経験過程理論の紹介

今野 良彦
千葉大学大学院自然科学研究科

はじめに

このノートは経験過程理論の紹介とその応用をまとめたものである.

目 次

1

準備：基本的な確率不等式

·····1.1·····

記号

$(\Omega, \mathscr{F}, \mathbb{P})$ を確率空間とする．ここで，Ω を集合，$\mathscr{F}$ を σ - 集合体，$\mathbb{P}$ を確率測度とする．$\mathscr{F}$ の元を事象と呼ぶ．また，Ω の元を ω と記すことにする．ある性質がほとんど確実に起こるとき，$a.s.$ と記す．Ω から $\mathbb{R}$ への実数値可測関数を確率変数と呼ぶ．$A \subset \Omega$ に対して，$A^c = \{\omega \in \Omega : \omega \notin A\}$ とする．また，$B \subset \Omega$ のとき，$B \setminus A = B \cap A^c$ とする．

確率変数 X に対して，その期待値を

$$\mathbb{E}[X] = \int X \, d\mathbb{P}, \qquad \mathbb{E}[X;\, A] = \int_A X \, d\mathbb{P}$$

と記す．1_A を指示関数とする．すなわち，

$$1_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \text{その他} \end{cases}$$

とする．また，$\{\omega : X(\omega) > a\}$ を $\{X > a\}$ と書く．与えられた確率変数に対し，$\mathbb{R}$ 上の確率測度を

$$\mathbb{P}_X(A) = \mathbb{P}\{\omega : X(\omega) \in A\}, \qquad A \subset \mathbb{R}$$

で定める．$\mathbb{P}_X$ を X の法則または分布と呼ぶ．特に，X の累積分布関数を

$$F_X(x) = \mathbb{P}\{(-\infty, x]\} = \mathbb{P}\{\omega \in \Omega : X(\omega) \le x\}, \qquad x \in \mathbb{R}$$

で定義する．

·····1.2·····

基本的な確率不等式

命題 1.1 非負確率変数 $X \geq 0,\ a.s.$ と正数 $p > 0$ に対して,

$$\mathbb{E}X^p = \int_0^\infty pt^{p-1}\mathbb{P}(X > t)\,dt$$

が成立する.

証明 Fubini の定理を用いて変形すれば,

$$\begin{aligned}\int_0^\infty pt^{p-1}\mathbb{P}(X > t)\,dt &= \int_0^\infty pt^{p-1}\mathbb{E}1_{(t,\,\infty)}(X)\,dt = \mathbb{E}\int_0^\infty pt^{p-1}1_{(t,\,\infty)}(x)\,dt \\ &= \mathbb{E}\int_0^X pt^{p-1}\,dt = \mathbb{E}X^p\end{aligned}$$

から命題は証明される. □

命題 1.2 （Markov の不等式） $X \geq 0,\ a.s.$ とする．任意の $a > 0$ に対して,

$$\mathbb{P}(X \geq a) = \frac{\mathbb{E}X}{a}$$

が成立する.

証明 $X \in [a,\,\infty)$ のとき，$X/a \geq 1$ に注意する：

$$\mathbb{P}\{X \geq a\} = \mathbb{E}[1_{[a,\,\infty)}(X)] \leq \mathbb{E}[\frac{X}{a}1_{[a,\,\infty)}(X)] \leq \frac{\mathbb{E}X}{a}.$$

□

注意 1.1 Y を確率変数とする．$X = (Y - \mathbb{E}Y)^2$ とおけば，Chebyshev の不等式

$$\mathbb{P}\{|Y - \mathbb{E}Y| \geq a\} = \mathbb{P}\{(Y - \mathbb{E}Y)^2 \geq a^2\} \leq \frac{\mathbb{VAR}(Y)}{a^2}$$

を得る.

命題 1.3 （Jensen の不等式） g は上に凸とし，X と $g(X)$ は可積分とする．このとき,

$$g(\mathbb{E}X) \leq \mathbb{E}g(X)$$

が成立する.

証明 任意の $x_0 \in \mathbb{R}$ とある定数 c に対し，$g(x) \geq g(x_0) + c(x - x_0)$ が成立する．$x = X(\omega)$ として，期待値をとれば，$\mathbb{E}g(x) \geq g(x_0) + c(\mathbb{E}X - x_0)$ を得る．さらに，$x_0 = \mathbb{E}X$ とすれば，命題は証明される. □

·····1.3·····

基本不等式の応用

補題 1.1 $Y_1, Y_2 \ldots, Y_n (n \geq 2)$ を実数値確率変数とし，ある定数 $\sigma > 0$ とすべての $s > 0$ に対し，

$$\mathbb{E}e^{sY_i} \leq e^{s^2\sigma^2/2}, \qquad i = 1, 2, \ldots n \tag{1.1}$$

を満足するとする．このとき，

$$\mathbb{E}\{\max_{1 \leq i \leq n} Y_i\} \leq \sigma\sqrt{2\log n}$$

が成立する．さらに，

$$\mathbb{E}e^{-sY_i} \leq e^{s^2\sigma^2/2}, \qquad i = 1, 2, \ldots n$$

も成立すれば，

$$\mathbb{E}\{\max_{1 \leq i \leq n} Y_i\} \leq \sigma\sqrt{2\log(2n)}$$

が成立する．

証明 Jensen の不等式を用いる：すべての $s > 0$ に対して，

$$\exp[\mathbb{E}\{s \max_{1 \leq i \leq n} Y_i\}] \leq \mathbb{E}[\exp\{s \max_{1 \leq i \leq n} Y_i\}] \leq \sum_{i=1}^{n} \mathbb{E}e^{sY_i} \leq ne^{s^2\sigma^2/2}$$

を得る[1)]．したがって，

$$\mathbb{E}\{\max_{1 \leq i \leq n} Y_i\} \leq \frac{\log n}{s} + \frac{s\sigma^2}{2}$$

となる．$s = \sqrt{(2/\sigma^2)\log n}$ のとき，上式の右辺は最小となることから補題は証明される． □

補題 1.2 定数 $a, b\,(a < 0 < b)$ に対し，$a \leq X \leq b,\ a.s.$ とし，$\mathbb{E}X = 0$ が成立するとする．このとき，任意の $t > 0$ に対して，

$$\mathbb{E}e^{tX} \leq \exp\left[\frac{t^2}{8}(b-a)^2\right]$$

が成立する．

証明 任意の $t > 0$ に対して，

$$e^{tX} \leq e^{ta}\frac{b-X}{b-a} + e^{tb}\frac{X-a}{b-a}$$

が成立する．$\mathbb{E}X = 0$ から

$$\mathbb{E}e^{tX} \leq e^{ta}\frac{b}{b-a} - e^{tb}\frac{a}{b-a} \tag{1.2}$$

となる．ここで，

$$\alpha = 1 - \beta = \frac{-a}{b-a}, \qquad u = t(b-a)$$

とおけば，$a < 0 < b$ から $\alpha > 0$，$\beta > 0$ ，$\alpha + \beta = 1$ となる．また，(1.2) の両辺に対数をとれば，

$$\log\mathbb{E}e^{tX} \leq \log\{\beta e^{-\alpha u} + \alpha e^{\beta u}\} = -\alpha u + \log(\beta + \alpha e^u) \equiv f(u)$$

---

1) $\exp\{s\max Y_i\} \leq \sum \exp sY_i$ に注意せよ．

となる．$f(u)$ を u に関して微分すれば，

$$f'(u) = -\alpha + \frac{\alpha e^u}{\beta + \alpha e^u} = -\alpha + \frac{\alpha}{\alpha + \beta e^{-u}},$$

$$f''(u) = \frac{\alpha\beta e^{-u}}{(\alpha + \beta e^{-u})^2} = \frac{\alpha}{\alpha + \beta e^{-u}} \cdot \frac{\beta e^{-u}}{\alpha + \beta e^{-u}} \leq \frac{1}{4}$$

となる．テーラー展開から

$$f(u) = f(0) + uf'(0) + \frac{1}{2}u^2 f''(u^*)$$

を得る．ただし，u^* は 0 と u の間の数である．$\alpha + \beta = 1$ から $f'(0) = 0$ になるので，

$$f(u) \leq \frac{1}{8}u^2 = \frac{t^2}{8}(b-a)^2$$

となり補題は証明された． □

補題 1.3 定数 $a, b\,(a < 0 < b)$ に対し，$a \leq X \leq b,\ a.s.$ とし，$\mathbb{E}X = 0$ が成立するとする．このとき，任意の $s > 0$ に対して，

$$\mathbb{P}\{X \geq s\} \leq \exp\left[-\frac{2s^2}{(b-a)^2}\right]$$

が成立する．

証明 Markov の不等式と補題 1.2 を用いる：任意の $t > 0$ に対して，

$$\mathbb{P}\{X \geq s\} = \mathbb{P}\{e^{tX} \geq e^{st}\} \leq e^{-st}\mathbb{E}e^{tX} \leq \exp\left[-st + \frac{t^2}{8}(b-a)^2\right]$$

となる．いま，

$$f(t) = -st + \frac{t^2}{8}(b-a)^2$$

は $t = 4s/(b-a)^2$ のとき，最小値 $f(4s/(b-a)^2) = -2s^2/(b-a)^2$ をとることから補題は証明された． □

補題 1.4 $C > 1$，$K > 0$ とする．非負確率変数 X は，すべての $t > 0$ に対し，

$$\mathbb{P}\{X > t\} \leq Ce^{-Kt} \tag{1.3}$$

を満足する．このとき，

$$\mathbb{E}X \leq \frac{1 + \log C}{K}$$

が成立する．

証明 $u > 0$ とする．命題 1.1 を用いる：

$$\begin{aligned}\mathbb{E}X &= \int_0^\infty \mathbb{P}\{X > t\}\,dt = \int_0^u \mathbb{P}\{X > t\}\,dt + \int_u^\infty \mathbb{P}\{X > t\}\,dt \leq u + C\int_u^\infty e^{-Kt}\,dt \\ &= u + \frac{C}{K}e^{-Ku}\end{aligned}$$

となり，上式の最右辺は $u = K^{-1}\log C$ のときに最小値 $K^{-1}(1 + \log C)$ をとることから補題は証明される． □

補題 1.5 $C > 1$，$K > 0$ とする．非負確率変数 X は，すべての $t > 0$ に対し，

$$\mathbb{P}\{X > t\} \leq Ce^{-Kt^2} \tag{1.4}$$

を満足する．このとき，

$$\mathbb{E}X \leq \sqrt{\frac{1 + \log C}{K}}$$

が成立する．

証明　(1.4) と補題 1.4 と同様な議論から，任意の $u > 0$ に対して，

$$\begin{aligned}\mathbb{E}X^2 &= \int_0^u \mathbb{P}\{X^2 > t\}\,dt + \int_u^\infty \mathbb{P}\{X^2 > t\}\,dt \le u + \int_u^\infty \mathbb{P}\{X > \sqrt{t}\}\,dt \\ &\le u + C\int_u^\infty e^{-Kt}\,dt\end{aligned}$$

を得る．したがって，上式の最右辺は $u = K^{-1}\log C$ のとき，最小値 $K^{-1}\log(1+C)$ を取ることおよび Cauchy - Schwarz の不等式から $\mathbb{E}X \le \sqrt{\mathbb{E}X^2}$ から補題は証明される． □

·····1.4·····

指数上限と Azuma の不等式

定理 1.1 $Z_1, Z_2, \ldots, Z_n$ を互いに独立な確率変数列で $\mathbb{E}Z_i = 0 (i = 1, 2, \ldots, n)$ とし，$a_i \le Z_i \le b_i$ を満足する．ただし，$a_i, b_i (a_i < b_i)$ は定数とする．このとき，すべての $t > 0$ に対して，

$$\mathbb{P}\left\{\sum_{i=1}^n Z_i > t\right\} \le \exp\left\{-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right\}$$

を満足する．

証明 Markov の不等式を用いる：各 $s > 0$ に対して，

$$\begin{aligned}
\mathbb{P}\{\sum_{i=1}^n Z_i > t\} &= \mathbb{P}\{\exp[s\sum_{i=1}^n Z_i > e^{st}\} \\
&\le e^{-st}\mathbb{E}[\exp\{s\sum_{i=1}^n Z_i\}] \\
&= e^{-st}\Pi_{i=1}^n \mathbb{E}[\exp(sZ_i)] \\
&\le e^{-st}\Pi_{i=1}^n \exp\left[\frac{s^2}{8}(b_i - a_i)^2\right]
\end{aligned}$$

となる．最後の不等式は補題 1.2 からわかる．上式の右辺は $s = 4t/\sum_{i=1}^n (b_i - a_i)^2$ のとき，最小値 $\exp[-st^2/\sum_{i=1}^n (b_i - a_i)^2]$ をとることより定理は示された． □

定理 1.2 確率変数 X は確率測度 P をもち，$f \in L_1(P)$ とする．$\{(d_i, \mathscr{M}_i\}_{i=1}^n\}$ はマルチンゲール差とする．すなわち，$\mathbb{E}(d_i|\mathscr{M}_{i-1}) = 0, i = 1, 2, \ldots, n$ である．さらに，$f - Pf = \sum_{i=1}^n d_i$ と書け，$\|d_i\|_\infty < \infty$ とする．また，$a^2 = \sum_{i=1}^n \|d_i\|_\infty^2$ とおく．このとき，すべての $t > 0$ に対して，

$$\begin{aligned}
\mathbb{P}\{f(X) - Pf > t\} &\le \exp\left\{-\frac{t^2}{2a^2}\right\}, \\
\mathbb{P}\{-(f(X) - Pf) > t\} &\le \exp\left\{-\frac{t^2}{2a^2}\right\}, \\
\mathbb{P}\{|f(X) - Pf| > t\} &= 2\exp\left\{-\frac{t^2}{2a^2}\right\}
\end{aligned}$$

が成立する．

証明 $-1 \le d_i/\|d_i\|_\infty < 1$ に注意して，補題 1.2 を用いれば，任意の $s > 0$ に対して，

$$\begin{aligned}
\mathbb{E}[\exp(sd_i)|\mathscr{M}_{i-1}] &= \mathbb{E}\left[\exp\left\{s\|d_i\|_\infty \frac{d_i}{\|d_i\|_\infty}\middle|\mathscr{M}_{i-1}\right\}\right] \\
&\le \exp\left(\frac{s^2\|d_i\|_\infty^2}{2}\right)
\end{aligned}$$

となる．この不等式を繰り返し用いれば，

$$\begin{aligned}\mathbb{E}[\exp\{s(f(X)-Pf)\}] &= \mathbb{E}[\exp\{s\sum_{i=1}^{n} d_i\}]\\ &\le \mathbb{E}[\exp\{s\sum_{i=1}^{n-1} d_i\}\mathbb{E}(sd_n|\mathscr{M}_{n-1})]\\ &\le \exp\left(\frac{s^2\|d_n\|_\infty^2}{2}\right)\mathbb{E}[\exp\{s\sum_{i=1}^{n-1} d_i\}]\\ &\quad\cdots\\ &\le \exp\left(\frac{s^2a^2}{2}\right)\end{aligned}$$

となる．つぎに，Markov の不等式を用いる：

$$\begin{aligned}\mathbb{P}\{f(X)-Pf>t\} &= \mathbb{P}\{\exp(s(f(X)-Pf))>e^{st}\}\\ &\le e^{-st}\mathbb{E}[\exp(s\sum_{i=1}^{n} d_i)]\\ &\le \exp\left[-st+\frac{s^2a^2}{2}\right]\\ &\le \exp\left(-\frac{t^2}{2a^2}\right)\end{aligned}$$

を得る．最後の不等号は $g(s)=-st+s^2a^2/2$ は $s=t/a^2$ のとき，最小値 $-t^2/2a^2$ を取ることよりわかる． □

補題 1.6 V と Z を確率変数とし，$\mathbb{E}(V|Z)=0$, *a.s.* とする．$c>0$ を定数とし，関数 f は

$$f(Z)\le V\le f(Z)+c$$

を満足するものとする．このとき，すべての $t>0$ に対して，

$$\mathbb{E}(e^{tV}|Z)\le \exp\left(\frac{t^2c^2}{8}\right)$$

が成立する．

証明 補題 1.2 よりわかる． □

補題 1.7 $X_1, X_2, \ldots, X_n$ を確率変数列とする．$\mathscr{M}_i=\sigma\{X_1, X_2, \ldots, X_i\}$, $i=1, 2, \ldots, n$ とする．$\{V_i, \mathscr{M}_i\}_{i=1}^n$ はマルチンゲール差列とする．さらに，$Z_1, Z_2, \ldots, Z_n$ と非負の定数 $c_1, c_2, \ldots, c_n$ が存在し，Z_i は $X_1, X_2, \ldots, X_{i-1}$ の関数で $Z_i\le V_i\le Z_i+c_i$ を満足するものとする．このとき，すべての $t>0$ に対して，

$$\mathbb{P}\{\sum_{i=1}^{n} V_i>t\}\le \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right),$$

$$\mathbb{P}\{-\sum_{i=1}^{n} V_i>t\}\le \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right)$$

が成立する．

証明　任意 $s > 0$ に対して，

$$\begin{aligned}\mathbb{P}\{\sum_{i=1}^{n} V_i > t\} &= \mathbb{P}\{\exp(s\sum_{i=1}^{n} V_i) > e^{st}\}\\ &\le e^{-st}\mathbb{E}[\exp\{s\sum_{i=1}^{n} V_i\}]\\ &= e^{-st}\mathbb{E}[\exp\{(s\sum_{i=1}^{n-1} V_i\mathbb{E}(e^{sV_n}|\mathscr{M}_{n-1})\}]\\ &\quad\cdots\\ &\le e^{-st}\exp\left[s^2\sum_{i=1}^{n}\frac{c_i^2}{8}\right]\\ &\le \exp\left[-\frac{2t^2}{\sum_{i=1}^{n} c_i^2}\right]\end{aligned}$$

となる．最後の不等号は $s = 4t/\sum_{i=1}^{n} c_i^2$ とおくことで得られる．□

定理 1.3　　$X_1, X_2, \ldots, X_n$ を互いに独立な確率変数列とし，その値域を $\mathcal{A}$ とする．関数 $g : \mathcal{A}^n \to \mathbb{R}$ は各 $i (i = 1, 2, \ldots, n)$ に対し，ある定数 $c_i > 0$ が存在し，

$$\sup_{x_1, x_2, \ldots, x_n \in \mathcal{A}, x_i' \in \mathcal{A}} |g(x_1, x_2, \ldots, x_n) - g(x_1, x_2, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n| \le c_i \tag{1.5}$$

このとき，すべての $t > 0$ に対して，

$$\mathbb{P}\{g(X_1, X_2, \ldots, X_n) - \mathbb{P}g > t\} \le \exp\left(-\frac{2t^2}{\sum_{i=1}^{n} c_i^2}\right)$$

$$\mathbb{P}\{-(g(X_1, X_2, \ldots, X_n) - \mathbb{P}g) > t\} \le \exp\left(-\frac{2t^2}{\sum_{i=1}^{n} c_i^2}\right)$$

を満足するとする．

証明　$V = g(X_1, X_2, \ldots, X_n) - Pg$ とし，

$$V_i = \mathbb{E}(V|\mathscr{M}_i) - \mathbb{E}(V|\mathscr{M}_{i-1}), \qquad i = 1, 2, \ldots, n$$

とおく．ただし，$\mathscr{M}_i = \sigma(X_1, X_2, \ldots, X_n)$ である．また，$\mathbb{E}(V|\mathscr{M}_0) = \mathbb{E}(V)$ とする．明らかに，$\{(V_i, \mathscr{M}_i)\}_{i=1}^{n}$ はマルチンゲール差列である．いま，

$$V_i = H_i(X_1, X_2, \ldots, X_i) - \int H_i(X_1, X_2, \ldots, X_{i-1}, u)\, dF_i(u)$$

となる．ただし，$F_i(\cdot)$ は X_i の累積分布関数である．さらに，

$$W_i = \sup_u \left\{H_i(X_1, X_2, \ldots, X_{i-1}, u) - \int H_i(X_1, X_2, \ldots, X_{i-1}, u')\, dF_i(u')\right\}$$

