L^2(Q)})$ を $L^2(Q)$ ノルム $\|\cdot\|_{L^2(Q)}$ によって誘導された距離に関する計量エントロピーとする.

条件 14.1: すべての確率測度 Q に対して, ある函数 $H(u)$ が存在して,

$$\begin{aligned} H(u\|G\|_Q, \mathcal{G}, \|\cdot\|_{L^2(Q)}) &\leq H(u) \qquad (u>0), \\ \mathcal{J} &:= \int_0^1 \sqrt{2\mathcal{H}(u)}\,du < \infty \end{aligned}$$

をみたすとする[1]. このとき,

$$\mathcal{J}(\rho) := 2\int_0^\rho \sqrt{2\mathcal{H}(u)}\,du \qquad (\rho>0)$$

と定義する.

定理 14.2 条件 14.1 が成立すると仮定し,

$$G(x) = \sup_{g\in\mathcal{G}} |g(x)| \in L^2(P)$$

とする. このとき,

$$\begin{aligned} &\mathbb{P}\left\{\|P_n - P\|_{\mathcal{G}} > \frac{4\|G\|\mathcal{J}(4R\|G\|)}{\sqrt{n}} + 32R\sqrt{\frac{1+t}{n}}\right\} \\ &\qquad \leq 8e^{-t} + 4\mathbb{P}\left\{\sup_{g\in\mathcal{G}\cup\{G\}} |(P_n-P)g^2| > R^2\right\} \qquad (t>0). \end{aligned}$$

証明　自然数 S は

$$2^{-S} \leq \frac{1}{\sqrt{n}}$$

をみたす最小のものとする. このとき,

$$\|g_\theta - g_\theta^{S+1}\|_{L^2(P_n)} \leq 2^{-(S+1)}R_n \leq \frac{R_n}{2\sqrt{n}}.$$

[1] $\mathcal{J}$ の定義では, $\sqrt{2\mathcal{H}(u)}$ としていないようだ.

$\{R_n \le 2R\} \cap \{\|G\|_{L^2(P_n)} \le 2\|G\|_{L^2(P)}\}$ 上では,

$$2\int_{2^{-(S+2)}R_n}^{R_n} \sqrt{H(u, \mathcal{G}, \|\cdot\|_{L^2(P_n)})}\, du \le \|G\|_{L^2(P)} \mathcal{J}(4R\|G\|_{L^2(P)}) \quad (14.4)$$

となる. これは,

$$\begin{aligned}
&2\int_{2^{-(S+2)}R_n}^{R_n} \sqrt{H(u, \mathcal{G}, \|\cdot\|_{L^2(P_n)})}\, du \\
&\quad = 2\int_{2^{-(S+2)}R_n/\|G\|_{L^2(P_n)}}^{R_n/\|G\|_{L^2(P_n)}} \sqrt{H(u\|G\|_n, \mathcal{G}, \|\cdot\|_{L^2(P_n)})} \|G\|_n\, du \\
&\quad \le 2\|G\|_{L^2(P_n)} \int_0^{4R\|G\|_{L^2(P)}} \sqrt{2\mathcal{H}(u)}\, du \qquad \text{(変数変換の計算がうまくいかない.} \\
&\quad \le \|G\|_{L^2(P)} \mathcal{J}(4R\|G\|_{L^2(P)})
\end{aligned}$$

からわかる. $\boldsymbol{X}$ を与えたとき, (14.1) と (14.4) より,

$$\begin{aligned}
J_n &= \sum_{s=0}^{S} 2^{-s} R_n \sqrt{H_{s+1}} \\
&= \sum_{s=0}^{S} 2^{-s} R_n \sqrt{2H(2^{-(s+1)}R_n, \mathcal{G}, \|\cdot\|_{L_2(P_n)})} \\
&\le 2\int_{2^{-(S+2)R_n}}^{R_n} \sqrt{2H(u, \mathcal{G}, \|\cdot\|_{L_2(P_n)})}\, du \\
&\le \|G\|_{L^2(P)} \mathcal{J}(4R\|G\|_{L^2(P)})
\end{aligned}$$

となることと定理 14.1 より, $t > 0$ に対して, $\{R_n \le 2R\} \cap \{\|G\|_{L^2(P_n)} \le 2\|G\|_{L^2(P)}\} =: A_n$ 上では,

$$\mathbb{P}_{\boldsymbol{X}}\left(\|P_n^\sigma\|_{\mathcal{G}} \ge \frac{\|G\|_{L^2(P)} J(4R\|G\|_{L^2(P)}))}{\sqrt{n}} + 8R\sqrt{\frac{1+t}{n}} \right) \le 2e^{-t} \quad (14.5)$$

となることがわかる．これは,

$$
\begin{aligned}
&\mathbb{P}_{\boldsymbol{X}}\left(\|P_n^{\sigma}\|_{\mathcal{G}} \geq \frac{\|G\|_{L^2(P)} J(4R\|G\|_{L^2(P)}))}{\sqrt{n}} + 8R\sqrt{\frac{1+t}{n}}\right) \\
&\quad \leq \mathbb{P}_{\boldsymbol{X}}\left(\sup_{g\in\mathcal{G}} \|g - g_\theta^{S+2}\|_n + \max_{g_\theta^1,\ \dots,\ g_\theta^S \in \mathcal{G}_S} \left|\sum_{s=0}^{S} P_n^{\sigma}(g_\theta^{s+1} - g_\theta^{s})\right| \right. \\
&\qquad\qquad \left. \geq \frac{\|G\|_{L^2(P)} \mathcal{J}(4R\|G\|_{L^2(P)})}{\sqrt{n}} + 8R\sqrt{\frac{1+t}{n}}\right) \\
&\quad \leq \mathbb{P}_{\boldsymbol{X}}\left(\frac{R_n}{2\sqrt{n}} + \max_{g_\theta^1,\ \dots,\ g_\theta^S \in \mathcal{G}_S} \left|\sum_{s=0}^{S} P_n^{\sigma}(g_\theta^{s+1} - g_\theta^{s})\right| \right. \\
&\qquad\qquad \left. \geq \frac{\|G\|_{L^2(P)} \mathcal{J}(4R\|G\|_{L^2(P)})}{\sqrt{n}} + 8R\sqrt{\frac{1+t}{n}}\right) \\
&\quad \leq \mathbb{P}_{\boldsymbol{X}}\left(\frac{R}{\sqrt{n}} + \max_{g_\theta^1,\ \dots,\ g_\theta^S \in \mathcal{G}_S} \left|\sum_{s=0}^{S} P_n^{\sigma}(g_\theta^{s+1} - g_\theta^{s})\right| \right. \\
&\qquad\qquad \left. \geq \frac{\|G\|_{L^2(P)} \mathcal{J}(4R\|G\|_{L^2(P)})}{\sqrt{n}} + 8R\sqrt{\frac{1+t}{n}}\right) \\
&\quad \leq \mathbb{P}_{\boldsymbol{X}}\left(\max_{g_\theta^1,\ \dots,\ g_\theta^S \in \mathcal{G}_S} \left|\sum_{s=0}^{S} P_n^{\sigma}(g_\theta^{s+1} - g_\theta^{s})\right| \geq \frac{J_n}{\sqrt{n}} + 4R\sqrt{\frac{1+t}{n}}\right) \\
&\quad \leq 2e^{-t}
\end{aligned}
$$