と

$$Z_i = \inf_u \left\{H_i(X_1, X_2, \ldots, X_{i-1}, u) - \int H_i(X_1, X_2, \ldots, X_{i-1}, u')\, dF_i(u')\right\}$$

とおけば，

$$Z_i \le V_i \le W_i, \qquad a.s.$$

と

$$W_i - Z_i \le \sup_u \sup_{u'} (H_i(X_1, X_2, \ldots, X_{i-1}, u) - H_i(X_1, X_2, \ldots, X_{i-1}, u')) \le c_i$$

となる．よって，補題 1.2 より定理は示される．□

2

経験過程理論と VC - 族

·····2.1·····

経験測度とエントロピー数

$X_1, X_2, \ldots, X_n$ を可測空間 $(\mathscr{X}, \mathscr{A})$ 上の確率測度 P からのランダム標本とする．可測関数 $f : \mathscr{X} \mapsto \mathbb{R}$ に対して

$$
\begin{aligned}
\mathbb{P}_n f &= \frac{1}{n}\sum_{i=1}^{n} f(X_i), \\
P f &= \int f\, dP, \\
\mathbb{G}_n f, &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(f(X_i) - Pf) = \sqrt{n}(\mathbb{P}_n - P)f, \\
\|f\|_{P,r} &= (P|f|^r)^{1/r}, \qquad r \geq 1
\end{aligned}
$$

とする．可測関数の族 $\mathcal{F}$ が与えられたとき，$\mathbb{P}_n$ を $\mathcal{F}$ から $\mathbb{R}$ への写像 $f \mapsto \mathbb{P}_n f\,(f \in \mathcal{F})$ とみなすことができる．関数族 $\mathcal{F}$ に対する envelope 関数 $F_{\mathcal{F}}(\cdot)$ は次を満足するものとする：すべての $x \in \mathscr{X}$ と $f \in \mathcal{F}$ に対して，

$$|f(x)| \leq F_{\mathcal{F}}(x)$$

が成立することである．$\mathcal{F}$ 上の実数値関数 $z : \mathcal{F} \mapsto \mathbb{R}$ に対して，$\|z\|_{\mathcal{F}}$ を sup ノルムとする：

$$\|z\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |z(f)|.$$

定義 2.1 （被覆数） 任意の正数 $\epsilon > 0$ に対して，$\mathcal{F}$ を覆うために必要な半径 ϵ の $\|\cdot\|_{P,r}$ - 球（中心 f ）$\{g : \|g - f\|_{P,r} \leq \epsilon\}$ の最小の数を被覆数と呼び，$N(\epsilon, \mathcal{F}, \|\cdot\|_{P,r})$ と記す．また，$\log N(\epsilon, \mathcal{F}, \|\cdot\|_{P,r})$ を括弧なしエントロピーと呼ぶ．

定義 2.2 （括弧数） 与えられた二つの関数 $l(x)$ と $u(x)$ （$l(x) < u(x)$，$\|l\|_{P,r} < \infty$，$\|u\|_{P,r} < \infty$）に対して，括弧 $[l, u]$ を $l \leq f \leq u$ なるすべての関数 $f\,(f \in \mathcal{F}$ の集まりとする．ϵ - 括弧は $[l, u]$ で $\|u - l\|_{P,r} < \epsilon$ なるものをいう．$\mathcal{F}$ を覆うために必要な ϵ - 括弧の最小数を括弧数と呼び，$N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|_{P,r})$ と記す．さらに，$\log N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|_{P,r})$ を括弧付きエントロピーと呼ぶ．

定義 2.3 envelope 関数 $F_{\mathcal{F}}$ に関する L_p - 一様被覆数を次で定義する：任意の $\epsilon > 0$ に対して，

$$\sup_Q N(\epsilon\|F_{\mathcal{F}}\|_{Q,p}, \mathcal{F}, \|\cdot\|_{Q,r}).$$

ただし， supremum は $(\mathscr{X}, \mathscr{A})$ 上のすべての離散測度 Q で $\|F_{\mathcal{F}}\|_{Q,p} < \infty$ なるものに関して取る．

注意 2.1 $N(\epsilon, \mathcal{F}, \|\cdot\|_{P,r})$ と $N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|_{P,r})$ を $N(\epsilon, \mathcal{F}, L_r(P))$ と $N_{[]}(\epsilon, \mathcal{F}, L_r(P))$ とそれぞれ記すこともある．

·····2.2·····

Glivenko - Cantelli 族

定義 2.4 $\mathcal{F}$ を可測関数 $f : \mathscr{X} \mapsto \mathbb{R}$ の族とする．関数族 $\mathcal{F}$ が P - Glivenko - Cantelli であるとは，

$$\|\mathbb{P}_n - P\|_{\mathcal{F}} \to 0, \qquad a.s.$$

が成立することである．

注意 2.2 $\|\mathbb{P}_n - P\|_{\mathcal{F}}$ の可測性は保証されないので，$a.s.$ は以下のように解釈する．$\{Z_n\}_{n=1}^{\infty}$ をある距離空間に値を取る可測とはかぎらない写像とし，d をその距離とする．Z_n がある確率要素 Z にほとんど確実に収束するとは，ある可測関数 Δ_n が存在して，すべての n に対して $d(Z_n, Z) \leq \Delta_n$ を満足する Δ_n がほとんど確実に 0 に収束することである．

以下で．ある関数族 $\mathcal{F}$ が P - Glivenko - Cantelli であるための十分条件を述べる．

定理 2.1 すべての $\epsilon > 0$ に対して，

$$N_{[\,]}(\epsilon, \mathcal{F}, L_1(P)) < \infty$$

のとき，$\mathcal{F}$ は P - Glivenko - Cantelli である．

証明 $N = n_{[\,]}(\epsilon, \mathcal{F}, L_1(P))$ とおく．任意の $\epsilon > 0$ に対して，$\{[l_j, u_j]\}_{j=1}^{N}$ を $\mathcal{F}$ の ϵ - 括弧とする．すなわち，

$$N = N_{[\,]}(\epsilon, \mathcal{F}, L_1(P)), \quad \|u_j - l_j\|_{P,1} \leq \epsilon, \qquad j = 1, 2, \ldots, N$$

で各 $f \in \mathcal{F}$ に対して，ある j_0 が存在して，$f \in [l_{j_0}, u_{j_0}]$ とできる．このとき，

$$\begin{aligned}\int f\, d(\mathbb{P}_n - P) &= \int f\, d\mathbb{P}_n - \int f\, dP \leq \int u_{j_0}\, d\mathbb{P}_n - \int f\, dP \\ &= \int u_{j_0}\, d(\mathbb{P}_n - P) + \int (u_{j_0} - f)\, dP \\ &\leq \int u_{j_0}\, d(\mathbb{P}_n - P) + \epsilon\end{aligned}$$

となる．同様にすれば，

$$\int f\, d(\mathbb{P}_n - P) \geq \int l_{j_0}\, d(\mathbb{P}_n - P) - \int (f - l_{j_0})\, dP \geq \int l_{j_0}\, d(\mathbb{P}_n - P) - \epsilon$$

となる．このふたつの式を合わせれば，

$$\max_{1\leq j\leq N} \int u_j\, d(\mathbb{P}_n - P) + \epsilon \geq \sup_{f\in\mathcal{F}} (\mathbb{P}_n - P)f \geq \max_{1\leq j\leq N} \int l_j\, d(\mathbb{P}_n - P) - \epsilon$$

を得る．最後に大数の強法則を上の式の最左辺と最右辺に適用すれば，定理は証明される． □

定理 2.2 $\mathcal{F}$ の envelope 関数 $F_{\mathcal{F}}(x)$ は $PF_{\mathcal{F}} < \infty$ を満足し，$\mathcal{F}$ は P - 可測な族とする．さらに，すべての $\epsilon > 0$ に対して， L_1 - 一様被覆数が有限であるとする．すなわち，

$$\sup_{Q} N(\epsilon, \mathcal{F}, L_1(Q)) < \infty$$

である．ここで，上の式の sup の Q は $(\mathscr{X}, \mathscr{A})$ 上の離散測度で $\|F_{\mathcal{F}}\|_{Q,1} < \infty$ なるものである．このとき，$\mathcal{F}$ は P - Glivenko - Cantelli である．

注意 2.3 $\mathcal{F}$ が P - 可測であるとは，$e_i \in \mathbb{R}, i = 1, 2, \ldots, n,$ に対して

$$(X_1, X_2, \ldots, X_n) \mapsto \left\| \sum_{i=1}^n e_i f(X_i) \right\|_{\mathcal{F}}$$

が $(\mathscr{X}^n, \mathscr{A}^n, P^n)$ の完備測度空間上で可測であることである．

$\mathcal{F}$ が可算部分集合 $\mathcal{G}$ を含み，すべての $f \in \mathcal{F}$ に対して，ある列 $\{g_m\}_{m=1}^{\infty}$ が $\mathcal{G}$ の中に存在して，各 x に対して，$m \nearrow \infty$ のとき，$g_m(x) \to f(x)$ ならば，

$$\left\| \sum_{i=1}^n e_i f(X_i) \right\|_{\mathcal{F}} = \left\| \sum_{i=1}^n e_i g(X_i) \right\|_{\mathcal{F}}$$

となるので，$\mathcal{F}$ は P - 可測な族となる．したがって，$\mathcal{F} = \{1_C : C$ は $\mathbb{R}^d$ のボレル集合 $\}$ は P - 可測な族となる．

定理 2.3 $\mathcal{F}$ は P - 可測な族とし，$\mathcal{F}$ の envelope 関数 $F_{\mathcal{F}}(x)$ は $PF_{\mathcal{F}} < \infty$ を満足し，すべての $\epsilon > 0$ に対して，

$$\frac{1}{n} \log N(\epsilon, \mathcal{F}, L_1(\mathbb{P}_n)) \xrightarrow{P} 0 \tag{2.6}$$

が成立するとする．このとき，$\mathcal{F}$ は P - Glivenko - Cantelli である．

証明 証明は 第 2.5 節 章で行う． □

·····2.3·····

VC - 族に対する被覆数

$\mathscr{X}$ の部分集合族 $\mathscr{C}$ と $x_1, x_2, \ldots, x_n \in \mathscr{X}$ に対し，

$$\Delta_n^{\mathscr{C}}(x_1, x_2, \ldots, x_n) = \#\{C \cap \{x_1, x_2, \ldots, x_n\} : C \in \mathscr{C}\}$$

とする．さらに，

$$\mathcal{S}_{\mathscr{C}}^{(n)} = \max_{x_1, x_2, \ldots, x_n} \Delta_n^{\mathscr{C}}(x_1, x_2, \ldots, x_n)$$

とし，

$$V(\mathscr{C}) \equiv \inf\{n : \mathcal{S}_{\mathscr{C}}(n) < 2^n\}$$

とする．$V(\mathscr{C})$ を VC - 次元という．

定義 2.5 $V(\mathscr{C}) < \infty$ のとき，$\mathscr{C}$ を VC 族と呼ぶ．

補題 2.1 $\mathscr{X}$ の部分集合族 $\mathscr{C}$ の VC - 次元は有限とする．すなわち，$V(\mathscr{C}) < \infty$ とする．このとき，すべての $n \geq V(\mathscr{C})$ に対して，

$$\mathcal{S}_{\mathscr{C}}(n) \leq \sum_{j=0}^{\tilde{V}} \binom{n}{j} \leq \left(\frac{ne}{\tilde{V}}\right)^{\tilde{V}}$$

が成立する．ただし，$\tilde{V} = V(\mathscr{C}) - 1$ である．

証明 VC - 次元の定義より，n 個を最大 $\tilde{V}$ の組にしか分類できないので，分類できるすべての組の数は

$$\sum_{j=0}^{\tilde{V}} \binom{n}{j}$$

となる．

第二番目の不等式の証明：Y を二項分布 $Bi(n, 1/2)$ に従う確率変数とする．$0 < r < 1$ とする．このとき，

$$\begin{aligned}
\sum_{j=0}^{\tilde{V}} \binom{n}{j} &= 2^n \sum_{j=0}^{\tilde{V}} \binom{n}{j} \left(\frac{1}{2}\right)^{\tilde{V}} \\
&= 2^n \mathbb{P}\{Y \leq \tilde{V}\} = 2^n \mathbb{P}\{r^Y \geq r^n\} \\
&= 2^n r^{-\tilde{V}} \mathbb{E} r^Y \qquad \text{（Markov の不等式）} \\
&= 2^n r^{-\tilde{V}} \sum_{j=0}^{n} \binom{n}{j} \left(\frac{1}{2}\right)^n r^j \\
&= 2^n r^{-\tilde{V}} \sum_{j=0}^{\tilde{V}} \binom{n}{j} \left(\frac{1}{2}\right)^{n-j} \left(\frac{r}{2}\right)^j \\
&= 2^n r^{-\tilde{V}} \left(\frac{1}{2} + \frac{r}{2}\right)^n \\
&= r^{-\tilde{V}} (1+r)^n \\
&= \left(\frac{n}{\tilde{V}}\right)^{\tilde{V}} \left(1 + \frac{\tilde{V}}{n}\right)^n \qquad \text{（}r = \tfrac{\tilde{V}}{n}\text{ とおく））} \\
&\leq \left(\frac{n}{\tilde{V}}\right)^{\tilde{V}} e^{\tilde{V}}
\end{aligned}$$

よりわかる． □

$(\mathscr{X}, d)$ を距離空間とする．

定義 2.6 $\epsilon > 0$ とする．$\mathscr{X}$ の点の集まり $\mathscr{D}$ が ϵ - 分離されているとは，任意の $x, y \in \mathscr{D}$ に対して，$d(x, y) > \epsilon$ が成立することである．つめ込み数（*packing number*）$D(\epsilon, \mathscr{X}, d)$ とは，$\mathscr{X}$ の中の ϵ - 分離されている点の集まりの最大数とする．

注意 2.4 $N(\epsilon, \mathscr{X}, d) \le D(\epsilon, \mathscr{X}, d) \le N(\epsilon/2, \mathscr{X}, d)$ となる．なぜならば，ϵ - 分離されている点を中心とした半径 ϵ の球は $\mathscr{X}$ を覆うので，第一番目の不等式はわかる．半径 $\epsilon/2$ の球で $\mathscr{X}$ を覆えるならば，各球の中心に点を配置すれば，あとはどこに点を置いても距離が ϵ 以下になるので，第二番目の不等式はわかる．

定理 2.4 $r \ge 1$ とし，$\mathscr{C}$ を $\mathscr{X}$ の部分集合族とする．$\mathscr{C}$ が VC - 族でその VC - 次元が $V(\mathscr{C})$ ならば，ある定数 K が存在して，任意の確率測度 P と $0 < \epsilon < 1$ に対して，

$$N(\epsilon, \mathscr{C}, L_r(P)) \le \left(\frac{K \log(K/\epsilon^r)}{\epsilon^r}\right)^{V(\mathscr{C})-1} \le \left(\frac{K'}{\epsilon}\right)^{r(V(\mathscr{C})-1)+\delta} \tag{2.7}$$

が成立する．ただし，K' は定数で $\delta > 0$ とする．

さらに，

$$N(\epsilon, \mathscr{C}, L_r(P)) \le K'' V(\mathscr{C})(4e)^{V(\mathscr{C})} \left(\frac{1}{\epsilon}\right)^{r(V(\mathscr{C})-1)} \tag{2.8}$$

が成立する．ただし，K'' はある定数である．

証明 $\mathscr{C}$ は VC - 族なので，$N(\epsilon, \mathscr{C}, L_1(P)) < \infty$ である．したがって，$D(\epsilon, \mathscr{C}, L_1(P)) = m$ と仮定する．すなわち，$C_1, C_2, \dots, C_m \in \mathscr{C}$ が存在して，$i \ne j$ に対して，

$$P\{C_i \triangle C_j\} = \mathbb{E}|1_{C_i} - 1_{C_j}| > \epsilon$$

とできる．このとき，大数の強法則から

$$\mathbb{P}_n\{C_i \triangle C_j\} \xrightarrow{as} P\{C_i \triangle C_j\}$$

となるので，ある n と ω が存在して，$X_1(\omega), X_2(\omega), \dots, X_n(\omega)$ に対して，

$$|\mathbb{P}_n\{C_i \triangle C_j\} - P\{C_i \triangle C_j\}| \le \frac{1}{2}|P\{C_i \triangle C_j\} - \epsilon|$$

とできる．したがって，$\mathbb{P}_n\{C_i \triangle C_j\} > \epsilon$ のときに，$P\{C_i \triangle C_j\} \le \epsilon$ と仮定すると

$$\mathbb{P}_n\{C_i \triangle C_j\} - P\{C_i \triangle C_j\} \ge \frac{1}{2}|P\{C_i \triangle C_j\} - \epsilon|$$

となり，矛盾する．したがって，$\mathbb{P}_n\{C_i \triangle C_j\} > \epsilon$ ならば，$P\{C_i \triangle C_j\} > \epsilon$ である．このことより，

$$D(\epsilon, \mathscr{C}, L_1(P)) \le D(\epsilon, \mathscr{C}, L_1(\mathbb{P}_n))$$

となるので，P を任意の離散測度として定理を証明すれば十分であることがわかる．

C_i と C_j が $\{X_1, X_2, \dots, X_n\}$ から同じ部分集合を取り出すための必要十分条件は $X_k \in C_i \triangle C_j$ なる元 X_k が存在しないことである．対偶を考える．もし，すべての $C_i \triangle C_j$ がある X_k を取り出せば，すべての $\{C_i\}_{i=1}^m$

は $\{X_1, X_2, \ldots, X_n\}$ から異なる部分集合を取り出せることになる．したがって，$\mathscr{C}$ は $\{X_1, X_2, \ldots, X_n\}$ から すくなくとも m 個の部分集合を取り出せることになる．したがって，

$$\begin{aligned}
&P\left[\{\text{ すべての } i \neq j \text{ に対し，ある } k \text{ が存在して，} X_k \in C_i \triangle C_j \text{ となる }\}^c\right] \\
&\quad = P\left[\{\text{ ある } i \neq j \text{ に対し，すべての } k(1 \leq k \leq n) \text{ に対して，} X_k \notin C_i \triangle C_j \ \}\right] \\
&\quad \leq \sum_{i<j} P\left[\{\text{ すべての } k \text{ に対して，} X_k \notin C_i \triangle C_j \ \}\right] \\
&\quad \leq \left(\frac{m(m-1)}{2}\right)\{\max_{i<j}(1 - P\{C_i \triangle C_j\})\}^n \\
&\quad \leq \left(\frac{m(m-1)}{2}\right)(1-\epsilon)^n < 1
\end{aligned}$$

が十分大きな n に対して成立することがわかる．実際，$-\log(1-\epsilon) < \epsilon$ と $m^3 > m(m-1)/2$ から

$$n > \frac{\log m^3}{\log \epsilon} = \frac{\log\{m(m-1)/2\}}{-\log(1-\epsilon)}$$

ならばよいことがわかる．したがって，

$$n = \lfloor 3 \log m / \epsilon \rfloor$$

とすればよい．ただし，$\lfloor x \rfloor$ は x を越えない最大の整数とする．この n に対して，

$$P\left[\{\text{ すべての } i \neq j \text{ に対し，ある } k \text{ が存在して，} X_k \in C_i \triangle C_j \text{ となる }\}\right] > 0$$

となる．したがって，ある $X_1(\omega), X_2(\omega), \ldots, X_n(\omega)$ が存在して，

$$\begin{aligned}
m &= \Delta_n^{\{C_1, C_2, \ldots, C_m\}}(X_1(\omega), X_2(\omega), \ldots, X_n(\omega)) \\
&\leq \Delta_n^{\mathscr{C}}(X_1(\omega), X_2(\omega), \ldots, X_n(\omega)) \\
&\leq \max_{x_1, x_2, \ldots, x_n} \Delta_n^{\mathscr{C}}(x_1, x_2, \ldots, x_n) \\
&\leq \left(\frac{ne}{\tilde{V}}\right)^{\tilde{V}}
\end{aligned}$$

となることがわかる．最後の不等号は補題 2.1 よりわかり，$\tilde{V} = V(\mathscr{C}) - 1$ とした．また，$n < 3\log m/\epsilon$ より

$$m < \left(\frac{3e \log m}{\epsilon \tilde{V}}\right)^{\tilde{V}}$$

となる．これを解けば，

$$\frac{m^{1/\tilde{V}}}{\log m} \leq \frac{3e}{\epsilon \tilde{V}}$$

となる．上の式を評価するために，$g(x) = x/\log x (x > 0)$ を考える．これは $x = e$ で最小値をとるので，

$$g(\log x) = \frac{\log x}{\log \log x} \geq g(e) = e$$

となる．したがって，y が不等式 $y \geq g(x)$ を満足すれば，

$$\log y \geq \log g(x) = \log x - \log\log x = \log x \left(1 - \frac{\log\log x}{\log x}\right) \geq \left(1 - \frac{1}{e}\right) \log x$$

となる．したがって，$y \geq x/\log x$ ならば，

$$x \leq y \log x \leq \left(1 - \frac{1}{e}\right)^{-1} y \log y$$

が成立する．ここで，$x = m^{1/\tilde{V}}$ と $y = 3e/\epsilon$ とおけば，

$$g(x) = \frac{x}{\log x} = \frac{\tilde{V} m^{1/\tilde{V}V}}{\log m} \le \frac{3e}{\epsilon} = y$$

となるので，仮定を満足する．したがって，

$$m^{1/\tilde{V}} \le \frac{e}{e-1} \frac{3e}{\epsilon} \log \frac{3e}{\epsilon}$$

となり，

$$D(\epsilon, \mathscr{C}, L_1(P)) = m \le \left\{ \frac{e}{e-1} \frac{3e}{\epsilon} \log \frac{3e}{\epsilon} \right\}^{\tilde{V}}$$

となる．$K = 3e^2/(e-1)$ とおけば，$3e \ge K$ ともなるので，

$$D(\epsilon, \mathscr{C}, L_1(P)) \le \left(\frac{K}{\epsilon} \log \frac{K}{\epsilon} \right)^{\tilde{V}}$$

を得る．よって，$r = 1$ のとき，(2.7) が証明された．

つぎに，$r > 1$ の場合について (2.7) を示す．

$$\|1_C - 1_D\|_{Q,1} = Q(C \bigtriangleup D) = \|1_C - 1_D\|_{Q,r}^r$$

より

$$N(\epsilon, \mathscr{C}, L_r(Q)) = N(\epsilon^r, \mathscr{C}, L_1(Q)) \le \left(\frac{K}{\epsilon^r} \log \frac{K}{\epsilon^r} \right)^{\tilde{V}} \le \left(\frac{K''}{\epsilon} \right)^{r\tilde{V}+\delta}$$

がわかる．

(2.8) の証明については van der Vaart and Wellner(1996) の p. 137 - 139 を参照. □

定義 2.7 関数 $f : \mathscr{X} \to \mathbb{R}$ のサブグラフを $\{(x, t) \in \mathscr{X} \times \mathbb{R} : |t| < |f(x)|\}$ で与えられる $\mathscr{X} \times \mathbb{R}$ の部分集合で定義する．$\mathscr{X}$ から $\mathbb{R}$ への関数族 $\mathcal{F}$ が VC - サブグラフ族であるとは，サブグラフの集まりが VC - 族であることである．VC - サブグラフ族に対し，その VC - 次元を $V(\mathcal{F}) = V(\mathcal{F}$ のサブグラフ $)$ と記す．

定理 2.5 $r \ge 1$ とする．$\mathcal{F}$ を envelope 関数 $F_{\mathcal{F}}(x)$ をもつ VC - サブグラフとし，その VC - 次元を $V(\mathcal{F})$ とする．P を $\|F_{\mathcal{F}}\|_{P,r} < \infty$ を満足する任意の確率測度とする．このとき，

$$N(2\epsilon \|F_{\mathcal{F}}\|_{P,r}, \mathcal{F}, L_r(P)) \le KV(\mathcal{F})(4e)^{V(\mathcal{F})} \left(\frac{1}{\epsilon} \right)^{r(V(\mathcal{F})-1)}$$

が成立する．ただし，K は定数とし，$0 < \epsilon < 1$ とする．

証明 $C_f = \{(x, t) : |t| \le |f(x)|\}$ とし，$\mathscr{C}$ は $f \in \mathcal{F}$ のすべてのサブグラフ C_f の集まりとする．Fubini の定理より

$$P|f - g| = (P \otimes \lambda)(C_f \bigtriangleup C_g)$$

となる．ただし，λ は $\mathbb{R}$ 上のルベーグ測度である．$P \otimes \lambda$ が $\{(x, t) : |t| \le |f(x)|\}$ 上の確率測度になるように基準化する：

$$R = \frac{P \otimes \lambda}{2PF_{\mathcal{F}}}.$$

(2.8) から

$$N(2\epsilon PF_{\mathcal{F}}, \mathcal{F}, L_1(P)) = N(\epsilon, \mathscr{C}, L_1(R)) \le KV(\mathcal{F}) \left(\frac{4e}{\epsilon} \right)^{V(\mathcal{F})-1}$$

となる．

つぎに，$r > 1$ の場合を考える．確率測度 Q を P に関する密度関数が $F_{\mathcal{F}}^{r-1}/PF_{\mathcal{F}}^{r-1}$ となるものとする．$f, g \in \mathcal{F}$ に対して，

$$P|f-g|^r \le P|f-g|(2F_{\mathcal{F}})^{r-1} \le 2^{r-1}Q|f-g|PF_{\mathcal{F}}^{r-1}$$

となる．よって，f と g をつぎのようにとる．

$$\|f-g\|_{Q,1} \le \epsilon^r QF_{\mathcal{F}}. \tag{2.9}$$

Q の定義より

$$QF_{\mathcal{F}} = \int F_{\mathcal{F}} \frac{F_{\mathcal{F}}^{r-1}}{PF_{\mathcal{F}}^{r-1}}\, dP = \frac{PF_{\mathcal{F}}^r}{PF_{\mathcal{F}}^{r-1}}$$

となることに注意すれば，

$$\begin{aligned}
\|f-g\|_{P,r}^r &\le 2^{r-1}Q|f-g|PF_{\mathcal{F}}^{r-1} \\
&= 2^{r-1}\|f-g\|_{Q,1}PF_{\mathcal{F}}^{r-1} \\
&\le 2^{r-1}\epsilon^r QF_{\mathcal{F}} \cdot PF_{\mathcal{F}}^{r-1} \qquad \text{((2.9) より)} \\
&\le 2^{r-1}\epsilon^r PF_{\mathcal{F}}^r \\
&\le 2^r\epsilon^r \|F_{\mathcal{F}}\|_{P,r}^r
\end{aligned}$$

となるので，

$$\|f-g\|_{Q,1} \le \epsilon^r QF_{\mathcal{F}}$$

ならば，

$$\|f-g\|_{P,r} \le 2\epsilon\|F\|_{P,r}$$

が成立する．したがって，

$$\begin{aligned}
N(2\epsilon\|F_{\mathcal{F}}\|_{P,r}, \mathcal{F}, L_r(P)) &\le N(\epsilon^r QF_{\mathcal{F}}, \mathcal{F}, L_1(Q)) \\
&\le KV(\mathcal{F})\left(\frac{4e}{\epsilon^r QF_{\mathcal{F}}}\right)^{V(\mathcal{F})-1} \\
&= K'V(\mathcal{F})(4e)^{V(\mathcal{F})-1}\left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{F})-1)}
\end{aligned}$$

が $r = 1$ の場合の不等式を利用すればわかる． □

·····2.4·····

Convex Hull と VC - Hull 族

定義 2.8 関数族 $\mathcal{F}$ の convex hull とは，$f_i \in \mathcal{F}(i = 1, 2, \ldots, m)$ と $\alpha_i \geq 0, \sum_{i=1}^{m} \alpha_i = 1$ によって，$\sum_{i=1}^{m} \alpha_i f_i$ の形で表現される関数の集合である．これを $\mathrm{conv}(\mathcal{F})$ と記す．可測関数族 $\mathcal{F}$ が VC - hull 族であるとは，ある VC - 族 $\mathcal{G}$ が存在して，$\mathcal{F} \subset \overline{\mathrm{conv}(\mathcal{G})}$ となることである．ただし，閉包は各点収束で考える．

定理 2.6 P を $(\mathscr{X}, \mathscr{A})$ 上の確率測度とし，$\mathcal{F}$ を 2 乗可積分な envelope 関数 $F_{\mathcal{F}}(x)(\|F_{\mathcal{F}}\|_{P,2} < \infty)$ を持つ関数族で $0 < \epsilon < 1$ に対して，

$$N(\epsilon \|F_{\mathcal{F}}\|_{P,2}, \mathcal{F}, L_2(P)) \leq C \left(\frac{1}{\epsilon}\right)^{V}$$

を満足するとする．ただし，C と V は定数である．このとき，C と V にのみ依存した定数 K が存在して，

$$\log N(\epsilon \|F_{\mathcal{F}}\|_{P,2}, \overline{\mathrm{conv}(\mathcal{F})}, L_2(P)) \leq K \left(\frac{1}{\epsilon}\right)^{2V/(V+2)}$$

が成立する．

証明 van der Vaart and Wellner 1996) の p. 142 - 143 を参照. □

例 2.1 $\mathcal{G} = \{1_{[t,\infty)}(x) : t \in \mathbb{R}\}$ とする．$\mathcal{F}$ の envelope 関数は $F_{\mathcal{G}}(x) = 1$ となり，$\mathcal{G}$ は VC - 次元 $V(\mathcal{G}) = 2$ の VC - 族である．よって，$0 < \epsilon < 1$ に対して，

$$N(\epsilon, \mathcal{G}, L_2(P)) \leq \frac{K}{\epsilon^2}$$

となる．ただし，K はある定数である．$\mathcal{F} = \{g : \mathbb{R} \to [0, 1], g$ は単調 $\}$ とすれば，$\mathcal{F} \subset \overline{\mathrm{conv}(\mathcal{G})}$ となる．したがって，

$$\log N(\epsilon, \mathcal{F}, L_2(P)) \leq \frac{K}{\epsilon}$$

を得る．

例 2.2 $\mathcal{G} = \{1_{[t,\infty)}(x) : t \in \mathbb{R}^d\}$ とする．すると，$\mathcal{G}$ は VC - 次元 $V(\mathcal{G}) = d + 1$ の VC - 族である．よって，$0 < \epsilon < 1$ に対して，

$$N(\epsilon, \mathcal{G}, L_2(P)) \leq K\epsilon^{-2d}$$

となる．ただし，K はある定数である．$\mathcal{F} = \{g : \mathbb{R}^d \to [0, 1], g$ は $\mathbb{R}^d$ 上の累積分布関数$\}$ とすれば，$\mathcal{F} \subset \overline{\mathrm{conv}(\mathcal{G})}$ となる．したがって，

$$\log N(\epsilon, \mathcal{F}, L_2(P)) \leq K\epsilon^{-2d/(d+1)}$$

を得る．特に，$d = 2$ のとき，

$$\log N(\epsilon, \mathcal{F}, L_2(P)) \leq K\epsilon^{-1/3}$$

を得る．

·····2.5·····

Glivenko - Cantelli の定理の証明

2.5.1 重み付き和に対する最大不等式

Θ を添え字集合とし，$\mathcal{F} = \{f_\theta : \theta \in \Theta\}$ を $\mathscr{X}$ 上の関数族とする．

補題 2.2 $W = (W_1, W_2, \ldots, W_n) \in \mathbb{R}^n$ を確率ベクトルとする．ある定数 $C_1 > 0$ と $C_2 > 0$ が存在して，すべての $r = (r_1, r_2, \ldots, r_n)$ とすべての $t > 0$ に対して，

$$\mathbb{P}\left\{\left|\sum_{i=1}^n r_i W_i\right| > t\right\} \le C_1 \exp\left[-\frac{t^2}{C_2^2 \sum_{i=1}^n r_i^2}\right] \tag{2.10}$$

が成り立つとする．さらに，$R = \sup_{f\in\mathcal{F}} \|f\|_{\mathbb{P}_n, 2}$ とする．このとき，C_1 と C_2 にのみ依存する定数 C が存在して，すべての $0 < \epsilon < \delta$ と $K > 0$ に対して，

$$\sqrt{n}(\delta - \epsilon) \ge C\left(\int_{\epsilon/(4K)}^R \sqrt{\log N(\epsilon, \mathcal{F}, L_2(P))}\, d\epsilon\right) \tag{2.11}$$

と

$$\mathbb{P}\left\{\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^n W_i f(X_i)\right| \ge \delta \quad \text{かつ} \quad \frac{1}{n}\sum_{i=1}^n W_i^2 \le K^2\right\} \le C\exp\left[-\frac{n(\delta-\epsilon)^2}{C^2R^2}\right]$$

が成立する．

証明 $S = \min\{s \ge 1 : s \in \mathbb{N},\, 2^{-s}R \le \epsilon/K\}$ とおく．$1 \le s \le S$ に対して，$\{f_j^s\}_{j=1}^{N_s}$ を半径 $2^{-s}R$ の $\mathcal{F}$ の $L_2(\mathbb{P}_n)$ に関する最小被覆とする．したがって，$N_s = N(2^{-s}R, \mathcal{F}, L_2(\mathbb{P}_n))$ となる．いま，f_θ^s は $\{f_1^s, f_2^s, \ldots, f_{N_s}^s\}$ の中で f_θ に最も近いものとする．集合 $\{(1/n)\sum_{i=1}^n W_i^2 \le K^2\}$ 上で Cauchy - Schwarz の不等式を用いる：

$$\left|\frac{1}{n}\sum_{i=1}^n W_i(f_\theta(X_i) - f_\theta^s(X_i))\right| \le K\|f_\theta - f_\theta^s\|_{\mathbb{P}_n,2} \le K2^{-s}R \le \epsilon$$

となる．これより，

$$\begin{aligned}
&\mathbb{P}\left\{\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^n W_i f_\theta(X_i)\right| \ge \delta \quad \text{かつ} \quad \frac{1}{n}\sum_{i=1}^n W_i^2 \le K^2\right\} \\
&\quad \le \mathbb{P}\left\{\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^n W_i(f_\theta(X_i) - f_\theta^s(X_i))\right|\right. \\
&\qquad\quad \left. + \max_{j=1,2,\ldots,N_s}\left|\frac{1}{n}\sum_{i=1}^n W_i f_j^s(X_i)\right| \ge \delta \quad \text{かつ} \quad \frac{1}{n}\sum_{i=1}^n W_i^2 \le K^2\right\} \\
&\quad \le \mathbb{P}\left\{\max_{j=1,2,\ldots,N_s}\left|\frac{1}{n}\sum_{i=1}^n W_i f_j^s(X_i)\right| \ge \delta - \epsilon\right\}
\end{aligned}$$

となる．いま，$f_\theta^s = \sum_{s=1}^S (f_\theta^s - f_\theta^{s-1})$ とおく．三角不等式から

$$\|f_\theta^s - f_\theta^{s-1}\|_{\mathbb{P}_n,2} \le \|f_\theta^s - f_\theta\|_{\mathbb{P}_n,2} + \|f_\theta - f_\theta^{s-1}\|_{\mathbb{P}_n,2} \le 2^{-s}R + 2^{-s+1}R = 3(2^{-s}R) \tag{2.12}$$

となる．$\eta_s > 0$ で $\sum_{s=1}^{S} \eta_s \le 1$ となるように $\{\eta_s\}_{s=1}^{S}$ をとる．このとき，

$$\begin{aligned}
&\mathbb{P}\left\{\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n} W_i(f_\theta^s(X_i) - f_\theta^{s-1}(X_i))\right| \ge \delta - \epsilon\right\} \\
&\qquad \le \sum_{s=1}^{S}\mathbb{P}\left\{\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n} W_i(f_\theta^s(X_i) - f_\theta^{s-1}(X_i))\right| \ge \eta_s(\delta - \epsilon)\right\}
\end{aligned}$$

となる．$f_\theta^s - f_\theta^{s-1}$ の組み合わせは最大 $\{N(2^{-s}R, \mathcal{F}, L_2(\mathbb{P}_n))\}^2$ である．また，(2.12) より

$$\sum_{i=1}^{n}(f_\theta^s(X_i) - f_\theta^{s-1}(X_i))^2 \le n\|f_\theta^s - f_\theta^{s-1}\|_{\mathbb{P}_n, 2} \le n9(2^{-2s}R^2).$$

これと (2.10) から，$X_1, X_2, \dots, X_n$ が与えられたとき，

$$\begin{aligned}
&\mathbb{P}\left\{\sup_{\theta\in\Theta}\left|\frac{1}{n}\sum_{s=1}^{S}\sum_{i=1}^{n} W_i(f_\theta^s(X_i) - f_\theta^{s-1}(X_i))\right| \ge \eta_s(\delta - \epsilon)\right\} \\
&\qquad \le \{N(2^{-s}R, \mathcal{F}, L_2(\mathbb{P}_n))\}^2 C_1 \exp\left[-\frac{n^2\eta_s^2(\delta-\epsilon)^2}{C_2^2\sum_{i=1}^{n}(f_\theta^s(X_i) - f_\theta^{s-1}(X_i))^2}\right] \\
&\qquad \le C_1 \exp\left[2\log N(2^{-s}R, \mathcal{F}, L_2(\mathbb{P}_n)) - \frac{n(\delta-\epsilon)^2}{9C_2^2 2^{-2s}R^2}\right]
\end{aligned}$$

となる．さらに，

$$\sqrt{n}(\delta - \epsilon) \ge 12C_2\sum_{s=1}^{S} 2^{-s}R\sqrt{\log N(2^{-s}R, \mathcal{F}, L_2(\mathbb{P}_n))} \vee C_2R\sqrt{1152\log 2} \tag{2.13}$$

が成立するようにする．いま，

$$\eta_s = \frac{6C_2 2^{-s}R\sqrt{\log N(2^{-s}R, \mathcal{F}, L_2(\mathbb{P}_n))}}{\sqrt{n}(\delta-\epsilon)} \vee \frac{2^{-s}\sqrt{s}}{8}$$

とする．すると

$$\sum_{s=1}^{S}\eta_s \le \sum_{s=1}^{S}\frac{6C_2 2^{-s}R\sqrt{\log N(2^{-s}R, \mathcal{F}, L_2(\mathbb{P}_n))}}{\sqrt{n}(\delta-\epsilon)} + \sum_{s=1}^{S}\frac{2^{-s}\sqrt{s}}{8} \le \frac{1}{2} + \frac{1}{2} \le 1$$

となる．最後の不等号は (2.13)

$$\begin{aligned}
\sum_{s=1}^{S} 2^{-s}\sqrt{s} &\le 1 + \int_1^\infty \frac{\sqrt{x}}{2^x}\,dx \\
&\le 1 + \int_1^\infty \frac{x}{2^x}\,dx \\
&= 1 - \frac{2^{-x}x}{\log 2}\bigg|_1^\infty + \int_1^\infty \frac{2^{-x}}{\log 2}\,dx \\
&= 1 - \frac{2^{-x}}{(\log 2)^2}\bigg|_1^\infty = 1 + \frac{1}{2(\log 2)^2} \le 4
\end{aligned}$$

□

よりわかる．さらに，

$$\eta_s \ge \frac{6C_2 2^{-s}R\sqrt{\log N(2^{-s}R, \mathcal{F}, L_2(\mathbb{P}_n))}}{\sqrt{n}(\delta-\epsilon)}$$

より

$$\log N(2^{-s}R, \mathcal{F}, L_2(\mathbb{P}_n)) \le \frac{n(\delta-\epsilon)^2\eta_s^2}{36C_2^2 2^{-2s}R^2}$$

である．これより

$$\sum_{s=1}^{S} C_1 \exp\left[2\log N(2^{-s}R,\,\mathcal{F},\,L_2(\mathbb{P}_n)) - \frac{n(\delta-\epsilon)^2\eta_s^w}{9C_2^2 2^{-2s}R^2}\right]$$
$$\leq \sum_{s=1}^{S} C_1 \exp\left[-\frac{n(\delta-\epsilon)^2\eta_s^2}{18C_2^2 2^{-2s}R^2}\right]$$