からわかる．さらに,

$$
R_n = \sup_{g\in\mathcal{G}} \|g\|_n, \quad 2R = \sup_{g\in\mathcal{G}} P|g|, \quad \|G\|_{L^2(P_n)}, \quad 2\|G\|_{L^2(P)}
$$

であったことに注意する．$2R < R_n$ のとき,

$$
\begin{aligned}
\sup_{g\in\mathcal{G}\cup\{G\}} |(P_n - P)g^2| &\geq \sup_{g\in\mathcal{G}\cup\{G\}} P_n g^2 - PG^2 = R_n - PG^2 \\
&\geq 4R^2 - PG^2 \geq R^2
\end{aligned}
$$

となる．また, $\|G\|_{L^2(P_n)} > 2\|G\|_{L^2(P)}$ のとき,

$$
\sup_{g\in\mathcal{G}\cup\{G\}} |(P_n - P)g^2| \geq P_n G^2 - \|G\|_2 \geq 3\|G\|_2 \geq R^2
$$

となる. したがって,

$$A_n^c = \{R_n \le 2R \text{ かつ } \|G\|_{L^2(P_n)} \le 2\|G\|_{L^2(P)}\}^c \subset \{\sup_{g\in\mathcal{G}\cup\{G\}} |(P_n - P)g^2| > R^2\}$$

である. このことと (14.5) より

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{X}}\left(\|P_n^\sigma\|_{\mathcal{G}} \ge \frac{\|G\|_{L^2(P)} J(4R\|G\|_{L^2(P)}))}{\sqrt{n}} + 8R\sqrt{\frac{1+t}{n}}\right)\\
&\quad \le \mathbb{P}_{\boldsymbol{X}}\left(\left\{\|P_n^\sigma\|_{\mathcal{G}} \ge \frac{\|G\|_{L^2(P)} J(4R\|G\|_{L^2(P)}))}{\sqrt{n}} + 8R\sqrt{\frac{1+t}{n}}\right\} \bigcap A_n\right)\\
&\qquad + \mathbb{P}_{\boldsymbol{X}}(A_n^c)\\
&\quad \le \mathbb{P}_{\boldsymbol{X}}\left(\left\{\|P_n^\sigma\|_{\mathcal{G}} \ge \frac{\|G\|_{L^2(P)} J(4R\|G\|_{L^2(P)}))}{\sqrt{n}} + 8R\sqrt{\frac{1+t}{n}}\right\} \bigcap A_n\right)\\
&\qquad + \mathbb{P}_{\boldsymbol{X}}\left(\sup_{g\in\mathcal{G}\cup\{G\}} |(P_n - P)g^2| > R^2\right)\\
&\quad \le 2e^{-t} + \mathbb{P}_{\boldsymbol{X}}\left(\sup_{g\in\mathcal{G}\cup\{G\}} |(P_n - P)g^2| > R^2\right)
\end{aligned}$$

を得る. 系 10.1 より

$$\begin{aligned}
&\mathbb{P}\left(\|(P_n - P)g\|_{\mathcal{G}} \ge \frac{4\|G\|_{L^2(P)} \mathcal{J}(4R\|G\|_{L^2(P)})}{\sqrt{n}} + 32R\sqrt{\frac{1+t}{n}}\right)\\
&\quad \le 4\mathbb{P}\left(\|P_n^\sigma g\|_{\mathcal{G}} \ge \frac{4\|G\|_{L^2(P)} \mathcal{J}(4R\|G\|_{L^2(P)})}{\sqrt{n}} + 8R\sqrt{\frac{1+t}{n}}\right)\\
&\quad = 4\mathbb{E}\left[\mathbb{P}_{\boldsymbol{X}}\left(\|P_n^\sigma g\|_{\mathcal{G}} \ge \frac{4\|G\|_{L^2(P)} \mathcal{J}(4R\|G\|_{L^2(P)})}{\sqrt{n}} + 8R\sqrt{\frac{1+t}{n}}\right) \middle| \boldsymbol{X}\right]\\
&\quad \le 8e^{-t} + 4\mathbb{P}\left(\sup_{g\in\mathcal{G}\cup\{G\}} |(P_n - P)g^2| > R^2\right)
\end{aligned}$$

を得る. □

4　経験過程の漸近連続性

定理 14.3 条件 14.1 を仮定し, 函数族 $\mathcal{G}$ は包絡函数 G で $PG^2 < \infty$ なるものをもつとする. このとき, $\{\nu_n(g);\, g \in \mathcal{G}\}$ は漸近連続である.

証明　各 $\delta > 0$ と $g_0 \in \mathcal{G}$ に対して,

$$\mathcal{G}(\delta) := \{g \in \mathcal{G};\, \|g - g_0\| < \delta\}.$$

定理 14.2 より, $t > 0$ に対して,

$$\begin{aligned}\mathbb{P}&\left(\|P_n - P\|_{\mathcal{G}(\delta)} > \frac{8\|G\|_{L^2(P)}\mathcal{J}(4\delta\|G\|_{L^2(P)})}{\sqrt{n}} + 32\delta\sqrt{\frac{1+t}{n}}\right)\\ &\qquad \le 8e^{-t} + 4\mathbb{P}\left(\sup_{g\in\mathcal{G}(\delta)\cup\{2G+g_0\}} |(P_n - P)(g - g_0)^2| > \delta^2\right).\end{aligned}$$

となる. 任意の $\epsilon > 0$ を取り, $t := \log(\delta/\epsilon)$ とおく. $\mathcal{J}(4\delta\|G\|_{L^2(P)}) \downarrow 0\,(\delta \to 0)$ なので, ある $\delta > 0$ が存在して,

$$\begin{aligned}\mathbb{P}&\left(\sqrt{n}\|P_n - P\|_{\mathcal{G}(\delta)} > \epsilon\right)\\ &\qquad \le \epsilon + 4\mathbb{P}\left(\sup_{g\in\mathcal{G}(\delta)\cup\{2G+g_0\}} |(P_n - P)(g - g_0)^2| > \delta^2\right).\end{aligned}$$

函数族 $\{(g - g_0)^2 :\, g \in \mathcal{G}(\delta) \cup \{2G + g_0\}\}$ に対して, 大数の一様弱法則を持ちいれて,

$$\mathbb{P}\left(\sup_{g\in\mathcal{G}(\delta)\cup\{2G+g_0\}} (P_n - P)(g - g_0)^2| > \delta^2\right) \longrightarrow 0 \qquad (n \to \infty).$$

□

定理 14.4 函数族 $\mathcal{G}$ は VC 族とし, 包絡函数 G で $G \in L^2(P)$ なるものをもつとする. このとき,　$\{\nu_n(g);\, g \in \mathcal{G}\}$ は漸近連続である.

証明　定理　11.5 より, $\mathcal{G}$ の VC 次元に依存する定数 $A > 0$ が存在して, 任意の確率測度 Q に対して,

$$N_1(\delta QG,\, \mathcal{G},\, Q) \le \max\{A\delta^{-2V},\, e^{\delta/4}\}$$

となる．任意の 2 つの函数 $g, \tilde{g} \in \mathcal{G}$ に対して,

$$\int (g - \tilde{g})^2 \, dQ \leq \int |g - \tilde{g}| \, d\overline{Q}$$

となる．ただし, $d\overline{Q} := 2G \, dQ$ である．これより

$$N(\delta \|G\|_{L^2(Q)}, \mathcal{G}, \|\cdot\|_{L^2(Q)}) \leq N(\delta^2 \|G\|_{L^1(\overline{Q})}, \mathcal{G}, \|\cdot\|_{L^1(\overline{Q})}),$$

となる．ただし,

$$\|G\|_{L^2(Q)} := \sqrt{\int G^2(x) \, dQ(x)} \qquad \|G\|_{L^1(\overline{Q})} := \int |G(x)| \, d\overline{Q}$$

である． 測度 $\overline{Q}$ は確率測度でないかもしれないが, 有界である． すなわち, $\overline{Q}(\mathcal{X}) < \infty$ である．よって,

$$\begin{aligned} H(\delta \|G\|_Q, \mathcal{G}, L^2(Q)) &\leq \log N(\delta^2 \|G\|_{L^1(\overline{Q})}, \mathcal{G}, L^1(\overline{Q})) \\ &\leq \log \left(\max\{A\delta^{-2V}, e^{\delta/4}\} \right) \end{aligned}$$

で

$$\int_0^1 \sqrt{\log \left(\max\{A\delta^{-2V}, e^{\delta/4}\} \right)} \, d\delta < \infty$$

となるので, 条件 14.1 をみたす．よって, 定理は証明された． □

第 15 章　M 推定量

1　M 推定量とは

$(\Omega, \mathcal{F}, \mathbb{P})$ を確率空間とする．$\mathcal{X}$ を距離空間とし，$\mathcal{B}(\mathcal{X})$ を $\mathcal{X}$ 上の開集合を含む最小の σ 加法族とし，$X : \Omega \longrightarrow \mathcal{X}$ は分布 P をもつ確率要素とする．すなわち，

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

である．

$X_1, X_2, \ldots, X_n$ は X の独立複製とする．(Θ, d) を距離空間とし，距離関数 d によって誘導されるノルムを $\|\cdot\|_d$ と書く．ある $\theta \in \Theta$ に対し，

$$\gamma_\theta : \mathcal{X} \longrightarrow \mathbb{R}$$

をある損失函数（可測）とする．すべての $\theta \in \Theta$ に対して，

$$P|\gamma_\theta| < \infty$$

とする．

未知の母数

$$\theta_0 := \operatorname*{argmin}_{\theta \in \Theta} P\gamma_\theta \tag{15.1}$$

を M 推定量

$$\widehat{\theta}_n := \widehat{\theta}_n(X_1, X_2, \ldots, X_n) := \operatorname*{argmin}_{\theta \in \Theta} P_n\gamma_\theta$$

で推定することを考える．

<u>仮定</u>

- θ_0 は一意的に存在．
- $\widehat{\theta}_n$ は存在（一意でないかもしれない）．

例 15.1 (i) 位置母数推定. $\mathcal{X} = \mathbb{R}$, $\Theta = \mathbb{R}$ とする.

(ia) $\gamma_\theta(x) = (x - \theta)^2$.

(ib) $\gamma_\theta(x) = |x - \theta|$.

(ii) 最尤推定. $\{p_\theta;\, \theta \in \Theta\}$ は σ 有限な測度 μ に支配されている密度函数の族とし,

$$\gamma_\theta := -\log p_\theta$$

とする.

もし,

$$\frac{dP}{d\mu} = p_{\theta_0} \qquad (\theta_0 \in \Theta)$$

ならば, θ_0 は

$$\Theta \ni \theta \mapsto P\gamma_\theta = \int_{\mathcal{X}} \gamma_\theta(x) p_\theta(x)\, d\mu(x) \in [0,\, \infty)$$

を最小化する.

(iia) ポアソン分布: $\theta > 0$ に対して,

$$p_\theta(x) = e^{-\theta} \frac{\theta^x}{x!} \qquad (x = 0,\, 1,\, 2,\, \ldots)$$

(iib) ロジィステック分布: $\theta \in \mathbb{R}$ に対して,

$$p_\theta(x) = \frac{e^{\theta - x}}{(1 + e^{\theta - x})^2} \qquad (x \in \mathbb{R}).$$

2　M 推定量の一致性

各 $\theta \in \Theta$ に対して

$$\begin{aligned} R(\theta) &= P\gamma_\theta = \int_{\mathcal{X}} \gamma_\theta(x)\, dP(x), \\ R_n(\theta) &= P_n\gamma_\theta = \frac{1}{n} \sum_{i=1}^{n} \gamma_\theta(X_i) \end{aligned}$$

とする.

定義 15.1 θ_0 はうまく分離されているとは, すべての $\eta > 0$ に対して

$$\inf\{R(\theta);\, d(\theta,\, \theta_0) > \eta\} > R(\theta_0)$$

が成り立つことである.

定理 15.1

$$\sup_{\theta\in\Theta} |R(\theta) - R_n(\theta)| \xrightarrow{P} 0 \qquad (n \to \infty) \tag{15.2}$$

を仮定する. すなわち, $\{\gamma_\theta;\, \theta \in \Theta\}$ は GC 族であることを仮定する. このとき,

$$R(\widehat{\theta}_n) \xrightarrow{P} R(\theta_0) \qquad (n \to \infty).$$

さらに, θ_0 がうまく分離されていれば,

$$\widehat{\theta}_n \xrightarrow{P} \theta_0 \qquad (n \to \infty).$$

証明 θ_0 と $\widehat{\theta}_n$ の定義

$$\sup_{\theta\in\Theta} |R_n(\theta) - R(\theta)| \xrightarrow{P} 0 \qquad (n \to \infty).$$

と (15.2) に注意すれば,

$$R(\widehat{\theta}_n) \xrightarrow{P} R(\theta_0) \qquad (n \to \infty) \tag{15.3}$$

がわかる. これは,

$$\begin{aligned}
0 &\le R(\widehat{\theta}_n) - R(\theta_0) \\
&= R(\widehat{\theta}_n) - R(\theta_0) - \{R_n(\widehat{\theta}_n) - R_n(\theta_0)\} + \{R_n(\widehat{\theta}_n) - R_n(\theta_0)\} \\
&\le R(\widehat{\theta}_n) - R(\theta_0) - \{R_n(\widehat{\theta}_n) - R_n(\theta_0)\} \qquad (\because R_n(\widehat{\theta}_n) - R_n(\theta_0) \le 0) \\
&\le |R_n(\widehat{\theta}_n) - R(\widehat{\theta}_n)| + |R_n(\theta_0) - R(\theta_0)| \\
&\le 2\sup_{\theta\in\Theta} |R_n(\theta) - R(\theta)| \xrightarrow{P} 0 \qquad (n \to \infty)
\end{aligned}$$

からわかる. さらに, θ_0 がうまく分離されていれば,

$$\widehat{\theta}_n \xrightarrow{P} \theta_0 \qquad (n \to \infty).$$

なぜならば, 任意の $\epsilon > 0$ に対して, 存在する $\eta > 0$ があって,

$$d(\theta,\, \theta_0) > \eta \text{ ならば } R(\theta) - R(\theta_0) > \epsilon$$

となる. これの対偶をとれば,

$$R(\theta) - R(\theta_0) \leq \epsilon \text{ ならば } d(\theta,\, \theta_0) \leq \eta$$

となるので,

$$\mathbb{P}\left(d(\widehat{\theta}_n,\, \theta_0) \leq \eta\right) \geq \mathbb{P}\left(R(\widehat{\theta}_n) - R(\theta_0) \leq \epsilon\right) \longrightarrow 1 \qquad (n \to \infty).$$

からわかる. □

注意 15.1 仮定 (15.2) はやや厳しい条件である. なぜならば, Θ はコンパクトであることを要求しているのに近い.