となる．さらに，$\eta_s \geq 2^{-s}\sqrt{s}/8$ に注意すれば，

$$\begin{aligned}\sum_{s=1}^{S} C_1 \exp\left[-\frac{n(\delta-\epsilon)^2\eta_s^2}{18C_2^2 2^{-2s}R^2}\right] &\leq \sum_{s=1}^{S} C_1 \exp\left[-\frac{n(\delta-\epsilon)^2 s^2}{1152C_2^2 2^{-2s}R^2}\right]\\ &\leq \sum_{s=1}^{\infty} C_1 \exp\left[-\frac{n(\delta-\epsilon)^2 s^2}{1152C_2^2 R^2}\right]\\ &= C_1\left\{1-\exp\left[-\frac{n(\delta-\epsilon)^2}{1152C_2^2R^2}\right]\right\}^{-1}\\ &\leq 2C_1\exp\left[-\frac{n(\delta-\epsilon)^2}{1152C_2^2R^2}\right]\end{aligned}$$

がわかる．最後の不等号は以下のようにすればわかる．まず，

$$\exp\left[-\frac{n(\delta-\epsilon)^2\eta_s^w}{1152C_2^2 2^{-2s}R^2}\right] < 1$$

に注意する．したがって，$0 < t < 2$ に対して

$$\frac{t}{1-t} \leq 2t$$

が成立することことよりわかる．

2.5.2 対称化

$X_1', X_2', \ldots, X_n'$ は $X_1, X_2, \ldots, X_n$ の独立複製とする．すなわち，$X_1', X_2', \ldots, X_n'$ も独立同一に確率測度 P に従い，$X_1, X_2, \ldots, X_n$ とも独立である．さらに，$\mathbb{P}_n'$ を $X_1', X_2', \ldots, X_n'$ から構成された経験測度とする．

補題 2.3 $\epsilon > 0$ とする．すべての $f \in \mathcal{F}$ に対して，

$$\mathbb{P}\left\{\left|\int f\,d(\mathbb{P}_n - P)\right| > \frac{\epsilon}{2}\right\} \leq \frac{1}{2}$$

が成立すると仮定する．このとき，

$$\mathbb{P}\left\{\sup_{f\in\mathcal{F}}\left|\int f\,d(\mathbb{P}_n - P)\right| > \epsilon\right\} \leq 2\mathbb{P}\left\{\sup_{f\in\mathcal{F}}\left|\int f\,d(\mathbb{P}_n - \mathbb{P}_n')\right| > \frac{\epsilon}{2}\right\}$$

が成立する．

証明 $X = (X_1, X_2, \ldots, X_n)$，$A_f = \{X : |\int f\,d(\mathbb{P}_n - P)| > \epsilon\}$，$\mathcal{A} = \cup_{f\in\mathcal{F}} A_f$ とする．$X \in \mathcal{A}$ ならば，ある f^* が存在して，$X \in A_{f^*}$ を満足する．さらに，$X' = (X_1', X_2', \ldots, X_n')$ としたとき，

$$\begin{aligned}\mathbb{P}\left\{A_{f^*}\quad \text{かつ}\quad \left|\int f\,d(\mathbb{P}_n' - P)\right| \leq \frac{\epsilon}{2}\right\} &= \mathbb{P}\left\{A_{f^*}\right\}\mathbb{P}\left\{\left|\int f\,d(\mathbb{P}_n' - P)\right| \leq \frac{\epsilon}{2}\right\}\\ &\leq \frac{1}{2}\mathbb{P}\left\{A_{f^*}\right\}\\ &= \frac{1}{2}\mathbb{P}\left\{\left|\int f\,d(\mathbb{P}_n - P)\right| > \epsilon\right\}\end{aligned}$$

となる．したがって，

$$\begin{aligned}
\mathbb{P}\left\{\sup_{f\in\mathcal{F}}\left|\int f\,d(\mathbb{P}_n-P)\right|>\epsilon\right\} &\le \mathbb{P}\left\{\left|\int f^*\,d(\mathbb{P}_n-P)\right|>\epsilon\right\}\\
&\le \mathbb{P}\left\{\left|\int f^*\,d(\mathbb{P}_n-P)\right|>\epsilon \quad \text{かつ} \quad \left|\int f^*\,d(\mathbb{P}'_n-P)\right|\le\frac{\epsilon}{2}\right\}\\
&\le 2\mathbb{P}\left\{\left|\int f^*\,d(\mathbb{P}_n-\mathbb{P}_n)\right|>\frac{\epsilon}{2}\right\}\\
&\le 2\mathbb{P}\left\{\sup_{f\in\mathcal{F}}\left|\int f\,d(\mathbb{P}_n-\mathbb{P}_n)\right|>\frac{\epsilon}{2}\right\}
\end{aligned}$$

より補題は証明された． □

次に，経験過程を対称化するためにつぎのような確率変数列を導入する：$W_1, W_2, \ldots, W_n$ は互いに独立で

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

とする．この確率変数列を Rademacher 列と呼ぶ．$W_1, W_2, \ldots, W_n$ を $X_1, X_2, \ldots, X_n$ および $X'_1, X'_2, \ldots, X'_n$ と独立にとる．すると，各 f と $i(i=1,2,\ldots,n)$ に対して，$f(X_i)-f(X'_i)$ と $W_i(f(X_i)-f(X'_i))$ は同じ分布に従う．同様に，

$$\{f(X_i)-f(X'_i):\, i=1,2,\ldots,n,\, f\in\mathcal{F}\}$$

と

$$\{W_i(f(X_i)-f(X'_i)):\, i=1,2,\ldots,n,\, f\in\mathcal{F}\}$$

は同じ分布に従う．また，

$$\begin{aligned}
&\mathbb{P}\left\{\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^n W_i(f(X_i)-f(X'_i))\right|>\frac{\epsilon}{2}\right\}\\
&\qquad\le \mathbb{P}\left\{\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^n W_if(X_i)\right|+\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^n W_if(X'_i)\right|>\frac{\epsilon}{2}\right\}\\
&\qquad\le \mathbb{P}\left\{\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^n W_if(X_i)\right|>\frac{\epsilon}{4} \quad \text{または} \quad \sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^n W_if(X'_i)\right|>\frac{\epsilon}{4}\right\}\\
&\qquad\le 2\mathbb{P}\left\{\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^n W_if(X_i)\right|>\frac{\epsilon}{4}\right\} \qquad (2.14)
\end{aligned}$$

となる．

補題 2.4 ある正数 R に対して，$\sup_{f\in\mathcal{F}}\|f\|_\infty\le R$ が成立するとする．このとき，$\epsilon>0$ と $n\ge 8R^2/\epsilon^2$ に対して，

$$\mathbb{P}\left\{\sup_{f\in\mathcal{F}}\left|\int f\,d(\mathbb{P}_n-P)\right|>\epsilon\right\}\le 4\mathbb{P}\left\{\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^n W_if(X_i)\right|>\frac{\epsilon}{4}\right\}$$

が成立する．

証明　Chebyshev の不等式を用いる：

$$\begin{aligned}\mathbb{P}\left\{\left|\int f\,d(\mathbb{P}_n - P)\right| > \frac{\epsilon}{2}\right\} &\le \mathbb{P}\left\{\sum_{i=1}^{n}\left|f(X_i) - Pf\right| > \frac{n\epsilon}{2}\right\} \\ &\le \frac{4}{n\epsilon^2}\mathbb{E}\left[\sum_{i=1}^{n}(f(X_i) - Pf)^2\right] \\ &= \frac{4}{n\epsilon}\mathbb{E}[(f(X_1) - Pf)^2] \\ &\le \frac{4R^2}{n\epsilon^2} \le \frac{1}{2}\end{aligned}$$

となる．これより，補題 2.3 と (2.14) を用いれば，

$$\begin{aligned}\mathbb{P}\left\{\sup_{f\in\mathcal{F}}\left|\int f\,d(\mathbb{P}_n - P)\right| > \epsilon\right\} &\le 2\mathbb{P}\left\{\sup_{f\in\mathcal{F}}\left|\int f\,d(\mathbb{P}_n - \mathbb{P}'_n)\right| > \frac{\epsilon}{2}\right\} \\ &\le 4\mathbb{P}\left\{\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}W_i f(X_i)\right| > \epsilon\right\}\end{aligned}$$

を得る． □

2.5.3 Glivenko - Cantelli の定理の証明

まず，f が有界のときに定理が成立することを示す．

補題 2.5　ある正数 $R > 0$ に対して，$\sup_{f\in\mathcal{F}}|f| \le R$ が成立するとする．このとき，任意の $\epsilon > 0$ に対して，

$$\frac{1}{n}\log N(\epsilon, \mathcal{F}, L_2(\mathbb{P}_n)) \xrightarrow{P} 0$$

ならば，$\mathcal{F}$ は Glivenko - Cantelli である．

証明　マルチンゲールの収束定理の議論から

$$\sup_{f\in\mathcal{F}}\left|\int f\,d(\mathbb{P}_n - P)\right|$$

はある定数に概収束することが保証される．(Pollard (1984) の p.39 の問題 [11] と p.215 の定理 24 の証明を参照.)

任意の $\epsilon > 0$ をとる．補題 2.4 を用いる： $n \ge 8R^2/\epsilon^2$ ならば，

$$\mathbb{P}\left\{\sup_{f\in\mathcal{F}}\left|\int f\,d(\mathbb{P}_n - P)\right| > \epsilon\right\} \le 4\mathbb{P}\left\{\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}W_i f(X_i)\right| > \frac{\epsilon}{4}\right\}$$

となる．$\sup_{f\in\mathcal{F}}|f| \le R$ に注意して，Hoeffding の不等式を用いる：

$$\mathbb{P}\left\{\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^{n}W_i f(X_i)\right| > \frac{n\epsilon}{4}\right\} \le 2\exp\left[-\frac{2\cdot n^2\epsilon^2/16}{\sum_{i=1}^{n}\{2f(X_i)\}^2}\right] \le 2\exp\left[-\frac{n^2\epsilon^2}{32R^2}\right]$$

より

$$\mathbb{P}\left\{\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}W_i f(X_i)\right| > \frac{\epsilon}{4}\right\} \le 2\exp\left[-\frac{n^2\epsilon^2}{32R^2}\right]$$

となる．補題 2.2 における ϵ，δ，k をそれぞれ $\epsilon/4$，$\epsilon/8$，1 におきかえれば，(2.10) の右辺

$$\int_{\epsilon/32}^{R}\sqrt{\log N(x, \mathcal{F}, L_2(\mathbb{P}_n))}\,dx \le R\sqrt{\log N(\epsilon/32, \mathcal{F}, L_2(\mathbb{P}_n))}$$

と評価できる．これは $N(x, \mathcal{F}, L_2(\mathbb{P}_n))$ の x に関しての単調減少性よりわかる．ここで

$$A_n = \left\{ (X_1, X_2, \ldots, X_n) : \frac{\epsilon\sqrt{n}}{8} \geq CR\left(\sqrt{\log N(\epsilon/32, \mathcal{F}, L_2(\mathbb{P}_n))}\right) \vee 1 \right\}$$

とおけば，A_n 上では (2.11) を満足する．また，補題の仮定から $\mathbb{P}(A_n^c) \to 0, (n \to \infty)$ となる．これらに注意して，補題 2.2 を用いれば，

$$\begin{aligned}
\mathbb{P}\left\{ \sup_{f\in\mathcal{F}} \left| \frac{1}{n}\sum_{i=1}^{n} W_i f(X_i) \right| > \frac{\epsilon}{4} \right\} &\leq \mathbb{P}\left\{ \sup_{f\in\mathcal{F}} \left| \frac{1}{n}\sum_{i=1}^{n} W_i f(X_i) \right| > \frac{\epsilon}{4} \quad \text{かつ} \quad A_n \right\} + \mathbb{P}\{A_n\} \\
&\leq C\exp\left[-\frac{n^2\epsilon^2}{32R^2} \right] + \mathbb{P}\{A_n\} \\
&\to 0, \qquad (n \nearrow \infty \text{ のとき})
\end{aligned}$$

を得る． □

最後に，条件 $\sup|f| \leq R$ を除くための議論をする．そのために，R に対して，

$$\mathcal{F}_R = \{ f\mathbf{1}\{F_{\mathcal{F}} \leq R :\}, f \in \mathcal{F} \}$$

とおく．任意の $f_1, f_2 \in \mathcal{F}$ に対して，

$$\int (f_1 - f_2)^2 \mathbf{1}\{F_{\mathcal{F}}(\,\cdot\,) \leq R\}\, d\mathbb{P}_n \leq 2R \int |f_1 - f_2|\, d\mathbb{P}_n$$

より，

$$N(\epsilon, \mathcal{F}_R, L_2(\mathbb{P}_n)) \leq N(\epsilon, \mathcal{F}, L_1(\mathbb{P}_n))$$

となる．したがって，(2.6) は

$$\frac{1}{n}\log N(\epsilon, \mathcal{F}_R, L_2(\mathbb{P}_n)) \xrightarrow{P} 0$$

を保証する．したがって，補題 2.4 から，$\mathcal{F}_R$ は Glivenko - Cantelli 族である．任意の $\epsilon > 0$ に対して，R を十分に大きくとれば，$F_{\mathcal{F}} \in L_1(P)$ より

$$\int_{F_{\mathcal{F}} > R} F_{\mathcal{F}}\, dP \leq \epsilon$$

とできる．次に，n を十分大きくとれば，大数の強法則から

$$\int_{F_{\mathcal{F}} > R} F_{\mathcal{F}}\, d\mathbb{P}_n \leq 2\epsilon, \qquad a.s.$$

とできる．また，$\mathcal{F}_R$ は Glivenko - Cantelli 族であることから

$$\sup_{f\in\mathcal{F}} \left| \int_{F_{\mathcal{F}} \leq R} f\, d(\mathbb{P}_n - P) \right| \leq \epsilon, \qquad a.s.$$

とできる．これらより

$$\begin{aligned}
\sup_{f\in\mathcal{F}} \left| \int f\, d(\mathbb{P}_n - P) \right| &\leq \sup_{f\in\mathcal{F}} \left| \int_{F_{\mathcal{F}} \leq R} f\, d(\mathbb{P}_n - P) \right| + \sup_{f\in\mathcal{F}} \left| \int_{F_{\mathcal{F}} > R} F_{\mathcal{F}}\, d\mathbb{P}_n \right| + \sup_{f\in\mathcal{F}} \left| \int_{F_{\mathcal{F}} > R} f\, dP \right| \\
&\leq 4\epsilon, \qquad a.s.
\end{aligned}$$

となり，定理 2.3 は証明された．

3

統計的２値識別問題

·····3.1·····

問題設定とベイズリスク

パターン認識（分類または判別）は観測値がどの群に属するかを推測もしくは予測することが目的である．観測を d - 次元ベクトル x とし，この観測の属性をクラスと呼ぶ．クラスは 2 つで構成さえているとする．クラスをそれぞれ 0 と 1 に対応させる．観測のの属性を y と記すことにする．パターン識別問題では与えられた観測 x に基づいて，y を予測する関数 $g(x) : \mathbb{R}^d \to \{0, 1\}$ を求める．関数 g を分類子と呼ぶ．すると，x が与えられたとに， $g(x) \neq y$ が分類子 g の誤りとなる．

学習問題をモデル化するために，$\mathbb{R}^d \times \{0, 1\}$ - 値確率変数 (X, Y) を導入する．任意のボレル集合 $A \subset \mathbb{R}^d$ に対して，

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

とし，任意の $x \in \mathbb{R}^d$ に対して，

$$\eta(x) = \mathbb{P}\{Y = 1|X = x\} = \mathbb{E}[Y|X = x]$$

とする．(X, Y) の同時分布は

$$\mathbb{P}\{X \in A, Y = y\} = \int_A \{\eta(x)\}^y \{1 - \eta(x)\}^{1-y} \mu(dx)$$

で与えられる．なぜなら，

$$\mathbb{E}[\mathbb{1}_A(X)\mathbb{1}\{Y = y\}] = \mathbb{E}[\mathbb{1}_A(X)\mathbb{E}(\mathbb{1}\{Y = y\}|X)] = \mathbb{E}[\mathbb{1}_A(X)\{\eta(X)\}^y\{1 - \eta(X)\}^{1-y}]$$

よりわかる．$\eta(x)$ を事後確率と呼ぶ．

任意の分類子 $g : \mathbb{R}^d \to \{0, 1\}$ に対して，$g(x) \neq y$ のときに誤りは起こり，分類子 g の誤りの確率は

$$\mathscr{R}(g) = \mathbb{P}\{g(X) \neq Y\}$$

となる．ベイズ分類子を

$$g^*(x) = \begin{cases} 1, & (\eta(x) > \frac{1}{2}) \\ 0, & (\text{その他}) \end{cases}$$

で定義すれば，g^* が誤りの確率 $\mathscr{R}$ を最小にすることが次の定理からわかる．すなわち，$g(x) = \ \mathrm{argmin}\ _y \mathbb{P}\{Y = y|X = x\}$ である．

定理 3.1 どんな分類子 $g : \mathbb{R}^d \to \{0, 1\}$ に対しても

$$\mathbb{P}\{g^*(X) \neq Y\} \leq \mathbb{P}\{g(X) \neq Y\}$$

が成立する．

証明 $X = x$ が与えられたときの誤りの条件付確率を考える：

$$\begin{aligned}
\mathbb{P}\{g(X) \neq Y|X = x\} &= 1 - \mathbb{P}\{g(X) = Y|X = x\} \\
&= 1 - [\mathbb{P}\{Y = 1, g(X) = 1|X = x\} + \mathbb{P}\{Y = 0, g(X) = 0|X = x\}] \\
&= 1 - [\mathbb{1}\{g(x) = 1\}\mathbb{P}\{Y = 1|X = x\} + \mathbb{1}\{g(x) = 0\}\mathbb{P}\{Y = 0|X = x\}] \\
&= 1 - [\mathbb{1}\{g(x) = 1\}\eta(x) + \mathbb{1}\{g(x) = 0\}(1 - \eta(x))]
\end{aligned}$$

となる．したがって，

$$\begin{aligned}
&\mathbb{P}\{g(X) \neq Y|X = x\} - \mathbb{P}\{g^*(X) \neq Y|X = x\} \\
&\quad = \eta(x)(\mathbf{1}\{g^*(x) = 1\} - \mathbf{1}\{g(x) = 1\}) + (1 - \eta(x))(\mathbf{1}\{g^*(x) = 0\} - \mathbf{1}\{g(x) = 0\}) \\
&\quad = (2\eta(x) - 1)(\mathbf{1}\{g^*(x) = 1\} - \mathbf{1}\{g(x) = 1\}) \geq 0 \qquad (3.15)
\end{aligned}$$

となることがわかる．上式の最後の等号は $\mathbf{1}\{g(x) = 0\} + \mathbf{1}\{g(x) = 1\}$ よりわかる．また，最後の不等号は $\eta(x) > 1/2$ ならば，$\mathbf{1}\{g^*(x) = 1\} = 1$ から，$\mathbf{1}\{g^*(x) = 1\} - \mathbf{1}\{g(x) = 1\} \geq 0$ となり，反対に $\eta(x) \leq 1/2$ ならば，$\mathbf{1}\{g^*(x) = 1\} = 0$ から，$\mathbf{1}\{g^*(x) = 1\} - \mathbf{1}\{g(x) = 1\} \leq 0$ となることよりわかる．最後に (3.15) の最右辺と最左辺を X に関して期待値をとれば，定理は示される． □

·····3.2·····

経験リスク最小化問題

$\mathcal{G}$ を分類子 $g:\mathbb{R}^d\to\{0,1\}$ の族とする．ある分類子 g の誤りの確率 $\mathscr{R}(g)=\mathbb{P}\{g(X)\neq Y\}$ を推定するために

$$\hat{\mathscr{R}}_n(g)=\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\{g(X_i)\neq Y_i\}$$

を用いるのが自然である．$\hat{\mathscr{R}}$ のことを g の経験誤り確率または経験リスクと呼ぶ.