補題 15.1 距離空間 $(\Theta,\, d)$ はコンパクト. 写像

$$\theta \mapsto \gamma_\theta$$

は連続[1]とする. さらに,

$$G(x) = \sup_{\theta \in \Theta} |\gamma_\theta(x)|$$

とし, $PG < \infty$ とする. このとき,

$$\sup_{\theta \in \Theta} |R_n(\theta) - R(\theta)| \xrightarrow{P} 0 \qquad (n \to \infty).$$

証明 補題 9.2 よりわかる. □

補題 15.2 X の分布を P とし, $X_1,\, X_2,\, \ldots,\, X_n\, (n \geq 1)$ を X の独立複製とする. Θ はノルム $\|\cdot\|_d$ を持つ凸集合とし, 写像

$$\Theta \ni \theta \mapsto \gamma_\theta \in \mathbb{R}$$

[1] 写像 $\theta \mapsto \gamma_\theta$ が $\theta = \theta_0$ で連続であるとは, 任意の $\epsilon > 0$ に対して, ある $\delta > 0$ が存在して, $|\theta - \theta_0| < \delta$ ならば, $\sqrt{P(\gamma_\theta - \gamma_{\theta_0})^2} < \epsilon$ とであることである.

は狭義凸とする．さらに，

$$\begin{aligned}
\theta_0 &= \operatorname*{argmin}_{\theta\in\Theta} P\gamma_\theta = \operatorname*{argmin}_{\theta\in\Theta} \int_{\mathcal{X}} \gamma_\theta(x)\,dP(x), \\
\widehat{\theta}_n &= \widehat{\theta}_n(X_1, X_2, \ldots, X_n) = \operatorname*{argmin}_{\theta\in\Theta} P_n\gamma_\theta = \operatorname*{argmin}_{\theta\in\Theta} \left\{ \frac{1}{n}\sum_{i=1}^n \gamma_\theta(X_i) \right\}
\end{aligned}$$

とする．ある $\epsilon > 0$ をとる．

- $\Theta_\epsilon = \Theta_\epsilon(\theta_0) := \{\theta \in \Theta;\ \|\theta - \theta_0\|_d \le \epsilon\}$ はコンパクト，
- $x \in \mathcal{X}$ に対して，$G_\epsilon(x) := G_{\Theta_\epsilon}(x) := \sup_{\theta\in\Theta_\epsilon} |\gamma_\theta(x)|$ としたとき，$PG_\epsilon = \int_{\mathcal{X}} G_\epsilon(x)\,dP(x) < \infty$，

を仮定する．このとき，

$$\widehat{\theta}_n \xrightarrow{P} \theta_0 \qquad (n \to \infty).$$

証明　補題 9.2 より

$$\sup_{\theta\in\Theta_\epsilon} |R_n(\theta) - R(\theta)| \xrightarrow{P} 0 \qquad (n \to \infty).$$

いま，

$$\alpha := \frac{\epsilon}{\epsilon + \|\widehat{\theta}_n - \theta_0\|_d}, \qquad \tilde{\theta}_n := \alpha\widehat{\theta}_n + (1-\alpha)\theta_0$$

とする．Θ が凸集合であることから，$\tilde{\theta} \in \Theta$ であることに注意する．さらに，

$$\begin{aligned}
\|\tilde{\theta}_n - \theta_0\|_d &\le \alpha\|\widehat{\theta}_n - \theta_0\|_d + (1-\alpha)\|\widehat{\theta}_n - \theta_0\|_d \\
&= \|\widehat{\theta}_n - \theta_0\|_d = \epsilon.
\end{aligned}$$

函数 $\Theta \ni \theta \mapsto \gamma_\theta \in \mathbb{R}$ は凸なので，函数 $\Theta \ni \theta \mapsto P\gamma_\theta \in \mathbb{R}$ も凸になることから，

$$\begin{aligned}
R_n(\widehat{\theta}_n) &\le \alpha R_n(\widehat{\theta}_n) + (1-\alpha)R_n(\theta_0) \\
&\le \alpha R_n(\theta_0) + (1-\alpha)R_n(\theta_0) = R_n(\theta_0).
\end{aligned}$$

すると，

$$R(\tilde{\theta}_n) \xrightarrow{P} R(\theta_0) \qquad (n \to \infty). \tag{15.4}$$

なぜならば,

$$
\begin{aligned}
0 &\leq R(\tilde{\theta}_n) - R(\theta_0) \\
&= R(\tilde{\theta}_n) - R(\theta_0) - \{R_n(\widehat{\theta}_n) - R_n(\theta_0)\} + \{R_n(\widehat{\theta}_n) - R_n(\theta_0)\} \\
&\leq R(\tilde{\theta}_n) - R(\theta_0) - \{R_n(\widehat{\theta}_n) - R_n(\theta_0)\} \qquad (\because R_n(\widehat{\theta}_n) - R_n(\theta_0)\} < 0) \\
&\leq |R_n(\tilde{\theta}_n) - R(\tilde{\theta}_n)| + |R_n(\theta_0) - R(\theta_0)| \\
&\leq 2 \sup_{\theta \in \Theta_\epsilon} |R_n(\theta) - R(\theta)| \xrightarrow{P} 0 \qquad (n \to \infty)
\end{aligned}
$$

よりわかる. 写像 $\Theta \ni \theta \mapsto \gamma_\theta \in [0, \infty)$ の狭義凸性と θ_0 は一意に存在することに注意する, (15.4) より,

$$\|\tilde{\theta}_n - \theta_0\|_d \xrightarrow{P} 0 \qquad (n \to \infty).$$

しかし, $0 \leq \alpha \leq 1$ なので,

$$\|\widehat{\theta}_n - \theta_0\| = \alpha \|\tilde{\theta}_n - \theta_0\|_d \xrightarrow{P} 0 \qquad (n \to \infty).$$

□

例 15.2 X は実数値確率変数とし, $\Theta = \mathbb{R}$ とする. X の分布を P と書く. ある $r \geq 1$ に対して,

$$\gamma_\theta(x) = |x - \theta|^r \qquad (x \in \mathbb{R}), \qquad \mathbb{E}[|X - \theta|^r] < \infty$$

とする. このとき, 任意の $\epsilon > 0$ に対して,

$$G_\epsilon(x) := G_{\epsilon,\, \theta_0}(x) := \sup_{|\theta - \theta_0| \leq \epsilon} |\gamma_\theta(x)| \leq 2^{r-1}\{|x - \theta_0|^r + \epsilon^r\}$$

なので,

$$G_\epsilon \in L^1(P).$$

したがって, θ_0 が一意ならば,

$$\widehat{\theta}_n \xrightarrow{P} \theta_0 \qquad (n \to \infty).$$

以上の議論を下記のように拡張できる. $p \in \mathbb{N}\,(p \geq 2)$ とする. $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$, $\boldsymbol{X}$ は $\mathbb{R}^p$ 値確率変数列ベクトル (共変量) で, Y は実数値確率変数 (応答変数) とする. $(\boldsymbol{X}, Y)$ の独立複製を考える. $r \geq 1$ とし, 損失函数を

$$\gamma_{\boldsymbol{\theta}}(\boldsymbol{x}, y) = |y - \boldsymbol{x}^\top \boldsymbol{\theta}|^r$$

とする. $\mathbb{E}[|Y - \boldsymbol{X}^\top \boldsymbol{\theta}|^r] < \infty$ であり,

$$\boldsymbol{\theta}_0 = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, P\gamma_{\boldsymbol{\theta}}, \qquad \widehat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, P_n\gamma_{\boldsymbol{\theta}},$$

とすれば,

$$\widehat{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}_0 \qquad (n \to \infty)$$

である. ただし, P_n は $(\boldsymbol{X}, Y)$ の独立複製 $(\boldsymbol{X}_i, Y_i)\,(i = 1, 2, \ldots, n)$

例 15.3 X は閉区間 $[0, 1]$ を値域にもつ確率変数で, Y は実数値確率変数で

$$Y = \theta_0(X) + \xi$$

とする. ただし, $\theta_0 : [0, 1] \longrightarrow \mathbb{R}$ は可測函数で, ξ は $N(0, \sigma^2)$ に従うとする. ここで, $\sigma^2\,(\sigma > 0)$ は未知とする. 可測函数 $\theta : [0, 1] \longrightarrow \mathbb{R}$ とし, 損失函数

$$\gamma_\theta(x, y) = (y - \theta(x))^2 \qquad (x \in [0, 1], y \in \mathbb{R})$$

を考える.

$$\Theta = \Bigg\{\theta : [0, 1] \longrightarrow \mathbb{R} \quad ; \quad \theta \text{ は } m \text{ 階連続微分可能}$$
$$\text{（その } m \text{ 階導関数を } \theta^{(m)} \text{ とかく）で} \quad \int_0^1 |\theta^{(m)}(x)|^2\,dx < 1\Bigg\}$$

とし, $\theta_0 \in \Theta$ を固定する. 集合 Θ は一様ノルム

$$\|\theta\|_\infty := \sup_{x \in [0, 1]} |\theta(x)|$$

をもつとする. このとき, 任意の $\epsilon > 0$ に対して,

$$\Theta_\epsilon := \Theta_{\epsilon, \theta_0} := \left\{\theta : [0, 1] \longrightarrow \mathbb{R}, \|\theta - \theta_0\|_\infty < \epsilon, \int_0^1 |\theta^{(m)}(x)|^2\,dx \leq 1\right\}$$

はコンパクトであることを示すことができる. θ_0 が一意的に存在（たとえば, X の密度が, 零にならない測度に関して絶対連続であればよい）すれば, 最小 2 乗推定量 $\widehat{\theta}_n$ は一様ノルムに関して一致性をもつこと知られている.

3　漸近正規性

この章の以下の節では, $\Theta \subset \mathbb{R}^p$ とする.

定義 15.2 母数 $\boldsymbol{\theta}_0$ の推定量列 $\{\widehat{\boldsymbol{\theta}}_n\}_{n=1}^{\infty}$ が漸近線形であるとは, 以下のように表現できるときをいう:

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \sqrt{n}P_n\underline{\boldsymbol{\ell}} + o_P(1).$$

ただし,

$$\underline{\boldsymbol{\ell}(x)} = \begin{pmatrix} \ell_1(x) \\ \ell_2(x) \\ \vdots \\ \ell_p(x) \end{pmatrix} : \mathcal{X} \longrightarrow \mathbb{R}^p$$

は

$$P\underline{\boldsymbol{\ell}} = \begin{pmatrix} P\ell_1 \\ P\ell_2 \\ \vdots \\ P\ell_p \end{pmatrix} = \begin{pmatrix} \int_{\mathcal{X}} \ell_1(x)\, dP(x) \\ \int_{\mathcal{X}} \ell_2(x)\, dP(x) \\ \vdots \\ \int_{\mathcal{X}} \ell_p(x)\, dP(x) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} =: \mathbf{0}$$

と $P\ell_k^2 = \int_{\mathcal{X}} \ell_k^2(x)\, dP(x) < \infty\, (k = 1, 2, \ldots, p)$ をみたす. 函数 $\underline{\boldsymbol{\ell}}$ を影響函数という. $p = 1$ のとき,

$$\sigma^2 := P\ell^2$$

を漸近分散という.

定義 15.3 $p = 1$ とする. 母数 θ_0 の 2 つの推定量列 $\{\widehat{\theta}_{n,1}\}_{n=1}^{\infty}$ と $\{\widehat{\theta}_{n,2}\}_{n=1}^{\infty}$ が漸近線形であり, 漸近分散が σ_1^2 と σ_2^2 とする. このとき,

$$e_{1,2} = \frac{\sigma_2^2}{\sigma_1^2}$$

を $\widehat{\theta}_{n,2}$ と比較した $\widehat{\theta}_{n,1}$ の漸近相対効率という.

4　M 推定量の漸近正規性のための強い十分条件

以下の条件は, 漸近正規性が成立するための十分条件である．これが, 十分条件であることを証明するのは易しいが, 厳しい条件である．次章では, よりゆるい十分条件を議論する.

- 条件 a: $\epsilon > 0$ が存在して, $|\boldsymbol{\theta} - \boldsymbol{\theta}_0| < \epsilon$ なる θ に対して, 写像

$$\boldsymbol{\theta} \mapsto \gamma_{\boldsymbol{\theta}}(x)$$

　はすべての x に対して, 微分可能で, 導函数

$$\boldsymbol{\psi}_{\boldsymbol{\theta}}(x) := \frac{\partial}{\partial \boldsymbol{\theta}} \gamma_{\boldsymbol{\theta}}(x) = \left(\frac{\partial}{\partial \theta_1} \gamma_{\boldsymbol{\theta}}(x), \ldots, \frac{\partial}{\partial \theta_p} \gamma_{\boldsymbol{\theta}}(x) \right)^\top \qquad (x \in \mathcal{X})$$

　をもつ．ここで, $\mathbb{R}^p$ の Euclid ノルムを $|\cdot|$ と書いた.

- 条件 b: $\boldsymbol{\theta} \to \boldsymbol{\theta}_0$ のとき,

$$P(\boldsymbol{\psi}_{\boldsymbol{\theta}} - \boldsymbol{\psi}_{\boldsymbol{\theta}_0}) = V(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o(1)|\boldsymbol{\theta} - \boldsymbol{\theta}_0|.$$

　ただし, V は $p \times p$ の正定値行列である.