分類子の族の中で誤りの確率をその中でもっとも小さくするを選べばよいだろう．もし，分類子 g の誤りの確率を経験リスクの推定を $\mathcal{G}$ に関して一様にうまくできれば，経験リスクを最小にする分類子のリスクが最小に近いことが期待できる.

g_n^* を経験リスクを最小にするものとする：すなわち，すべての $g\in\mathcal{G}$ に対して，

$$\hat{\mathscr{R}}_n(g_n^*)\leq\hat{\mathscr{R}}_n(g)$$

が成立する.

いま，$D_n=(X_1,Y_1,X_2,Y_2,\ldots,X_n,Y_n)$ とする.

補題 3.1　D_n を与えたときの g_n^* の誤りの条件付き確率

$$\mathscr{R}(g_n^*)=\mathbb{P}\{g_n^*(X)\neq Y|D_n\}$$

に対して，

$$\mathscr{R}(g_n^*)-\inf_{g\in\mathcal{G}}\mathscr{R}(g)\leq 2\sup_{g\in\mathcal{G}}|\hat{\mathscr{R}}_n(g)-\mathscr{R}(g)|,$$
$$|\hat{\mathscr{R}}(g_n^*)-\mathscr{R}(g_n^*)|\leq\sup_{g\in\mathcal{G}}|\hat{\mathscr{R}}_n(g)-\mathscr{R}(g)|$$

が成立する.

証明　第一の不等式は

$$\begin{aligned}\mathscr{R}(g_n^*)-\inf_{g\in\mathcal{G}}\mathscr{R}(g)&=\mathscr{R}(g_n^*)-\hat{\mathscr{R}}_n(g_n^*)+\hat{\mathscr{R}}_n(g_n^*)-\inf_{g\in\mathcal{G}}\mathscr{R}(g)\\&\leq\mathscr{R}(g_n^*)-\hat{\mathscr{R}}_n(g_n^*)+\sup_{g\in\mathcal{G}}|\hat{\mathscr{R}}_n(g)-\inf_{g\in\mathcal{G}}\mathscr{R}(g)|\\&\leq 2\sup_{g\in\mathcal{G}}|\hat{\mathscr{R}}_n(g)-\mathscr{R}(g)|\end{aligned}$$

よりわかる.

第二の不等式は自明である. □

$\sup_{g\in\mathcal{G}}|\hat{\mathscr{R}}_n(g)-\mathscr{R}(g)|$ の上限を求めることにより以下の 2 点に対する評価を与えることになる.

- $\mathcal{G}$ の中での最適なものに g_n^* のリスクがどれだけ近いか？すなわち，$\mathscr{R}(g_n^*)-\inf_{g\in\mathcal{G}}\mathscr{R}(g)$ の評価.
- g_n^* の経験リスクは g_n^* のリスクにどの程度近いか？すなわち，$|\mathscr{R}_n(g_n^*)-\mathscr{R}(g_n^*)|$ の評価．$\mathscr{R}_n(\hat{g}_n)$ は $\mathscr{R}(g_n^*)$ を小さめに推定するが，それがどの程度の偏りがわかる.

·····3.3·····

集中確率不等式

定理 3.2　$N = \#\mathcal{G}$ とする．このとき，任意の正数 ϵ に対して，

$$\mathbb{P}\left\{\sup_{g\in\mathcal{G}} |\hat{\mathscr{R}}_n(g) - \mathscr{R}(g)| > \epsilon\right\} \le 2Ne^{-2n\epsilon^2}$$

が成立する．

証明　$f(x, y) = 1\{g(x) \neq y\}$ とおけば，

$$-Pf \le f(x, y) - Pf \le 1 - Pf$$

になることに注意して Hoeffding の不等式を用いる：

$$\begin{aligned}\mathbb{P}\{|\hat{\mathscr{R}}_n(g) - \mathscr{R}(g)| > \epsilon\} &= \mathbb{P}\{|\sum_{i=1}^n f(X_i, Y_i) - Pf| > n\epsilon\} \\ &\le 2\exp\left[-\frac{2n^2\epsilon^2}{n}\right] = 2\exp[-2n\epsilon^2]\end{aligned}$$

を得る．これより，

$$\begin{aligned}\mathbb{P}\{\sup_{g\in\mathcal{G}} |\hat{\mathscr{R}}_n(g) - \mathscr{R}(g)| > \epsilon\} &\le \sum_{g\in\mathcal{G}} \mathbb{P}\{|\hat{\mathscr{R}}_n(g) - \mathscr{R}(g)| > \epsilon\} \\ &\le 2N\exp[-2n\epsilon^2]\end{aligned}$$

より定理は示せた． □

定理 3.3　$N = \#\mathcal{G}$ とする．このとき，

$$\mathbb{E}[\sup_{g\in\mathcal{G}} |\hat{\mathscr{R}}_n(g) - \mathscr{R}(g)|] \le \sqrt{\frac{\log(2N)}{2n}}$$

が成立する．

証明　$f(x, y) = 1\{g(x) \neq y\}$ とおく．任意の正数 ϵ に対して

$$\mathbb{E}[\exp\{t(\hat{\mathscr{R}}_n(g) - \mathscr{R}(g))\}] = \Pi_{i=1}^n \mathbb{E}[\exp\{(t/n)(f(X_i, Y_i) - Pf)\}]$$

となる．$\mathbb{E}[f(X_i, Y_i) - Pf] = 0$ と $-Pf \le f(x, y) \le 1 - Pf$ に注意して，補題 1.2 を用いれば，

$$\mathbb{E}\left[\exp\left\{\frac{t}{n}(f(X_i, Y_i) - Pf)\right\}\right] \le \exp\left[\frac{t^2}{8n^2}\right]$$

となる．したがって，

$$\mathbb{E}\left[\exp\{t(\hat{\mathscr{R}}(g) - \mathscr{R}(g))\}\right] \le \exp\left[\frac{t^2}{8n}\right]$$

となる．同様にすれば，

$$\mathbb{E}\left[\exp\{-t(\hat{\mathscr{R}}(g) - \mathscr{R}(g))\}\right] \le \exp\left[\frac{t^2}{8n}\right]$$

を得る．これより，補題 1.1（$\sigma = 1/\sqrt{4n}$，i を g として）を用いれば，定理は示される． □

·····3.4·····

Vapnik - Chervonenkis の不等式

分類子の集合 $\mathcal{G}$ が有限集合（$N = \#\mathcal{G}$ ）のとき，すべての $\epsilon > 0$ に対し，

$$\mathbb{P}\{\sup_{g\in\mathcal{G}} |\hat{\mathscr{R}}_n(g) - \mathscr{R}(g)| > \epsilon\} \leq 2N \exp(-2n\epsilon^2)$$

と

$$\mathbb{E}[\sup_{g\in\mathcal{G}} |\hat{\mathscr{R}}_n(g) - \mathscr{R}(g)| > \epsilon] \leq \sqrt{\frac{\log(2N)}{2n}}$$

を示した．しかし，N が大きいときには，これらの不等式は有効ではない．

ここでは，N が大きい場合の上のふたつの不等式の右辺の上限を求める．$X_1, X_2, \ldots, X_n$ は $\mathbb{R}^d$ - 値確率変数で独立同一に確率測度 μ に従うとする．さらに，$\mu_n = (1/n)\sum_{i=1}^n \delta_{X_i}$ とおく．

定理 3.4　任意の正数 t に対して，

$$\mathbb{P}\left\{\sup_{A\subset\mathbb{R}^d} |(\mu_n - \mu)(A)| - \mathbb{E} \sup_{A\subset\mathbb{R}^d} |(\mu_n - \mu)(A)| > t\right\} \leq 2e^{-2nt^2}$$

が成立する．

証明　$Z_i = 1\{X_i \in A\} - \mu(A)$ とおく．すると

$$V \equiv \sup_{A\in\mathbb{R}^d} |n(\mu_n - \mu)(A)| = \sup_{A\in\mathbb{R}^d} |\sum_{i=1}^n Z_i|$$

と表現できる．$\mathscr{M}_i = \sigma\{X_1, X_2, \ldots, X_i\}, i = 1, 2, \ldots, n$ とし，

$$V_i = \mathbb{E}[V_i|\mathscr{M}_i] - \mathbb{E}[V_i|\mathscr{M}_{i-1}]$$

とおく．ただし，$\mathbb{E}[V|\mathscr{M}_0] = \mathbb{E}[V]$ とする．さらに，

$$V = \sup_{A\in\mathbb{R}^d}\left|\sum_{i=1}^n Z_i\right| \equiv h(X_1, X_2, \ldots, X_n)$$

とおく．このとき，

$$\begin{aligned}
&|h(X_1, X_2, \ldots, X_n) - h(X_1, X_2, \ldots, X_{i-1}, X_i', X_{i+1}, \ldots, X_n)| \\
&= \sup_{A\in\mathbb{R}^d} |\sum_{j\neq i}^n Z_j + Z_i| - \sup_{A\in\mathbb{R}^d} |\sum_{j\neq i}^n Z_j + Z_i'| \\
&\leq \sup_{A\in\mathbb{R}^d} |Z_i - Z_i'| \\
&= \sup_{A\in\mathbb{R}^d} |1\{X_i \in A\} - 1\{X_i' \in A\}| \leq 1
\end{aligned}$$

となる．ただし，$Z_i' = 1\{X_i' \in A\} - \mu(A)$ とした．したがって，McDiarmid の定理（定理 1.3）の条件 (1.5) を満足するので，この定理から

$$P\{V - \mathbb{E}V > t\} \leq \exp\left(-\frac{2t^2}{n}\right)$$

を得る．故に，

$$P\left\{\sup_{A\in\mathbb{R}^d}(\mu_n-\mu)(A)-\mathbb{E}[\sup_{A\in\mathbb{R}^d}(\mu_n-\mu)(A)]>t\right\}=P\left\{V-\mathbb{E}V>nt\right\}\le\exp[-2nt^2]$$

を得る．同様な議論により

$$P\left\{\sup_{A\in\mathbb{R}^d}-(\mu_n-\mu)(A)+\mathbb{E}[\sup_{A\in\mathbb{R}^d}(\mu_n-\mu)(A)]>t\right\}\le\exp[-2nt^2]$$

を得る．このふたつの式を合わせれば，定理は証明される． □

定理 3.5 $\mathcal{A}$ を $\mathbb{R}^d$ の部分集合族とし，

$$\mathcal{S}_{\mathcal{A}}(n)=\max_{x_1,\,x_2,\,\ldots\,,\,x_n\in\mathbb{R}^d}\#\{\{x_1,\,x_2,\,\ldots\,,\,x_n\}\cap A:A\in\mathcal{A}\}$$

とする．このとき，

$$\mathbb{E}\left[\sup_{A\in\mathcal{A}}|(\mu_n-\mu)(A)|\right]\le 2\sqrt{\frac{\log\{2\mathcal{S}_{\mathcal{A}}(n)\}}{n}}$$

が成立する．

証明 $X_1',\,X_2',\,\ldots\,,\,X_n'$ を $X_1,\,X_2,\,\ldots\,,\,X_n$ の独立複製とし，$W_1,\,W_2,\,\ldots\,,\,W_n$ を $X_1,\,X_1,'\,\ldots\,,\,X_n,\,X_n'$ と独立な Rademacher 列とする：すなわち，$P(W_1=1)=P(W_1=-1)=1/2$ である．$\mu_n'=(1/n)\sum_{i=1}^n\delta_{X_i'}$ とおく．このとき，

$$\begin{aligned}\mathbb{E}\left[\sup_{A\in\mathcal{A}}|(\mu_n-\mu)(A)|\right]&=\mathbb{E}\left[\sup_{A\in\mathcal{A}}\Big|\mathbb{E}[(\mu_n-\mu_n')(A)|X_1,\,X_2,\,\ldots\,,\,X_n]\Big|\right]\\&\le\mathbb{E}\left[\sup_{A\in\mathcal{A}}\mathbb{E}[|(\mu_n-\mu_n')(A)||X_1,\,X_2,\,\ldots\,,\,X_n]\right]\qquad\text{（Jensen の不等式）}\\&\le\mathbb{E}\left[\sup_{A\in\mathcal{A}}|(\mu_n-\mu_n')(A)|\right]\qquad\text{（}\sup\mathbb{E}(\,\cdot\,)\le\mathbb{E}\sup(\,\cdot\,)\text{ から）}\\&=\frac{1}{n}\mathbb{E}\left[\sup_{A\in\mathcal{A}}\left|\sum_{i=1}^n W_i(\mathbb{1}\{X_i\in A\}-\mathbb{1}\{X_i'\in A\})\right|\right]\end{aligned}$$

となる．上の式の最後の等号は $W_i(\mathbb{1}\{X_i\in A\}-\mathbb{1}\{X_i'\in A\})$ と $\mathbb{1}\{X_i\in A\}-\mathbb{1}\{X_i'\in A\}$ が同じ分布に従うこと[1)]よりわかる．つぎに，$X_i=x_i,\,X_i'=x_i',\,i=1,\,2,\,\ldots\,,\,n$ を固定して

$$\mathbb{E}\left[\sup_{A\in\mathcal{A}}\left|\sum_{i=1}^n W_i(\mathbb{1}\{x_i\in A\}-\mathbb{1}\{x_i'\in A\})\right|\right]$$

を評価する．$\tilde{\mathcal{A}}\subset\mathcal{A}$ を $\tilde{\mathcal{A}}$ の中の任意のふたつの集合は $\{x_1,\,x_1',\,\ldots\,,x_n,\,x_n'\}$ と異なる積をもつように構成する：すなわち，任意の $A_1,\,A_2\in\tilde{\mathcal{A}}$ に対して

$$A_1\cap\{x_1,\,x_1',\,\ldots\,,x_n,\,x_n'\}\ne A_2\cap\{x_1,\,x_1',\,\ldots\,,x_n,\,x_n'\}$$

が成立する．したがって，$\#\tilde{\mathcal{A}}\le\mathcal{S}_{\mathcal{A}}(2n)$ となる．これより

$$\mathbb{E}\left[\sup_{A\in\mathcal{A}}\left|\sum_{i=1}^n W_i(\mathbb{1}\{x_i\in A\}-\mathbb{1}\{x_i'\in A\})\right|\right]=\mathbb{E}\left[\sup_{A\in\tilde{\mathcal{A}}}\left|\sum_{i=1}^n W_i(\mathbb{1}\{x_i\in A\}-\mathbb{1}\{x_i'\in A\})\right|\right]$$

となる．$P(W_i=\pm1)=1/2$ と

$$\mathbb{E}[W_i(\mathbb{1}\{x_i\in A\}-\mathbb{1}\{x_i'\in A\})]=0$$

---

1)同時分布の同一性には A が属する集合族 $\mathcal{A}$ に条件が必要である．

および

$$-1 \le W_i(\mathbb{1}\{x_i \in A\} - \mathbb{1}\{X_i' \in A\}) \le 1$$

に注意して，補題 1.2（$a = -1$，$b = 1$ として）をもちいれば，任意の正数 t に対して

$$\mathbb{E}\left[\exp\left\{t\sum_{i=1}^{n} W_i(\mathbb{1}\{x_i \in A\} - \mathbb{1}\{x_i' \in A\})\right\}\right]$$
$$= \Pi_{i=1}^{n}\mathbb{E}\exp\left[tW_i(\mathbb{1}\{x_i \in A\} - \mathbb{1}\{x_i' \in A\})\right] \le \exp\left[\frac{nt^2}{2}\right]$$

となる．さらに，$W_i(\mathbb{1}\{x_i \in A\} - \mathbb{1}\{x_i' \in A\})$ はゼロを中心に対称に分布することに注意すれば，

$$\mathbb{E}\left[\exp\left\{-t\sum_{i=1}^{n} W_i(\mathbb{1}\{x_i \in A\} - \mathbb{1}\{x_i' \in A\})\right\}\right] \le \exp\left[\frac{nt^2}{2}\right]$$

が成立するので，補題 1.1（$\sigma = \sqrt{n}$ として）を用いれば

$$\mathbb{E}\left\{\sup_{A\in\tilde{\mathcal{A}}}\left|\sum_{i=1}^{n} W_i(\mathbb{1}\{x_i \in A\} - \mathbb{1}\{x_i' \in A\})\right|\right\} \le \sqrt{2n\log(2\mathcal{S}_{\mathcal{A}}(2n))}$$

を得る．さらに，

$$\{x_1, x_1', \ldots, x_n, x_n'\} \cap A = (\{x_1, x_2, \ldots, x_n, \} \cap A) \cup (\{x_1', x_2', \ldots x_n'\} \cap A)$$

より

$$\max \#[\{x_1, x_1', \ldots, x_n, x_n'\} \cap A :\ A \in \mathcal{A}]$$
$$\le \max \#[\{x_1, x_2, \ldots, x_n\} \cap A :\ A \in \mathcal{A}] \max \#[\{x_1', x_2', \ldots, x_n'\} \cap A :\ A \in \mathcal{A}]$$

より

$$\mathcal{S}_{\mathcal{A}}(2n) \le \{\mathcal{S}_{\mathcal{A}}(n)\}^2$$

がわかる．これより，

$$\sqrt{2n\log\{2\mathcal{S}_{\mathcal{A}}(2n)\}} \le 2\sqrt{n\log\{2\mathcal{S}_{\mathcal{A}}(n)\}}$$

となるので，

$$\mathbb{E}\left[\sup_{A\in\mathcal{A}}|(\mu_n - \mu_n')(A)|\right] \le \frac{1}{n}2\sqrt{n\log\{2\mathcal{S}_{\mathcal{A}}(n)\}} = \sqrt{\frac{\log\{2\mathcal{S}_{\mathcal{A}}(n)\}}{n}}$$

を得る．最後に，$X_1, X_1', \ldots, X_n, X_n'$ に関して期待値を取れば，定理は証明される． □

つぎの定理は，$\sup_A \sqrt{\mu(A)(1-\mu(A))}$ が非常に小さいときには，定理 3.6 を改良するものを与える．

定理 3.6 $\psi = \{\sup_A \sqrt{\mu(A)(1-\mu(A))}\}^2$ とおく．このとき，

$$\mathbb{E}\left[\left|\sup_A(\mu_n - \mu)(A)\right|\right] \le \frac{8\log\{2\mathcal{S}_{\mathcal{A}}(2n)\}}{n} + \sqrt{\frac{32\psi\log\{2\mathcal{S}_{\mathcal{A}}(2n))}{n}}$$

が成立する．

証明 定理 3.6 の証明中の議論から

$$\mathbb{E}\left[\sup_{A\in\mathcal{A}}|(\mu_n - \mu)(A)|\right]$$
$$\le \frac{1}{n}\mathbb{E}\left[\sup_{A\in\mathcal{A}}\left|\sum_{i=1}^{n} W_i(\mathbb{1}\{X_i \in A\} - \mathbb{1}\{X_i' \in A\})\right| X_1, X_1', X_2, X_2', \ldots, X_n, X_n'\right] \quad (3.16)$$

となる．ただし，$X'_1, X'_2, \ldots, X_n$ は $X_1, X_2, \ldots, X_n$ の独立複製で，$W_1, W_2, \ldots, W_n$ は $X_1, X'_1, X_2, X'_2, \ldots, X_n, X'_n$ とは独立な Rademacher 列である．$s > 0$ とする．各 $A \in \mathcal{A}$ に対して

$$-|1\{X_i \in A\} - 1\{X'_i \in A\}| \le W_i(1\{X_i \in A\} - 1\{X'_i \in A\}) \le |1\{X_i \in A\} - 1\{X'_i \in A\}|$$

となることに注意して，補題 1.2 を用いる：

$$\begin{aligned}
&\mathbb{E}\left[\exp\left\{s\sum_{i=1}^n W_i(1\{X_i \in A\} - 1\{X'_i \in A\})\right\}\middle| X_1, X'_1, X_2, X'_2, \ldots, X_n, X'_n\right] \\
&\quad = \Pi_{i=1}^n \mathbb{E}\left[\exp\{sW_i(1\{X_i \in A\} - 1\{X'_i \in A\})\}\middle| X_1, X'_1, X_2, X'_2, \ldots, X_n, X'_n\right] \\
&\quad \le \Pi_{i=1}^n \exp\left[\frac{s^2(1\{X_i \in A\} - 1\{X'_i \in A\})^2}{2}\right] \\
&\quad = \exp\left[\sum_{i=1}^n \frac{s^2}{2}(1\{X_i \in A\} - 1\{X'_i \in A\})^2\right]
\end{aligned}$$