- 条件 c: ある $\epsilon > 0$ が存在して, 族

$$\mathcal{G} =: \{\boldsymbol{\psi}_{\boldsymbol{\theta}} : |\boldsymbol{\theta} - \boldsymbol{\theta}_0| < \epsilon\}$$

　を考える．$\sqrt{n}(P_n - P)$ は $\mathcal{G}$ 上で漸近連続である．さらに,

$$\begin{aligned} \lim_{\boldsymbol{\theta} \to \boldsymbol{\theta}_0} \|\boldsymbol{\psi}_{\boldsymbol{\theta}_0} - \boldsymbol{\psi}_{\boldsymbol{\theta}}\|_{L^2(P)} &= \lim_{\boldsymbol{\theta} \to \boldsymbol{\theta}_0} \sqrt{P\{|\boldsymbol{\psi}_{\boldsymbol{\theta}_0} - \boldsymbol{\psi}_{\boldsymbol{\theta}}|^2\}} \\ &= \lim_{\boldsymbol{\theta} \to \boldsymbol{\theta}_0} \sqrt{P\{(\boldsymbol{\psi}_{\boldsymbol{\theta}_0} - \boldsymbol{\psi}_{\boldsymbol{\theta}})^\top (\boldsymbol{\psi}_{\boldsymbol{\theta}_0} - \boldsymbol{\psi}_{\boldsymbol{\theta}})\}} = 0 \end{aligned}$$

　である.

補題 15.3 条件 a, b, c を仮定する．このとき, $\widehat{\boldsymbol{\theta}}_n$ は漸近線形で, 影響函数

$$\underline{\boldsymbol{\ell}}(x) = -V^{-1} \boldsymbol{\psi}_{\boldsymbol{\theta}_0}(x)$$

をもつ．したがって，

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N_p(\mathbf{0},\, V^{-1}JV^{-1}).$$

ただし，

$$J := P[\boldsymbol{\psi}_{\boldsymbol{\theta}_0}\boldsymbol{\psi}_{\boldsymbol{\theta}_0}^{\top}].$$

証明　$\boldsymbol{\theta}_0$ は Θ の内点であり，

$$\Theta \ni \boldsymbol{\theta} \mapsto P\gamma_{\boldsymbol{\theta}} \in [0,\, \infty)$$

を最小化する．したがって，

$$P\psi_{\boldsymbol{\theta}_0} = \mathbf{0}.$$

推定量 $\widehat{\boldsymbol{\theta}}_n$ は一致性を持つので，スコア函数の解としてよい:

$$P_n\psi_{\widehat{\boldsymbol{\theta}}_n} = \mathbf{0}.$$

このスコア函数を書きかえると

$$\begin{aligned}
\mathbf{0} &= P_n\psi_{\widehat{\boldsymbol{\theta}}_n} = P_n(\psi_{\widehat{\boldsymbol{\theta}}_n} - \psi_{\boldsymbol{\theta}_0}) + P_n\psi_{\boldsymbol{\theta}_0} \\
&= (P_n - P)(\psi_{\widehat{\boldsymbol{\theta}}_n} - \psi_{\boldsymbol{\theta}_0}) + P\psi_{\widehat{\boldsymbol{\theta}}_n} - P\psi_{\boldsymbol{\theta}_0} + P_n\psi_{\boldsymbol{\theta}_0} \\
&= (P_n - P)(\psi_{\widehat{\boldsymbol{\theta}}_n} - \psi_{\boldsymbol{\theta}_0}) + P\psi_{\widehat{\boldsymbol{\theta}}_n} + P_n\psi_{\boldsymbol{\theta}_0}. \qquad (\because P\psi_{\boldsymbol{\theta}_0} = \mathbf{0})
\end{aligned}$$

いま，条件 b と $\{\psi_{\boldsymbol{\theta}} : |\boldsymbol{\theta} - \boldsymbol{\theta}_0| \le \epsilon\}$ の漸近連続性より

$$\begin{aligned}
\mathbf{0} &= (P_n - P)(\psi_{\widehat{\boldsymbol{\theta}}_n} - \psi_{\boldsymbol{\theta}_0}) + P(\psi_{\widehat{\boldsymbol{\theta}}_n} - \psi_{\boldsymbol{\theta}_0}) + P_n\psi_{\widehat{\boldsymbol{\theta}}_n} \\
&= o_P(n^{-1/2}) + V(\psi_{\widehat{\boldsymbol{\theta}}_n} - \psi_{\boldsymbol{\theta}_0}) + o(|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0|) + P_n\psi_{\boldsymbol{\theta}_0}
\end{aligned}$$

となる．よって，

$$\begin{aligned}
o_P(1) &= V\bigl(\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\bigr) + o\bigl(\sqrt{n}|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0|\bigr) + \sqrt{n}P_n\psi_{\boldsymbol{\theta}_0} \\
&= o_P(1) + V\bigl(\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\bigr) + o\bigl(\sqrt{n}|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0|\bigr) + O_p(1)
\end{aligned}$$

となるので，

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_P(1).$$

したがって,

$$\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = -V\left\{\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\right\} + o_P(n^{-1/2})$$

がわかる. 中心極限定理と $P\boldsymbol{\psi}_{\boldsymbol{\theta}_0} = \mathbf{0}$ より,

$$\sqrt{n}P_n\psi_{\boldsymbol{\theta}_0} \xrightarrow{d} N\Bigl(\mathbf{0},\, P\bigl(\psi_{\boldsymbol{\theta}_0}\psi_{\boldsymbol{\theta}_0}^\top\bigr)\Bigr)$$

なので,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N\bigl(\mathbf{0},\, V^{-1}JV^{-1}\bigr).$$

□

例 15.4 Huber 型推定量の漸近分布. $\mathcal{X} = \mathbb{R}$, $\Theta = \mathbb{R}$ とし, $0 < k < \infty$ はある固定された定数（統計学者が選択するもの）とする. Huber 型損失函数

$$\begin{aligned} \gamma_\theta &= \gamma(x-\theta), \\ \gamma(x) &= x^2\mathbb{1}\{|x| \le k\} + (2k|x| - k^2)\mathbb{1}\{|x| > k\}, \qquad (x \in \mathbb{R}) \end{aligned}$$

に対応する推定量を Huber 型推定量とよぶ.

これが, 条件 a, b, c をみたすことを確認する.

<u>条件 a</u>:

$$\psi_\theta(x) = \begin{cases} 2k & (x-\theta < -k), \\ -2(x-\theta) & (|x-\theta| \le k), \\ -2k & (x-\theta > k). \end{cases}$$

<u>条件 b</u>:

$$\frac{d}{d\theta}\int \psi_\theta(x)\,dF(x) = 2\{F(k+\theta) - F(-k+\theta)\}.$$

ただし, $F(t) = \mathbb{P}(X \le t)\,(t \in \mathbb{R})$ は X の分布函数である. なぜならば, $|\psi_\theta(x)| \le k$ なので, 定理 3.8 を適用して,

$$\begin{aligned} \frac{d}{d\theta}\int \psi_\theta(x)\,dF(x) &= \int \frac{d}{d\theta}\psi_\theta(x)\,dF(x) = 2\int_{-k+\theta}^{k+\theta} dF(x) \\ &= 2\{F(k+\theta) - F(-k+\theta)\} \end{aligned}$$

よりわかる. したがって,

$$V = 2\{F(k+\theta_0) - F(-k+\theta_0)\}.$$

条件 c: 明らかに $\{\psi_\theta;\, \theta \in \mathbb{R}\}$ はVC グラフ族で, 包絡函数 $\Psi = \sup_{x\in\mathbb{R}} |\psi_\theta(x)|$ は $\Psi \le 2k$ となり, 有界.

以上より, 漸近正規性は補題 15.3 より従う.

例 15.5 確率変数 X の分布函数 F は中央値 θ_0 で微分可能とし, 中央値の漸近理論を述べる. $X_1, X_2, \ldots, X_n$ は X の独立複製とし,

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

とする. 標本中央値は Huber 型推定量の極限の場合（$k \to \infty$）とみなすことができる. しかし,

$$\gamma_\theta(x) = |x - \theta|$$

は $x = \theta$ で微分できないので, 条件 a をみたさない. そのため, 少し工夫が必要となる.

簡単のために, 標本数は奇数とする. 標本中央値 $\widehat{\theta}_n$ は, スコア方程式

$$\widehat{F}_n(\widehat{\theta}_n) - \frac{1}{2} = 0$$

をみたす. このとき,

$$\begin{aligned} 0 &= \widehat{F}_n(\widehat{\theta}_n) - F(\theta_0) \\ &= [\widehat{F}_n(\widehat{\theta}_n) - F(\widehat{\theta}_n)] + [F(\widehat{\theta}_n) - F(\theta_0)] \\ &= \frac{1}{\sqrt{n}} W_n(\widehat{\theta}_n) + [F(\widehat{\theta}_n) - F(\theta_0)] \end{aligned}$$

をえる. ただし,

$$W_n(x) = \sqrt{n}(\widehat{F}_n(x) - F(x))$$

は経験過程である. 分布函数 F は θ_0 で連続であり, $\mathbb{E}[|X - \theta|] < \infty$ のとき, 例 15.2 より,

$$\widehat{\theta}_n \xrightarrow{P} \theta_0 \qquad (n \to \infty)$$

となる. 半閉区間の集合族 $\{(-\infty, r];\, r \in \mathbb{R}\}$ は VC 族なので, 経験過程 W_n は漸近連続性をもつので,

$$W_n(\widehat{\theta}_n) = W_n(\theta_0) + o_P(1)$$

となる．したがって，

$$
\begin{aligned}
0 &= W_n(\widehat{\theta}_n) + \sqrt{n}(F(\widehat{\theta}_n) - F(\theta_0)) \\
&= W_n(\theta_\theta) + \sqrt{n}\{f(\theta_0) + o(1)\}(\widehat{\theta}_n - \theta_0) + o_P(1)
\end{aligned}
$$

である．ただし，

$$
f(\theta) = \frac{d}{d\theta}F(\theta).
$$

これより，

$$
\sqrt{n}(\widehat{\theta}_n - \theta_0) = -\frac{W_n(\theta_0)}{f(\theta_0)} + o_P(1).
$$

よって，

$$
\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} N(0,\, \sigma^2)
$$

となる．ただし，漸近分散は

$$
\sigma^2 = \frac{1}{\{f(\theta_0)\}^2}\mathbb{V}[W_n(\theta_0)] = \frac{1}{4\{f(\theta_0)\}^2}
$$

となる．なぜならば，

$$
\begin{aligned}
\mathbb{V}[W_n(\theta_0)] &= \mathbb{V}\left[\sqrt{n}\frac{1}{n}\sum_{i=1}^{n}(\mathbb{1}_{(-\infty,\,\theta_0]}(X_i) - F(\theta_0))\right] \\
&= \mathbb{V}[\mathbb{1}_{(-\infty,\,\theta_0]}(X_1)] = F(\theta_0)\{1 - F(\theta_0)\} \\
&= \mathbb{E}[\{\mathbb{1}_{(-\infty,\,\theta_0]}(X_1)\}^2] - \{\mathbb{E}[\mathbb{1}_{(-\infty,\,\theta_0]}(X_1)]\}^2 \\
&= F(\theta_0) - \{F(\theta_0)\}^2 = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4}
\end{aligned}
$$

よりわかる．

5 M 推定量の漸近正規性のためのより弱い十分条件

損失函数 $\gamma_{\boldsymbol{\theta}}$ の微分可能性についての条件をゆるめる．

- 条件 aa: 函数

$$
\begin{aligned}
&\boldsymbol{\psi}_{\boldsymbol{\theta}_0,\,i} : \mathcal{X} \longrightarrow \mathbb{R} \qquad (i = 1,\, 2,\, \ldots,\, p), \\
&\boldsymbol{\psi}_{\boldsymbol{\theta}_0} = (\boldsymbol{\psi}_{\boldsymbol{\theta}_0,\,1},\, \boldsymbol{\psi}_{\boldsymbol{\theta}_0,\,2},\, \ldots,\, \boldsymbol{\psi}_{\boldsymbol{\theta}_0,\,p})^\top
\end{aligned}
$$

は $P\boldsymbol{\psi}_{0,i}^2 < \infty$ で

$$\lim_{\boldsymbol{\theta}\to\boldsymbol{\theta}_0} \frac{\|\gamma_{\boldsymbol{\theta}} - \gamma_{\boldsymbol{\theta}_0} - (\boldsymbol{\theta}-\boldsymbol{\theta}_0)^\top \boldsymbol{\psi}_{\boldsymbol{\theta}_0}\|_{L^2(P)}}{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|} \longrightarrow 0$$

が成立する.

- 条件 bb: $\boldsymbol{\theta} \to \boldsymbol{\theta}_0$ のとき,

$$P(\gamma_{\boldsymbol{\theta}} - \gamma_{\boldsymbol{\theta}_0}) = \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^\top V(\boldsymbol{\theta}-\boldsymbol{\theta}_0) + o(1)|\boldsymbol{\theta}-\boldsymbol{\theta}_0|^2.$$

ただし, V は $p \times p$ の正値対称行列である.

- 条件 cc:

$$g_{\boldsymbol{\theta}}(x) = \begin{cases} \frac{\gamma_{\boldsymbol{\theta}}(x)-\gamma_{\boldsymbol{\theta}_0}(x)-(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^\top\boldsymbol{\psi}_{\boldsymbol{\theta}_0}(x)}{|\boldsymbol{\theta}-\boldsymbol{\theta}_0|} & (\boldsymbol{\theta} \neq \boldsymbol{\theta}_0) \\ 0 & (\boldsymbol{\theta} = \boldsymbol{\theta}_0) \end{cases}$$

と定義する. ある $\epsilon > 0$ が存在して, 函数族

$$\mathcal{G} := \{g_{\boldsymbol{\theta}};\, 0 < |\boldsymbol{\theta}-\boldsymbol{\theta}_0| < \epsilon\}$$

を考える. $\sqrt{n}(P_n - P)$ は $\mathcal{G}$ 上で漸近連続とする.

補題 15.4 条件 aa, bb, cc が成立すると仮定する. このとき, $\widehat{\boldsymbol{\theta}}_n$ は影響函数

$$\boldsymbol{\ell}(x) = -V^{-1}\boldsymbol{\psi}_{\boldsymbol{\theta}_0}$$

をもつ. したがって,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N_p(\mathbf{0},\, V^{-1}JV^{-1})$$

が成立する. ただし, $J = P(\boldsymbol{\psi}_{\boldsymbol{\theta}_0}\boldsymbol{\psi}_{\boldsymbol{\theta}_0}^\top)$ である.