となる．したがって，すべての $A \in \mathcal{A}$ に対して

$$\begin{aligned}
&\mathbb{E}\left[\exp\left\{s\sum_{i=1}^n W_i(1\{X_i \in A\} - 1\{X'_i \in A\})\right\}\middle| X_1, X'_1, X_2, X'_2, \ldots, X_n, X'_n\right] \\
&\quad \le \exp\left[\sup_{A \in \mathcal{A}} \frac{s^2}{2}\sum_{i=1}^n (1\{X_i \in A\} - 1\{X'_i \in A\})^2\right]
\end{aligned}$$

となる．ここで，

$$\sigma^2 = \sigma^2(X_1, X'_1, X_2, X'_2, \ldots, X_n, X'_n) = \sup_{A \in \mathcal{A}} \sum_{i=1}^n (1\{X_i \in A\} - 1\{X'_i \in A\})^2$$

とおく．$\tilde{A} \subset \mathcal{A}$ を $\tilde{A}$ の中の任意のふたつの集合は $\{X_1, X'_1, \ldots, X_n, X'_n\}$ と異なる積をもつように構成する．したがって，$\#\tilde{A} \le \mathcal{S}_{\mathcal{A}}(2n)$ となる．これに注意すれば，

$$\begin{aligned}
&\mathbb{E}\left[\sup_{A \in \mathcal{A}} \left|\sum_{i=1}^n W_i(1\{X_i \in A\} - 1\{X'_i \in A\})\right| \middle| X_1, X'_1, X_2, X'_2, \ldots, X_n, X'_n\right] \\
&\quad = \mathbb{E}\left[\sup_{A \in \tilde{A}} \left|\sum_{i=1}^n W_i(1\{X_i \in A\} - 1\{X'_i \in A\})\right| \middle| X_1, X'_1, X_2, X'_2, \ldots, X_n, X'_n\right] \\
&\quad \le \sigma\sqrt{2\log(2\mathcal{S}_{\mathcal{A}}(2n)}
\end{aligned}$$

となる．上式の最後の不等式は補題 1.1 よりわかる．したがって，

$$\begin{aligned}
&\mathbb{E}\left[\sup_{A \in \mathcal{A}} |(\mu_n - \mu)(A)|\right] \\
&\qquad \le \frac{1}{n}\sqrt{2\log(2\mathcal{S}_{\mathcal{A}}(2n)}\mathbb{E}[\sigma(X_1, X'_1, X_2, X'_2, \ldots, X_n, X'_n)] \qquad (3.17)
\end{aligned}$$

を得る．最後に $\sigma(X_1, X_1', X_2, X_2', \ldots, X_n, X_n')$ を評価する．

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{A\in\mathcal{A}}\sqrt{\sum_{i=1}^{n}(1\{X_i\in A\}-1\{X_i'\in A\})^2}\right]\\
&\le\sqrt{\mathbb{E}\left[\sup_{A\in\mathcal{A}}\sum_{i=1}^{n}(1\{X_i\in A\}-1\{X_i'\in A\})^2\right]}\qquad \text{Jensen の不等式}\\
&\le\sqrt{\mathbb{E}\left[\sup_{A\in\mathcal{A}}\sum_{i=1}^{n}\{(1\{X_i\in A\}-\mu(A))-(1\{X_i'\in A\}-\mu(A))\}^2\right]}\\
&\le\sqrt{\mathbb{E}\left[\sup_{A\in\mathcal{A}}2\sum_{i=1}^{n}\{(1\{X_i\in A\}-\mu(A))^2+(1\{X_i'\in A\}-\mu(A))^2\}\right]}\\
&\le\sqrt{4\mathbb{E}\left[\sup_{A\in\mathcal{A}}\sum_{i=1}^{n}(1\{X_i\in A\}-\mu(A))^2\right]}\\
&\le 2\Bigg\{\mathbb{E}\Bigg[\sup_{A\in\mathcal{A}}\sum_{i=1}^{n}\{(1\{X_i\in A\}-\mu(A))(1-\mu(A))+\mu(A)(\mu(A)-1\{X_i\in A\})\\
&\qquad +\mu(A)(1-\mu(A))\}\Bigg]\Bigg\}^{1/2}\\
&\le 2\sqrt{n\psi}+2\sqrt{n\mathbb{E}\sup_{A\in\mathcal{A}}|(\mu_n-\mu)(A)|}
\end{aligned}
\tag{3.18}
$$

となる．$M=\mathbb{E}\sup_{A\in\mathcal{A}}|(\mu_n-\mu)(A)|$ とおき，(3.18) を (3.17) に代入すれば，

$$M\le\sqrt{\frac{8\log(2\mathcal{S}_{\mathcal{A}}(2n))}{n}}(\sqrt{\psi}+\sqrt{M})$$

を得る．

$$a=\sqrt{\frac{8\log(2\mathcal{S}_{\mathcal{A}}(2n))}{n}}$$

とおいて，$\sqrt{M}$ について解けばてん

$$\sqrt{M}\le\frac{a+\sqrt{a^2+4a\sqrt{\psi}}}{2}$$

となり

$$M=\frac{1}{4}\{a+\sqrt{a^2+4a\sqrt{\psi}}\}^2\le\frac{1}{2}\{a^2+a\ 2+4a\sqrt{\psi}\}=a^2+\sqrt{4a^2\psi}$$

となり，定理は証明された． □

つぎの定理も VC 不等式を改良するものである．そのために以下の記号を準備する．$\mathbb{R}^d$ の部分集合 $\mathcal{A}$ と $X=(X_1, X_2, \ldots, X_n)\in\mathbb{R}^d\times\mathbb{R}^n$ に対して，

$$\mathcal{A}(X)=\{b=(b_1, b_2, \ldots, b_n)\in\{0, 1\}^n:\ b_i=1\{X_i\in A\},\ i=1, 2, \ldots, n,\ A\in\mathcal{A}\}$$

と定める．$\mathcal{A}(X)$ 上の距離を

$$\rho(b, c)=\sqrt{\frac{1}{n}\sum_{i=1}^{n}1\{b_i\ne c_i\}},\qquad b, c\in\mathcal{A}(X)$$

で定める．$\{0, 1\}^n$ の部分集合 B を覆うために必要な半径 $\epsilon(\epsilon>0)$ の ρ - 球の最小の個数を $N(\epsilon, B, \rho)$ と記す．

定理 3.7

$$\mathbb{E}\left[\sup_{A}|(\mu_n-\mu)(A)|\right] \le \frac{24}{\sqrt{n}}\max_{X\in\mathbb{R}^d\times\mathbb{R}^n}\int_0^1\sqrt{\log\{2N(\epsilon,\,\mathcal{A}(X),\,\rho)\}}\,d\epsilon$$

が成立する．

証明　$X_1', X_2', \dots, X_n'$ を $X_1, X_2, \dots, X_n$ の独立複製とし，$W_1, W_2, \dots, W_n$ を Rademacher 列とし，$X_1, X_1', \dots, X_n, X_n'$ とは独立なものとする．このとき，

$$\begin{aligned}
\mathbb{E}\left[\sup_{A}|(\mu_n-\mu)(A)|\right] &\le \frac{1}{n}\mathbb{E}\left[\sup_{A\in\mathcal{A}}\left|\sum_{i=1}^n W_i(\mathbb{1}\{X_i\in A\}-\mathbb{1}\{X_i'\in A\})\right|\right]\\
&\le \frac{2}{n}\mathbb{E}\left[\sup_{A\in\mathcal{A}}\left|\sum_{i=1}^n W_i\mathbb{1}\{X_i\in A\}\right|\right]\\
&= \frac{2}{n}\mathbb{E}\left[\mathbb{E}\big[\sup_{A\in\mathcal{A}}\big|\sum_{i=1}^n W_i\mathbb{1}\{X_i\in A\}\big|\,\Big|X_1, X_2, \dots, X_n\big]\right]
\end{aligned}$$

となる．一番目の不等式は定理 3.4 の証明中の議論と同様にすればわかる．$X_1=x_1, X_2=x_2, \dots, X_n=x_n$ を固定する：

$$\mathbb{E}\left[\sup_{A\in\mathcal{A}}\left|\sum_{i=1}^n W_i\mathbb{1}\{X_i\in A\}\right|\right] = \mathbb{E}\left[\sup_{b\in\mathcal{A}(x)}\left|\sum_{i=1}^n W_i b_i\right|\right]$$

となる．ただし，$x=(x_1, x_2, \dots, x_n)$ とした．

いま，$B_0=\{(\underbrace{0, 0, \dots, 0}_{n})\}$ とし，$B_1, B_2, \dots, B_M\in\{0, 1\}^n$ を半径 2^{-k} の $\mathcal{A}(x)$ の ρ - 被覆とする．ただし，$M=\lfloor\log_2\sqrt{n}+1\rfloor$ とする．B_0 は半径 2^0 の被覆であり，$B_M=\mathcal{A}(x)$ となる．$b^*=(b_1^*, b_2^*, \dots, b_n^*)\in\mathcal{A}(x)$ とし，

$$\big|\sum_{i=1}^n W_i b_i^*\big| = \max_{b\in\mathcal{A}(x)}\big|\sum_{i=1}^n W_i b_i\big|$$

を満足するものとする．各 $k\le M$ に対して，$b^{(k)}\in B_k$ を k 番目の被覆中で b^* の最も近いものとする：すべての $b\in B_i$ に対して，

$$\rho(b^{(k)}, b^*)\le\rho(b, b^*).$$

B_k の定義から $\rho(b^{(k)}, \rho^*)\le 2^{-k}$ となることに注意すれば，

$$\rho(b^{(k)}, b^{(k-1)})\le\rho(b^{(k)}, b^*)+\rho(b^*, b^{(k-1)})\le 3\cdot 2^{-k}$$

となる．さらに，$b^{(0)}=(0, 0, \dots, 0)$ なので，

$$\sum_{i=1}^n W_i b_i^* = \sum_{i=1}^n W_i b_i^{(0)} + \sum_{k=1}^M\sum_{i=1}^n W_i(b_i^{(k)}-b_i^{(k-1)}) = C$$

を得る．したがって，

$$\begin{aligned}
\mathbb{E}\left[\max_{b\in\mathcal{A}(x)}\left|\sum_{i=1}^n W_i b_i\right|\right] &= \mathbb{E}\left[\max_{b\in\mathcal{A}(x)}\left|\sum_{k=1}^M\sum_{i=1}^n W_i(b_i^{(k)}-b_i^{(k-1)})\right|\right]\\
&\le \sum_{k=1}^M\mathbb{E}\left[\big|\sum_{i=1}^n W_i(b_i^{(k)}-b_i^{(k-1)})\big|\right]\\
&\le \sum_{k=1}^M\mathbb{E}\left[\max_{b\in B_k,\,c\in B_{k-1}:\,\rho(b,\,c)\le 3\cdot 2^{-1}}\big|\sum_{i=1}^n W_i(b_i-c_i)\big|\right]
\end{aligned}$$

となる．いま，$\mathbb{E}W_i(b_i - c_i) = 0$ と

$$-|b_i - c_i| \le W_i(b_i - c_i) \le |b_i - c_i|$$

に注意して，補題 1.2 を用いる：

$$\mathbb{E}\exp\left[s\sum_{i=1}^n W_i(b_i - c_i)\right] \le \exp\left[\frac{s^2\sum_{i=1}^n (b_i - c_i)^2}{2}\right] \le \exp\left[\frac{s^2 n(3\cdot 2^{-k})}{2}\right]$$

を得る．一方，$b \in B_k$ と $c \in B_{k-1}$ の組の上限は

$$\#(B_i)\#(B_{k-1}) \le \{\#(B_k)\}^2 \le \{N(2^{-k},\, \mathcal{A}(x),\, \rho)\}^2$$

となる．同様の議論から

$$\mathbb{E}\exp\left[-s\sum_{i=1}^n W_i(b_i - c_i)\right] \le \exp\left[\frac{s^2 n(3\cdot 2^{-k})}{2}\right]$$

を得る．したがって，補題 1.1 を適用（$\sigma = \sqrt{n}3\cdot 2^{-k}$ として）すれば，

$$\mathbb{E}\left[\max_{b\in B_k,\, c\in B_{k-1}:\, \rho(b,\,c)\le 3\cdot 2^{-1}} \Big|\sum_{i=1}^n W_i(b_i - c_i)\Big|\right] \le 3\sqrt{n}2^{-k}\sqrt{2\log[2\{N(2^{-k},\, \mathcal{A}(x),\, \rho)\}^2]}$$

を得る．k に関して合算すれば，

$$\begin{aligned}\mathbb{E}\left[\max_{b\in\mathcal{A}(x)}\Big|\sum_{i=1}^n W_i b_i\Big|\right] &\le 3\sqrt{n}2^{-k}\sum_{k=1}^M \sqrt{2\log[2\{N(2^{-k},\, \mathcal{A}(x),\, \rho)\}^2]} \\ &\le 12\sqrt{n}\sum_{k=1}^M 2^{-(k+1)}\sqrt{\log\{2N(2^{-k},\, \mathcal{A}(x),\, \rho)\}} \\ &\le 12\sqrt{n}\int_0^1 \sqrt{\log\{2N(t,\, \mathcal{A}(x),\, \rho)\}}\, dt\end{aligned}$$

を得る．最後の不等式は $N(t,\, \mathcal{A}(x),\, \rho)$ の t に関する単調性からわかる． □

定理 3.8 $\mathcal{A}$ を有限の VC - 次元 V をもつ集合族とする．すべての $x_1, x_2, \ldots, x_n \in \mathbb{R}^n$ と $0 < \rho < 1$ に対して，

$$N(\epsilon,\, \mathcal{A}(x),\, \rho) \le \left(\frac{4e}{\epsilon^2}\right)^{V/(1-e^{-1})}$$

が成立する．ただし $x = (x_1, x_2, \ldots, x_n)$ である．

証明 $x_1, x_2, \ldots, x_n$ を固定し，$B_0 \in \mathcal{A}(x_1, x_2, \ldots, x_n) \in \{0, 1\}^n$ とする．$\epsilon \in (0, 1)$ を固定し，$B_\epsilon \subset \{0, 1\}^n$ を半径 ϵ の B_0 の最小被覆とする．したがって，$\#(B_\epsilon) \le (4e/\epsilon^2)^{V/(1-V)}$ を示せばよい．

$\#(B_\epsilon) \le \#(C_\epsilon)$ なる packing set $C_\epsilon \subset B_0$ が存在することに注意する．すなわち，$b, c \in C_\epsilon$ は ϵ - 分離されたいる：$\rho(b\,c) > \epsilon$ である．これを示すために，C_ϵ は ϵ - 分離集合で最大の個数のものとする．このとき，任意の $b \in B_0$ に対して，ある $c \in C_\epsilon$ が存在して，$\rho(b, c) \le \epsilon$ である．そうでなけば，b を C_ϵ に加えても C_ϵ の個数は増えても ϵ - 分離である．したがって C_ϵ は半径 ϵ の被覆となる．よって，B_ϵ は ϵ - 被覆で最も個数の少ないものだから，$\#(B_\epsilon) \le \#(C_\epsilon)$ となる．C_ϵ の元を $c^{(1)}, c^{(2)}, \ldots, c^{(M)}$ と記す．任意の $1 \le i, j \le M (i \ne j)$ に対し，A_{ij} を $c^{(i)}$ と $c^{(j)}$ のベクトルの成分で値が異なる成分の集まりとする：

$$A_{ij} = \{1 \le m \le n :\, c_m^{(i)} \ne c_m^{(j)}\}.$$

ただし，$c^{(i)} = (c_1^{(i)}, c_2^{(i)}, \ldots, c_n^{(i)})$ とした．C_ϵ の任意のベクトル $c^{(i)}, c^{(j)}$ は $\|c^{(i)} - c^{(j)}\| \ge \epsilon$ より $n\epsilon^2$ 個の成分は異なる値を持つ．いま，$Y_1, Y_2, \ldots, Y_K$ を K 個の独立な確率変数で $\{1, 2, \ldots, n\}$ 上の一様分布に従うとする．このとき，任意の異なる $1 \le i, j \le M$ と $k \le K$ に対して，

$$P\{Y_k \in A_{ij}\} \ge \epsilon^2$$

となる．したがって，$Y_1, Y_2, \ldots, Y_K$ のどれもが A_{ij} に含まれない確率は $(1-\epsilon)^K$ 以下になる．よって，

$$
\begin{aligned}
&P\{\text{ すべての異なる } i, j \text{ に対して，少なくともひとつの } Y_k \text{ が } A_{ij} \text{ に含まれる }\} \\
&\qquad \geq 1 - M^2(1-\epsilon)^2 \geq 1 - M^2 e^{-K\epsilon^2}
\end{aligned}
$$

となる．ここで，$K = \lfloor (2/\epsilon^2)\log M \rfloor + 1$ とすれば，上の確率は正になる．すなわち，$Y_1(\omega), Y_2(\omega), \ldots, Y_K(\omega)$ は $\mathcal{A}(x)$ によって分離されることになる．よって，$K \leq V$ となる．□

定理 3.9 $\mathcal{A}$ を有限の VC - 次元 V をもつ集合族とする．このとき，

$$
\mathbb{E}\left[\sup_{A\in\mathcal{A}} |(\mu_n - \mu)(A)|\right] \leq c\sqrt{\frac{V}{n}}
$$

が成立する．ただし，c は適当な定数である．

証明 定理 3.7 と定理 3.8 を合わせれば，

$$
\begin{aligned}
\mathbb{E}\left[\sup_{A\in\mathcal{A}} |(\mu_n - \mu)(A)|\right] &\leq \frac{24}{\sqrt{n}} \max_{x\in\mathbb{R}^d\times\mathbb{R}^n} \int_0^n \sqrt{\log 2\left(\frac{4e}{t^2}\right)^{V/(1-e^{-1})}}\, dt \\
&\leq \frac{24}{\sqrt{n}} \int_0^1 \sqrt{\frac{V}{1-e^{-1}}\log\left(\frac{4e}{t^2}\right) + \log 2}\, dt \\
&\leq \frac{24}{\sqrt{n}} \int_0^1 \sqrt{\frac{V}{1-e^{-1}}\log\left(\frac{4e}{t^2}\right)}\, dt + \frac{24}{\sqrt{n}}\sqrt{\log 2} \\
&\leq \frac{24}{1-e^{-1}}\sqrt{\frac{V}{n}} \int_0^1 \sqrt{\log\left(\frac{4e}{t^2}\right)}\, dt + \frac{24}{\sqrt{n}}\sqrt{\log 2} \\
&\leq 24\frac{\sqrt{2e\pi}}{1-e^{-1}}\sqrt{\frac{V}{n}} + \frac{24}{\sqrt{n}}\sqrt{\log 2} \\
&\leq c\sqrt{\frac{V}{n}}
\end{aligned}
$$

となる．最後から 2 番目の不等式は以下の議論からわかる．$t = 2e^{(1-u)/2}$ とおけば $u = \log(4e/t^2)$ と $dt = -e^{(1-u)/2}$ となることに注意すれば，

$$
\begin{aligned}
\int_0^1 \sqrt{\log\frac{4e}{t^2}}\, dt &= \int_{1+2\log 2}^{\infty} \sqrt{u} e^{(1-u)/2}\, du \\
&\leq \sqrt{e}\int_0^{\infty} \sqrt{u} e^{-u/2}\, du \\
&= 2\sqrt{2e}\int_0^{\infty} \sqrt{u} e^{-u}\, du \\
&= 2\sqrt{2e}\frac{\pi}{4} = \sqrt{2\pi e}
\end{aligned}
$$

となる．最後の等式はガンマ関数の性質

$$
\int_0^{\infty} \sqrt{u} e^{-u}\, du = \sqrt{\frac{\pi}{4}}
$$

を用いた．□

定理 3.9 において，$\mathcal{A} = \{(-\infty, x] : x \in \mathbb{R}\}$ として，Dvoretzky-Kiefer-Wolfowitz の不等式を証明する．

定理 3.10 すべての n と任意の分布 F に対して,

$$\mathbb{E}[\sup_{x\in\mathbb{R}} |\mathbb{F}_n(x) - F(x)|] \leq \frac{c}{\sqrt{n}}$$

が成立する．ただし，$\mathbb{F}_n$ は経験分布関数で，c は n と F に依存しない定数である．

証明 $x_1, x_2, \ldots, x_n \in \mathbb{R}$ を固定する．一般性を失わず，$x_1 \leq x_2 \leq \ldots \leq x_n$ とできる．また，$x = (x_1, x_2, \ldots, x_n)$ とする．$b \in \mathcal{A}(x)$ は $\{1, 1, 1, 0, \ldots, 1\}$ という形になるので，$\#\mathcal{A}(x) \leq n + 1$ となる．したがって，$\epsilon > 0$ に対して，

$$N(\epsilon, \mathcal{A}(x), \rho) \leq \left\lfloor \frac{n}{\lfloor 2n\epsilon^2 \rfloor} \right\rfloor \leq \frac{1}{2\epsilon^2} + 1$$

となる．よって，

$$\int_0^1 \sqrt{\log\{2N(t, \mathcal{A}(x), \rho)\}}\, dt \leq \int_0^1 \sqrt{\log\left(\frac{1}{t^2} + \frac{1}{2}\right)}\, dt \leq \int_0^1 \sqrt{\frac{1}{t^2}}\, dt \leq 24\sqrt{\frac{\pi}{n}}$$

となる．最後の不等号は以下からわかる．$t = 2e^{-u}$ とおけば，$u = -\log t^2$ と $dt = -2e^{-u}$ を得る．これより

$$\int_0^1 \sqrt{\log \frac{1}{t^2}}\, dt = 2\int_e^\infty \sqrt{u}e^{-u}\, du \leq 2\int_0^\infty \sqrt{u}e^{-u}\, du = \sqrt{\pi}$$

となる．最後の等式はガンマ関数の性質

$$\int_0^\infty \sqrt{u}e^{-u}\, du = \sqrt{\frac{\pi}{4}}$$

を用いた． □

·····3.5·····

相対偏差に対する不等式

定理 3.11　すべての正数 t に対して,

$$\mathbb{P}\left\{\sup_{A\in\mathcal{A}} \frac{(\mu-\mu_n)(A)}{\sqrt{\mu_n(A)}} > t\right\} \le 4\mathcal{S}_{\mathcal{A}}(2n)e^{-nt^2/4}$$

と

$$\mathbb{P}\left\{\sup_{A\in\mathcal{A}} \frac{(\mu_n-\mu)(A)}{\sqrt{\mu_n(A)}} > t\right\} \le 4\mathcal{S}_{\mathcal{A}}(2n)e^{-nt^2/4}$$

が成立する.