証明 $\epsilon > 0$ に対して, $\{g_{\boldsymbol{\theta}} : |\boldsymbol{\theta}-\boldsymbol{\theta}_0| \le \epsilon\}$ は漸近連続なので, $\boldsymbol{\theta} \to \boldsymbol{\theta}_0$ のとき,

$$P_n(\gamma_{\boldsymbol{\theta}} - \gamma_{\boldsymbol{\theta}_0}) = \frac{1}{2}\left|V^{1/2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0) + o\big(|\boldsymbol{\theta}-\boldsymbol{\theta}_0|\big) + o_P(n^{-1/2})\right|^2. \qquad (15.5)$$

を得る．これは,

$$
\begin{aligned}
P_n(\gamma_{\boldsymbol{\theta}} - \gamma_{\boldsymbol{\theta}_0}) &= (P_n - P)(\gamma_{\boldsymbol{\theta}} - \gamma_{\boldsymbol{\theta}_0}) + P(\gamma_{\boldsymbol{\theta}} - \gamma_{\boldsymbol{\theta}_0}) \\
&= (P_n - P)\big\{g_{\boldsymbol{\theta}}|\boldsymbol{\theta} - \boldsymbol{\theta}_0| + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \boldsymbol{\psi}_{\boldsymbol{\theta}_0}\big\} + P(\gamma_{\boldsymbol{\theta}} - \gamma_{\boldsymbol{\theta}_0}) \\
&= (P_n - P)g_{\boldsymbol{\theta}}|\boldsymbol{\theta} - \boldsymbol{\theta}_0| + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top P_n \boldsymbol{\psi}_{\boldsymbol{\theta}_0} - (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top P \boldsymbol{\psi}_{\boldsymbol{\theta}_0} \\
&\qquad + P(\gamma_{\boldsymbol{\theta}} - \gamma_{\boldsymbol{\theta}_0}) \\
&= o_P(n^{-1/2}|\boldsymbol{\theta} - \boldsymbol{\theta}_0| + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top P_n \boldsymbol{\psi}_{\boldsymbol{\theta}_0} \\
&\qquad + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top V(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o\big(|\boldsymbol{\theta} - \boldsymbol{\theta}_0|^2\big) \\
&= \frac{1}{2}\Big|V^{1/2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + o\big(|\boldsymbol{\theta} - \boldsymbol{\theta}_0|\big) + o_P(n^{-1/2})\Big|^2
\end{aligned}
$$

からわかる．

$$
P_n(\gamma_{\widehat{\boldsymbol{\theta}}_n} - \gamma_{\boldsymbol{\theta}_0}) \le 0
$$

なので, $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_n$ に対して, (15.2) を用いれば,

$$
|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0| = O_p(n^{-1/2}).
$$

さらに, $\tilde{\boldsymbol{\theta}}_n := \boldsymbol{\theta}_0 - V^{-1/2} P_n \boldsymbol{\psi}_{\boldsymbol{\theta}_0}$ に対して, (15.2) を用いれば,

$$
\begin{aligned}
P_n(\gamma_{\tilde{\boldsymbol{\theta}}_n} - \gamma_{\boldsymbol{\theta}_0}) &= o_P(n^{-1/2})|V^{-1/2} P_n \boldsymbol{\psi}_{\boldsymbol{\theta}_0}| - (P_n \boldsymbol{\psi}_0)^\top V^{-1}(P_n \boldsymbol{\psi}_{\boldsymbol{\theta}_0}) \\
&\qquad + \frac{1}{2}(P_n \boldsymbol{\psi}_0)^\top V^{-1}(P_n \boldsymbol{\psi}_0) + o(|V^{-1/2} P_n \boldsymbol{\psi}_{\boldsymbol{\theta}_0}|) \\
&= -\frac{1}{2}\big|V^{-1/2} P_n \boldsymbol{\psi}_{\boldsymbol{\theta}_0}\big|^2 + o_P(n^{-1})
\end{aligned}
$$

ここで,

$$
P_n(\gamma_{\widehat{\boldsymbol{\theta}}_n} - \gamma_{\boldsymbol{\theta}_0}) \le P_n(\gamma_{\tilde{\boldsymbol{\theta}}_n} - \gamma_{\boldsymbol{\theta}_0})
$$

に注意すれば,

$$
\begin{aligned}
&O_P(n^{-1/2})|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0| + (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top P_n \boldsymbol{\psi}_{\boldsymbol{\theta}_0} + \frac{1}{2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top V(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top \\
&+ o(|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0|^2) \\
&\le -\frac{1}{2}\big|V^{-1/2} P_n \boldsymbol{\psi}_{\boldsymbol{\theta}_0}\big|^2 + o_P(n^{-1/2})
\end{aligned}
$$

より

$$\left|V^{1/2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\right|^2 = o_P(n^{-1}). \tag{15.6}$$

これは

$$\begin{aligned}
& o_P(n^{1/2})O_P(n^{-1/2}) + (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top P_n \boldsymbol{\psi}_{\boldsymbol{\theta}_0} \\
& \quad + \frac{1}{2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top V(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_P(n^{-1}) \leq -\frac{1}{2}\left|V^{-1/2}P_n \boldsymbol{\psi}_{\boldsymbol{\theta}_0}\right|^2 + o_P(n^{-1}) \\
\Longleftrightarrow\ & o_P(n^{-1}) + (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top P_n \boldsymbol{\psi}_{\boldsymbol{\theta}_0} + \frac{1}{2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top V(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\
& \qquad + o_P(n^{-1}) \leq -\frac{1}{2}\left|V^{-1/2}P_n \boldsymbol{\psi}_{\boldsymbol{\theta}_0}\right|^2 + o_P(n^{-1}) \\
\Longleftrightarrow\ & o_P(n^{-1}) + (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top P_n \boldsymbol{\psi}_{\boldsymbol{\theta}_0} + \frac{1}{2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top V(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\
& \qquad + o_P(n^{-1}) \leq -\frac{1}{2}\left|V^{-1/2}P_n \boldsymbol{\psi}_{\boldsymbol{\theta}_0}\right|^2 + o_P(n^{-1}) \\
\Longleftrightarrow\ & (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top P_n \boldsymbol{\psi}_{\boldsymbol{\theta}_0} + \frac{1}{2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top V(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\
& \qquad + o_P(n^{-1}) \leq -\frac{1}{2}\left|V^{-1/2}P_n \boldsymbol{\psi}_{\boldsymbol{\theta}_0}\right|^2 + o_P(n^{-1}) \\
\Longleftrightarrow\ & \left|V^{1/2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\right|^2 = o_P(n^{-1})
\end{aligned}$$

よりわかる. したがって, (15.6) より

$$V^{1/2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = -V^{-1/2}P_n \boldsymbol{\psi}_{\boldsymbol{\theta}_0} + o_P(n^{-1/2})$$

なので,

$$\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = -V^{-1}P_n \boldsymbol{\psi}_{\boldsymbol{\theta}_0} + o_P(n^{-1/2})$$

より補題は示された. □