証明

□

系 3.1　すべての $t \in (0,1)$ と $s > 0$ に対して,

$$\mathbb{P}\left\{\sup_{A\in\mathcal{A}} \frac{(\mu-\mu_n)(A)}{(\mu+\mu_n)(A)+s/2} > \epsilon\right\} \le 4\mathcal{S}_{\mathcal{A}}(2n)e^{-n\epsilon^2/4}$$

と

$$\mathbb{P}\left\{\sup_{A\in\mathcal{A}} \frac{(\mu_n-\mu)(A)}{(\mu+\mu_n)(A)+s/2} > \epsilon\right\} \le 4\mathcal{S}_{\mathcal{A}}(2n)e^{-n\epsilon^2/4}$$

が成立する.

証明

□

系 3.2　任意の $\epsilon > 0$ と $0 < t \le 1$ に対して,

$$\mathbb{P}\left\{{}^{\exists}A \in \mathcal{A} : \mu(A) > \epsilon \quad \text{かつ} \quad \mu_n(A) = 0\right\} \le 4\mathcal{S}_{\mathcal{A}}(2n)e^{-n\epsilon^2/4}$$

が成立する.

証明

□

·····3.6·····

経験リスク最小化問題と VC 不等式

$\mathcal{G}$ は関数 $g : \mathbb{R}^d \to \{0, 1\}$ の族である．観測 $D_n = \{(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ を用いて，経験リスク

$$\hat{\mathscr{R}}_n(g) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{g(X_i) \neq Y_i\}$$

の最小化を $g \in \mathcal{G}$ に関して考える．関数 g_n^* を $\hat{\mathscr{R}}_n$ を最小化するものとする：すべての $g \in \mathcal{G}$ に対して，

$$\hat{\mathscr{R}}_n(g_n^*) \leq \hat{\mathscr{R}}_n(g)$$

である．$\mathcal{G}$ の中の最適分類子による誤りの確率を $\mathscr{R}^*$ と記す：

$$\mathscr{R}^* = \inf_{g \in \mathcal{G}} \mathscr{R}(g).$$

補題 3.1 から

$$\mathscr{R}(g_n^*) - \mathscr{R}^* \leq 2 \sup_{g \in \mathcal{G}} |\mathscr{R}_n(g) - \mathscr{R}(g)|$$

である．

いま，

$$C_g = \{(x, y) \in \mathbb{R}^d \times \{0, 1\}, g(x) \neq y\}$$

とし，$\mathscr{C} = \{C_g : g \in \mathcal{G}\}$ とする．すると

$$\sup_{g \in \mathcal{G}} |\hat{\mathscr{R}}_n(g) - \mathscr{R}(g)| = \sup_{C \in \mathscr{C}} |(\mathbb{P}_n - P)(C)|$$

となる．

さらに，集合族

$$\mathcal{A} = \{A_g : g \in \mathcal{A}\} \qquad A_g = \{x \in \mathbb{R}^d, g(x) = 1\}$$

を考える．

補題 3.2　すべての n に対して，

$$\mathcal{S}_{\mathscr{C}}(n) = \mathcal{S}_{\mathcal{A}}(n)$$

が成立する．したがって，集合族 $\mathcal{A}$ と $\mathscr{C}$ の VC - 次元は同じである．

証明　N を正の整数とする．$\mathscr{C}$ の N 個の異なる集合が $\mathbb{R}^n \times \{0, 1\}$ 上の異なる n 個の点からなる N 個の部分集合を取り出すとき，$\mathcal{A}$ の中にそれに対応する N 個の異なる集合が存在して，$\mathbb{R}^d$ 上の異なる n 個の点から N 個のことなる部分集合を取り出せることとその反対を示せばよい．

$1 \leq m \leq n$ を固定し，点 $(x_1, 0), \ldots, (x_m, 0), (x_{m+1}, 1), \ldots, (x_n, 1)$ を考える．ある $A \in \mathcal{A}$ に対して，

$$C = (A \times \{0\}) \cup (A^c \times \{0\})$$

とする．すると，$C \in \mathscr{C}$ となる．この C は

$$(x_1, 0), \ldots, (x_k, 0), (x_{m+1}, 1), \ldots, (x_{m+l}, 1)$$

を取り出せたと仮定する．ただし，$1 \leq k \leq m, 1 \leq l \leq n-m$ である．すると A は $(x_{k+1}, \ldots, x_m, x_{m+1}, \ldots, x_{m+l})$ を取り出す．このふたつの部分集合は 1 対 1 対応するので，$\mathcal{S}_{\mathscr{C}}(n) = \mathcal{S}_{\mathcal{A}}(n)$ となる．□

定理 3.12 $V_{\mathscr{C}}$ を $\mathscr{C}$ の VC - 次元とする．このとき,

$$\mathbb{E}[\mathscr{R}(g_n^*)] - \mathscr{R}^* \leq 2\sqrt{\frac{\log\{2\mathcal{S}_{\mathscr{C}}(n)\}}{n}} \leq 2\sqrt{\frac{V_{\mathscr{C}}\log(n+1) + \log 2}{n}}$$

が成立する．

証明 補題 2.1 から

$$\mathbb{E}[\mathscr{R}(g_n^*)] - \mathscr{R}^* \leq 2\mathbb{E}\sup_{g\in\mathcal{G}}|\hat{\mathscr{R}}_n(g) - \mathscr{R}(g)|$$

となる．また,

$$\sup_{g\in\mathcal{G}}|\hat{\mathscr{R}}_n(g) - \mathscr{R}(g)| = \sup_{C\in\mathscr{C}}|(\mathbb{P}_n - P)(C)|$$

に注意して，定理 3.4 を用いれば,

$$\sup_{C\in\mathscr{C}}|(\mathbb{P}_n - P(C)| \leq 2\sqrt{\frac{\log\mathcal{S}_{\mathscr{C}}(n)}{n}}$$

となる．V を $\mathscr{C}$ の VC - 次元とすれば，補題 2.1 と 補題 3.2 から

$$\sum_{k=0}^{V}\begin{pmatrix} n \\ k \end{pmatrix} \geq \mathcal{S}_{\mathcal{A}}(n) = \mathcal{S}_{\mathscr{C}}(n)$$

となる．さらに,

$$(n+1)^V = \sum_{k=0}^{V}\begin{pmatrix} V \\ k \end{pmatrix} n^k = \sum_{k=0}^{V}\frac{n^k V!}{k!(n-k)!} \geq \sum_{k=0}^{V}\frac{n^k}{k!} \geq \sum_{k=0}^{V}\frac{n!}{k!(n-k)!} = \sum_{k=0}^{V}\begin{pmatrix} n \\ k \end{pmatrix}$$

となることより定理は示された． □

定理 3.13 $V_{\mathscr{C}}$ を $\mathscr{C}$ の VC - 次元とする．このとき,

$$\mathbb{E}[\mathscr{R}(g_n^*)] - \mathscr{R}^* \leq 2c\sqrt{\frac{V_{\mathscr{C}}}{n}}$$

が成立する．

証明 定理 3.9 を用いれば，定理 3.12 の証明と同様の議論からわかる． □

·····3.7·····

モデル選択と複雑量に対する罰則

経験リスク最小化法では，与えられた関数族 $\mathcal{G}$ の中で経験リスクを最小にするものを見つける．ただし，$\mathcal{G}$ はリスクを最小にするものを含むとは限らない．この方法では，次のような状況においてうまく機能すると期待できる．

(i) 関数族 $\mathcal{G}$ は十分大きく， $\mathcal{G}$ の中で経験リスクを最小にするもののリスクは最小リスクに近い．
(ii) $\mathcal{G}$ は十分小さく，与えられたデータを用いて，経験リスクを最小にするものをうまく見つけることができる．

しかし，(i) と (ii) は相反することがつぎの式から見て取れる：

$$\mathbb{E}[\mathscr{R}(g_n^*)] - \mathscr{R}^* = (\mathbb{E}[\mathscr{R}(g_n^*)] - \inf_{g\in\mathcal{G}} \mathscr{R}(g)) + (\inf_{g\in\mathcal{G}} \mathscr{R}(g) - \mathscr{R}^*).$$

右辺の第一項目を推定誤差と呼び，第二項目を近似誤差と呼ぶ．

いま，モデルの列 $\{\mathcal{G}_k\}_{k=1}^{\infty}$ を考える．k が大きくなるとき，モデル $\mathcal{G}_k$ はより大きくなるとする．観測 D_n が与えられたとき，これらのモデルの中でよいモデルを選択するのがモデル選択の目的である．

$\hat{g}_k$ をモデル $\mathcal{G}_k$ の中で最小経験リスクを持つものとする．ここで，超過誤差 $\mathbb{E}[\mathscr{R}(\hat{g}_k)] - \mathscr{R}^*$ を

$$\min_k \mathbb{E}[\mathscr{R}(\hat{g}_k)] - \mathscr{R}^* = \min_k [\{\mathbb{E}[\mathscr{R}(\hat{g}_k)] - \inf_{g\in\mathcal{G}} \mathscr{R}(g)\} + \{\inf_{g\in\mathcal{G}} \mathscr{R}(g) - \mathscr{R}^*\}]$$

にできるだけ近いモデルを選択することを目指す．構造的リスク最小化の考え方は超過適合の補償をするために各経験リスク $\mathscr{R}_n(\hat{g}_k)$ に対して複雑量罰則を科すことである．この罰則は $\sup_{g\in\mathcal{G}_k} |\hat{\mathscr{R}}_n(g) - \mathscr{R}(g)|$ に対する分布に依存しない上限に関係し，罰則は過剰適合の影響を取り除く．

つぎの定理は，$\hat{g}_k$ のリスク $\mathscr{R}(\hat{g}_k)$ の上限の推定量がモデル $\mathcal{G}_k$ に対して複雑量罰則 $T_{n,k}$ を与え，超過誤差が

$$\min_k \left[\mathbb{E}[T_{n,k}] + (\inf_{g\in\mathcal{G}_k} \mathscr{R}(g) - \mathscr{R}^*)\right]$$

に近くなるアルゴリズムを与えることを示す．

仮定 1 ある定数 $c > 1/2$ と正整数 m が存在して，各 k に対して，$\mathscr{R}(\hat{g}_k)$ の推定量 $T_{n,k}$ が存在して，つぎを満足する：すべての $\epsilon > 0$ に対して，

$$\mathbb{P}\{\mathscr{R}(\hat{g}_k) > T_{n,k} + \epsilon\} < ce^{-2m\epsilon^2}$$

が成立する．

いま，複雑量に対する罰則を

$$\mathrm{pen}_n(k) = T_{n,k} - \hat{\mathscr{R}}(\hat{g}_k) + \sqrt{\frac{\log k}{m}}$$

で定める．上式の右辺の最後の項は技術的理由から加えただけである．$T_{n,k} - \hat{\mathscr{R}}(\hat{g}_k)$ は正しい罰則 $\mathscr{R}(\hat{g}_k) - \hat{\mathscr{R}}(\hat{g}_k)$ に対する推定量である．最後に予測則を

$$\hat{g}_n^* = \mathrm{argmin}_{k=1,2,\ldots,} \hat{\mathscr{R}}_n^{\mathrm{pen}}(\hat{g}_k)$$

で定義する．ただし，

$$\hat{\mathscr{R}}_n^{\mathrm{pen}}(\hat{g}_k) = \hat{\mathscr{R}}_n(\hat{g}_k) + \mathrm{pen}_n(k) = T_{n,k} + \sqrt{\frac{\log k}{m}}$$

である．

定理 3.14 $\mathscr{R}(\hat{g}_k)$ の推定量 $T_{n,k}$ は仮定 1 を満足するとする．このとき，

$$\mathbb{E}[\mathscr{R}(\hat{g}_n^*)] - \mathscr{R}^+ \le \min_k \left[\mathbb{E}[\mathrm{pen}_n(k) + (\inf_{g\in\mathcal{G}_k} \mathscr{R}(g) - \mathscr{R}^*) + \sqrt{\frac{\log(2ce)}{2m}}\right]$$

が成立する．

証明 $\mathscr{R}_k^* = \inf_{g\in\mathcal{G}_k} \mathscr{R}(g)$ とおく．仮定 1 を用いる：任意の $\epsilon > 0$ に対して，

$$\begin{aligned}
P\{\mathscr{R}(\hat{g}_n^*) - \hat{\mathscr{R}}_n^{\mathrm{pen}}(\hat{g}_n^*) > \epsilon\} &\le P\big\{\sup_{k=1,2,\dots} (\mathscr{R}(\hat{g}_k) - \hat{\mathscr{R}}_n^{\mathrm{pen}}(\hat{g}_k) > \epsilon\big\} \\
&\le \sum_{k=1}^{\infty} P\big\{(\mathscr{R}(\hat{g}_k) - \hat{\mathscr{R}}_n^{\mathrm{pen}}(\hat{g}_k) > \epsilon\big\} \\
&= \sum_{k=1}^{\infty} P\big\{\mathscr{R}(\hat{g}_k) - T_{n,k} > \epsilon + \sqrt{\frac{\log k}{m}}\big\} \\
&\le \sum_{k=1}^{\infty} c\exp\left[-2m\left(\epsilon + \sqrt{\frac{\log k}{m}}\right)^2\right] \\
&\le \sum_{k=1}^{\infty} c\exp\left[-2m\left(\epsilon^2 + \frac{\log k}{m}\right)\right] \\
&= ce^{-2m\epsilon^2}\sum_{k=1}^{\infty}\frac{1}{k^2} < 2ce^{-2m\epsilon^2}
\end{aligned}$$

となる．最後の不等式は

$$\sum_{k=1}^{\infty}\frac{1}{k^2} = 1 + \sum_{k=2}^{\infty}\frac{1}{k^2} < 1 + \int_1^{\infty}\frac{dx}{x^2} = 2$$

よりわかる．定理を証明するために，$\mathscr{R}(\hat{g}_n^*) - \mathscr{R}_k^*$ をつぎのように書きなおす：

$$\mathscr{R}(\hat{g}_n^*) - \mathscr{R}_k^* = (\mathscr{R}(\hat{g}_n^*) - \inf_{k=1,2,\dots} \hat{\mathscr{R}}_n(\hat{g}_k)) + (\inf_{k=1,2,\dots} \hat{\mathscr{R}}_n(\hat{g}_k) - \mathscr{R}_k^*).$$

まず，上式の右辺の第一項目を評価する：

$$P\left(\mathscr{R}(\hat{g}_n^*) - \inf_{k=1,2,\dots} \hat{\mathscr{R}}_n(\hat{g}_k) > \epsilon\right) \le 2ce^{-2m\epsilon^2}$$

に注意して，補題 1.5 を用いれば，

$$\mathbb{E}\big[\mathscr{R}(\hat{g}_n^*) - \inf_{k=1,2,\dots} \hat{\mathscr{R}}_n(\hat{g}_k)\big] \le \sqrt{\frac{1+\log(2c)}{2m}}$$

となる．つぎに，$g_k^* \in \mathcal{G}$ を選び，$\mathscr{R}(g_k^*) = \mathscr{R}_k^*$ となる[1]ようにする．このとき，

$$\begin{aligned}
\mathbb{E}\big[\inf_k \hat{\mathscr{R}}_n^{\mathrm{pen}}(\hat{g}_k)\big] - \mathscr{R}_k^* &\le \mathbb{E}[\hat{\mathscr{R}}_n(\hat{g}_k)] - \mathscr{R}_k^* + \mathbb{E}[\mathrm{pen}_n(k)] \\
&\le \mathbb{E}[\hat{\mathscr{R}}_n(\hat{g}_k)] - \mathscr{R}(g_k^*) + \mathbb{E}[\mathrm{pen}_n(k)] = \mathbb{E}[\mathrm{pen}_n(k)]
\end{aligned}$$

となる．最後の等号は

$$\mathbb{E}[\hat{\mathscr{R}}(g_k^*)] = \frac{1}{n}\mathbb{E}\big[\sum_{i=1} -\frac{1}{n}\mathbb{1}\{g_k^*(X_i) \ne y_i\}\big] = \mathscr{R}(g_k^*)$$

よりわかる．これより，定理は示された． □

[1] ここで，$\inf \mathscr{R}(g)$ に達する g_k^* の存在を仮定しているが，存在しない場合には，$\inf \mathscr{R}(G)$ に到達する列を考えれば，議論は簡単に修正できる．

3.7.1 データ分割によるモデル選択

(X'_i, Y'_i), $i = 1, 2, \ldots, m$, の観測をさらに得たとする．$(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ に基づいて $\mathcal{G}_k$ の中で経験リスクを最小にする $\hat{g}_k$ のリスクの推定量は

$$T_{n,k} = \frac{1}{m}\sum_{i=1}^{m} 1\{\hat{g}_k(X'_i) \neq Y'_i\}$$

である．$D_n = \{(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\}$ と書けば，$\mathbb{E}[T_{n,k}|D_n] = \mathscr{R}(\hat{g}_k)$ となる．また，D_n が与えられたとき，

$$\mathscr{R}(\hat{g}_k) - 1 \leq \mathscr{R}(\hat{g}_k) - T_{n,k} \leq \mathscr{R}(\hat{g}_k)$$

になることに注意して，定理 3.7 を用いれば，任意の $\epsilon > 0$ に対して，

$$\mathbb{P}\{\mathscr{R}(\hat{g}_k) - T_{n,k} > \epsilon\} = \mathbb{P}\{-(\sum_{i=1}^{m} 1\{\hat{g}_k(X'_i) \neq Y'_i\} - \mathscr{R}(\hat{g}_k)) > m\epsilon\} \leq e^{-2m\epsilon^2}$$

となる．したがって，仮定 1（$c = 1$ として）を満足することより，つぎの系を得る．

系 3.3

$$\mathbb{E}[\mathscr{R}(\hat{g}_n^*)] - \mathscr{R}^* \leq \min\left[\mathbb{E}[\mathscr{R}(\hat{g}_k) - \hat{\mathscr{R}}(\hat{g}_k)] + (\inf_{g\in\mathcal{G}_k}\mathscr{R}(g) - \mathscr{R}^*) + \sqrt{\frac{\log k}{m}}\right] + \sqrt{\frac{1+\log 2}{2m}}$$

が成立する．

証明

$$\begin{aligned}
\mathbb{E}[\mathrm{pen}_n(k)] &= \mathbb{E}[T_{n,k} - \hat{\mathscr{R}}_n(\hat{g}_k)] + \sqrt{\frac{\log k}{m}} \\
&= \mathbb{E}[\mathbb{E}[T_{n,k}|D_n] - \hat{\mathscr{R}}_n(\hat{g}_k)] + \sqrt{\frac{\log k}{m}} \\
&= \mathbb{E}[\mathscr{R}(\hat{g}_k) - \hat{\mathscr{R}}_n(\hat{g}_k)] + \sqrt{\frac{\log k}{m}}
\end{aligned}$$

と定理 3.14 より，この系は証明される． □

3.7.2 VC - 次元による罰則

Vapnik - Chervonenkis の不等式（定理 3.12）より，

$$\mathbb{E}\{\sup_{g\in\mathcal{G}}|\hat{\mathscr{R}}_n(g) - \mathscr{R}(g)|\} \leq 2\sqrt{\frac{V_{\mathcal{G}}\log(n+1) + \log 2}{n}} \tag{3.19}$$

となることから，

$$T_{n,k} = \hat{\mathscr{R}}_n(\hat{g}_k) + 2\sqrt{\frac{V_{\mathcal{G}}\log(n+1) + \log 2}{n}} \tag{3.20}$$

を用いる．$T_{n,k}$ は，仮定 1（$c=1$ と $m=n$ として）を満足することが以下の議論よりわかる：任意の $\epsilon > 0$ に対して，

$$\begin{aligned}
\mathbb{P}\{\mathscr{R}(\hat{g}_k) > T_{n,k} + \epsilon\} &= \mathbb{P}\left\{\mathscr{R}(\hat{g}_k) - \hat{\mathscr{R}}(\hat{g}_k) > 2\sqrt{\frac{V_{\mathcal{G}}\log(n+1)+\log 2}{n}} + \epsilon\right\} \\
&\le \mathbb{P}\left\{\sup_{g\in\mathcal{G}_k}|\mathscr{R}(g) - \hat{\mathscr{R}}_n(g)| > 2\sqrt{\frac{V_{\mathcal{G}}\log(n+1)+\log 2}{n}} + \epsilon\right\} \\
&\le \mathbb{P}\left\{\sup_{g\in\mathcal{G}_k}|\mathscr{R}(g) - \hat{\mathscr{R}}_n(g)| > \mathbb{E}\sup_{g\in\mathcal{G}_k}|\mathscr{R}(g) - \hat{\mathscr{R}}_n(g)| + \epsilon\right\} \\
&\le e^{-2n\epsilon^2}
\end{aligned}$$

が成立する．最後から二番目の不等式は (3.19) よりわかり，最後の不等式は定理 3.9 をつぎのように適用すれば示すことができる．

$$h(x_1, x_2, \ldots, x_n) = \sup_{g\in\mathcal{G}_k}|\mathscr{R}(g) - \hat{\mathscr{R}}(g)|$$

とおけば，

$$\sup_{x_1, x_2, \ldots, x_n, x_i'}|h(x_1, x_2, \ldots, x_n) - h(x_1, x_2, \ldots, x_i', \ldots, x_n)| \le \frac{1}{n}$$

となる．したがって，$c_i = 1/n, i = 1, 2, \ldots, n,$ として，定理 3.9 を用いればよい．

以上の議論により，仮定 1が満たされるので，定理 3.14 からつぎの系を得る．

系 3.4 $T_{n,k}$ は (3.20) で与えられるとき，

$$\begin{aligned}
\mathbb{E}[\mathscr{R}(g_n^*)] - \mathscr{R}^* \le \min_k &\left[2\sqrt{\frac{V_{\mathcal{G}}\log(n+1)+\log 2}{n}} + (\inf_{g\in\mathcal{G}_k}\mathscr{R}(g) - \mathscr{R}^*) + \sqrt{\frac{\log k}{n}}\right] \\
&+\sqrt{\frac{1}{2n}}
\end{aligned}$$

が成立する．

3.7.3 最大 discrepancy による罰則

議論を簡単にするために，n を偶数とする．任意の分類子 g に対して，

$$\hat{\mathscr{R}}_n^{(1)}(g) = \frac{2}{n}\sum_{i=1}^{n/2}1\{g(X_i) \ne Y_i\}$$

と

$$\hat{\mathscr{R}}_n^{(2)}(g) = \frac{2}{n}\sum_{i=n/2+1}^{n}1\{g(X_i) \ne Y_i\}$$

とおく．いま，$T_{n,k}$ を

$$T_{n,k} = \hat{\mathscr{R}}_n(\hat{g}_k) + \sup_{g\in\mathcal{G}_k}(\hat{\mathscr{R}}_n^{(1)}(g) - \hat{\mathscr{R}}_n^{(2)}(g)) \tag{3.21}$$

で定める．(3.21) の右辺の第二項目を求めるために，

$$D_n' = \{(X_1,', Y_1'), \ldots, (X_n', Y_n')\}$$

として，$i=1,2,\dots,n$ に対して，

$$Y_i' = \begin{cases} 1-Y_i, & i \le n/2, \\ Y_i, & i > n/2 \end{cases}, \qquad X_i' = X_i, \qquad i=1,2,\dots,n$$

と定める．このとき，任意の g に対して，

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{g(X_i') \ne Y_i'\} &= \frac{1}{2} - \frac{1}{n}\sum_{i=1}^{n/2} \mathbb{1}\{g(X_i) \ne Y_i\} + \frac{1}{n}\sum_{i=n/2+1}^{n} \mathbb{1}\{g(X_i) \ne Y_i\} \\ &= \frac{1}{2}\left\{1 - (\hat{\mathscr{R}}_n^{(1)}(g) - \hat{\mathscr{R}}_n^{(2)}(g))\right\}\end{aligned}$$

となることより，データ D_n' に関して，$\hat{\mathscr{R}}_n(g)$ を最小にする分類子は $\hat{\mathscr{R}}_n^{(1)}(g) - \hat{\mathscr{R}}_n^{(2)}(g)$ を最大にすることがわかる．

系 3.5 $T_{n,k}$ が (3.21) で与えられたとき，

$$\begin{aligned}\mathbb{E}[\mathscr{R}(g_n^*)] - \mathscr{R}^* \le \min_k &\left[\mathbb{E}\sup_{g\in\mathcal{G}_k}(\hat{\mathscr{R}}_n^{(1)}(g) - \hat{\mathscr{R}}_n^{(2)}(g)) + (\inf_{g\in\mathcal{G}_k}\mathscr{R}(g) - \mathscr{R}^*) + \sqrt{\frac{2\log k}{n}}\right] \\ &+\sqrt{\frac{21\log(1+\log 3)}{2n}}\end{aligned}$$

が成立する．

証明 仮定 1 を確認して，定理 3.14 を適用すればよい．仮定 1 を確認するために，$(X_1,', Y_1'), \dots, (X_n', Y_n')$ を観測値とは独立な複製とし，

$$\hat{\mathscr{R}}_n'(g) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{g(X_i') \ne Y_i'\}$$

とおく．各 k に対して，

$$\begin{aligned}\mathbb{E}\left[\sup_{g\in\mathcal{G}_k}(\hat{\mathscr{R}}_n'(g) - \hat{\mathscr{R}}_n(g)\right] &= \frac{1}{n}\mathbb{E}\left[\sup_{g\in\mathcal{G}_k}\sum_{i=1} -n(\mathbb{1}\{g(X_i') \ne Y_i'\} - \mathbb{1}\{g(X_i) \not Y_i\})\right] \\ &\le \frac{1}{n}\Bigg[\mathbb{E}\{\sup_{g\in\mathcal{G}_k}\sum_{i=1}^{n/2}(\mathbb{1}\{g(X_i') \ne Y_i'\} - \mathbb{1}\{g(X_i) \ne Y_i\})\} \\ &\qquad + \mathbb{E}\{\sup_{g\in\mathcal{G}_k}\sum_{i=n/2+1}^{n}(\mathbb{1}\{g(X_i') \ne Y_i'\} - \mathbb{1}\{g(X_i) \ne Y_i\})\}\Bigg] \\ &= \frac{2}{n}\mathbb{E}\{\sup_{g\in\mathcal{G}_k}\sum_{i=1}^{n/2}(\mathbb{1}\{g(X_i') \ne Y_i'\} - \mathbb{1}\{g(X_i) \ne Y_i\})\} \\ &= \mathbb{E}\big[\sup_{g\in\mathcal{G}_k}(\hat{\mathscr{R}}_n^{(1)}(g) - \hat{\mathscr{R}}_n^{(2)}(g)\big]\end{aligned}$$

となる．ここで，定理 1.3 を用いる：

$$h(x_1, x_2, \dots, x_n, x_1', x_2', \dots, x_n') = \sup_{g\in\mathcal{G}_k}(\hat{\mathscr{R}}_n'(g) - \hat{\mathscr{R}}_n(g))$$

とおけば，

$$\begin{aligned}\sup_{x_1,x_2,\dots,x_n,x_1',x_2',\dots,x_n',y} &h(x_1, x_2, \dots, x_n, x_1', x_2', \dots, x_n') \\ &-h(x_1, x_2, \dots, x_n, x_1', x_2', \dots, y, \dots, x_n') \le \frac{1}{n}\end{aligned}$$

となる．したがって，$c_i = 1/n(i = 1, 2, \ldots, 2n)$ より

$$\mathbb{P}\left\{\sup_{g\in\mathcal{G}_k}(\hat{\mathscr{R}}'_n(g) - \hat{\mathscr{R}}_n(g)) > \mathbb{E}\big[\sup_{g\in\mathcal{G}_k}(\hat{\mathscr{R}}'_n(g) - \hat{\mathscr{R}}_n(g))\big] + \epsilon\right\} \le e^{-n\epsilon^2} \tag{3.22}$$

を得る．同様に，

$$\mathbb{P}\left\{\sup_{g\in\mathcal{G}_k}(\hat{\mathscr{R}}^{(1)}_n(g) - \hat{\mathscr{R}}^{(2)}_n(g)) > \mathbb{E}\big[\sup_{g\in\mathcal{G}_k}(\hat{\mathscr{R}}^{(1)}_n(g) - \hat{\mathscr{R}}^{(2)}_n(g))\big] + \epsilon\right\} \le e^{-n\epsilon^2} \tag{3.23}$$

を得る．これらより

$$\begin{aligned}
\mathbb{P}\{\hat{\mathscr{R}}(g) > T_{n,k} + \epsilon\} &= \mathbb{P}\left\{\mathscr{R}(\hat{g}_k) - \hat{\mathscr{R}}_n(\hat{g}_k) > \sup_{g\in\mathcal{G}_k}(\hat{\mathscr{R}}^{(1)}_n(g) - \hat{\mathscr{R}}^{(2)}_n(g)) + \epsilon\right\} \\
&\le \mathbb{P}\left\{\hat{\mathscr{R}}'_n(\hat{g}_k) - \hat{\mathscr{R}}_n(\hat{g}_k) > \sup_{g\in\mathcal{G}_k}(\hat{\mathscr{R}}^{(1)}_n(g) - \hat{\mathscr{R}}^{(2)}_n(g)) + \epsilon\right. \\
&\qquad\qquad \left.\text{かつ}\quad \mathscr{R}(\hat{g}_k) - \hat{\mathscr{R}}'_n(\hat{g}_k) \le \frac{2\epsilon}{9}\right\} \\
&\quad + \mathbb{P}\left\{\mathscr{R}(\hat{g}_k) - \hat{\mathscr{R}}'_n(\hat{g}_k) > \frac{2\epsilon}{9}\right\} \\
&\le \mathbb{P}\left\{\hat{\mathscr{R}}'_n(\hat{g}_k) - \hat{\mathscr{R}}_n(\hat{g}_k) > \sup_{g\in\mathcal{G}_k}(\hat{\mathscr{R}}^{(1)}_n(g) - \hat{\mathscr{R}}^{(2)}_n(g)) + \frac{7\epsilon}{9}\right\} \\
&\quad + \mathbb{P}\left\{\mathscr{R}(\hat{g}_k) - \hat{\mathscr{R}}'_n(\hat{g}_k) > \frac{2\epsilon}{9}\right\} \\
&\le \mathbb{P}\left\{\hat{\mathscr{R}}'_n(\hat{g}_k) - \hat{\mathscr{R}}_n(\hat{g}_k) > \sup_{g\in\mathcal{G}_k}(\hat{\mathscr{R}}^{(1)}_n(g) - \hat{\mathscr{R}}^{(2)}_n(g)) + \frac{7\epsilon}{9}\right\} \\
&\quad + \exp\left[-\frac{8n\epsilon^2}{81}\right]
\end{aligned}$$

を得る．最後の不等式は Hoefffding の不等式よりわかる．さらに，最右辺の第一項目を評価する：

$$\begin{aligned}
&\mathbb{P}\left\{\hat{\mathscr{R}}'_n(\hat{g}_k) - \hat{\mathscr{R}}_n(\hat{g}_k) > \sup_{g\in\mathcal{G}_k}(\hat{\mathscr{R}}^{(1)}_n(g) - \hat{\mathscr{R}}^{(2)}_n(g)) + \frac{7\epsilon}{9}\right\} \\
&\quad\le \mathbb{P}\left\{\sup_{g\in\mathcal{G}_k}\hat{\mathscr{R}}'_n(g) - \hat{\mathscr{R}}_n(g) > \sup_{g\in\mathcal{G}_k}(\hat{\mathscr{R}}^{(1)}_n(g) - \hat{\mathscr{R}}^{(2)}_n(g)) + \frac{7\epsilon}{9}\right\} \\
&\quad\le \mathbb{P}\left\{\sup_{g\in\mathcal{G}_k}\hat{\mathscr{R}}'_n(g) - \hat{\mathscr{R}}_n(g) > \sup_{g\in\mathcal{G}_k}(\hat{\mathscr{R}}^{(1)}_n(g) - \hat{\mathscr{R}}^{(2)}_n(g)) + \frac{7\epsilon}{9}\right. \\
&\qquad\qquad \left.\text{かつ}\quad \sup_{g\in\mathcal{G}_k}(\hat{\mathscr{R}}^{(1)}_n(g) - \hat{\mathscr{R}}^{(2)}_n(g)) \ge \mathbb{E}\big[\sup_{g\in\mathcal{G}_k}(\hat{\mathscr{R}}^{(1)}_n(g) - \hat{\mathscr{R}}^{(2)}_n(g))\big] - \frac{4\epsilon}{9}\right\} \\
&\quad + \mathbb{P}\left\{\sup_{g\in\mathcal{G}_k}(\hat{\mathscr{R}}^{(1)}_n(g) - \hat{\mathscr{R}}^{(2)}_n(g)) < \mathbb{E}\big[\sup_{g\in\mathcal{G}_k}(\hat{\mathscr{R}}^{(1)}_n(g) - \hat{\mathscr{R}}^{(2)}_n(g))\big] - \frac{4\epsilon}{9}\right\} \\
&\quad\le \mathbb{P}\left\{\sup_{g\in\mathcal{G}_k}\hat{\mathscr{R}}'_n(g) - \hat{\mathscr{R}}_n(g) > \sup_{g\in\mathcal{G}_k}(\hat{\mathscr{R}}^{(1)}_n(g) - \hat{\mathscr{R}}^{(2)}_n(g)) + \frac{\epsilon}{3}\right\} \\
&\quad + \mathbb{P}\left\{\sup_{g\in\mathcal{G}_k}(\hat{\mathscr{R}}^{(1)}_n(g) - \hat{\mathscr{R}}^{(2)}_n(g)) < \mathbb{E}\big[\sup_{g\in\mathcal{G}_k}(\hat{\mathscr{R}}^{(1)}_n(g) - \hat{\mathscr{R}}^{(2)}_n(g))\big] - \frac{4\epsilon}{9}\right\} \\
&\quad \exp\left[-\frac{n\epsilon^2}{9}\right] + \exp\left[\frac{8n\epsilon^2}{81}\right]
\end{aligned}$$

となる．最後の不等式は (3.21) と (3.22) からわかる．最後に得られた不等式は合わせれば，

$$\mathbb{P}\big\{\hat{\mathscr{R}}(g) > T_{n,k} + \epsilon\big\} \le 3\exp\left[-\frac{8n\epsilon^2}{81}\right]$$

を得る．したがって，$m = n/21$ と $C = 3$ として，定理 3.14 そ用いれば，この系は証明される． □

4

経験過程理論の数理統計への応用

·····4.1·····

距離空間における分布収束

定義 4.1 ：空でない集合 $\mathbb{D}$ とつぎの性質 (i) ∼ (iii) をもつ写像 $d : \mathbb{D} \times \mathbb{D} \longrightarrow [0, \infty)$ があるとき、d を $\mathbb{D}$ 上の**距離**といい、$\mathbb{D}$ を**距離空間**という．

(i) ．任意の $x, y \in \mathbb{D}$ に対し、$d(x, y) = 0 \Leftrightarrow x = y$ が成立する．
(ii) ．任意の $x, y \in \mathbb{D}$ に対し、$d(x, y) = d(y, x)$ が成立する．
(iii) ．任意の $x, y, z \in \mathbb{D}$ に対し、$d(x, z) \leq d(x, y) + d(y, z)$ が成立する.

r を正の数とし、集合 $\{y : d(x, y) < r\}$ を**開球**という．A を $\mathbb{D}$ の部分集合とし、$x \in \mathbb{D}$ が A の**内点**であるとは、ある $r > 0$ に対して、$\{y : d(x, y) < r\} \subset A$ が成立することである．A の各点が A の内点であるとき、A は**開集合**であるという．A の補集合が開集合のとき、A を**閉集合**であるという．A に含まれる最大の開集合を A の**内部**といい、$\mathring{A}$ で記す．点 $x \in \mathbb{M}$ が A の**集積点**であるとは、任意の r に対しても $0 < d(z, x) < r$ を満たす点 $z \in A$ が存在することである．A とその集積点全体の集合の和集合を A の**閉包**とよび、$\overline{A}$ で記す．点列 $\{x_n\}_{n=1}^{\infty}$ に対して、$d(x_n, x) \to 0 (n \to \infty)$ なる点 $x \in \mathbb{D}$ が存在するとき、$\{x_n\}$ は x に**収束する**といい、$x_n \to x$ と記す．部分集合 A が $\mathbb{D}$ において**稠密**であるとは、$\overline{A} = \mathbb{D}$ が成立することである．$\mathbb{D}$ が**可分**であるとは、$\mathbb{D}$ のあるたかだか可算部分集合が $\mathbb{D}$ において稠密であることをいう．$\mathbb{D}$ のどんな Cauchy 列[1)] も $\mathbb{M}$ のある点に収束するとき、$\mathbb{D}$ は**完備**であるという．$\mathbb{D}$ の部分集合 K が**コンパクト**であるとは、K の任意の開被覆から K の有限開被覆を選び出せることである．K が**全有界**であるとは、任意の $\epsilon > 0$ に対しても K の中から有限個の点 $x_1, \ldots, x_m$ を選び出して、$K \subset \bigcup_{i=1}^{m} \{y : d(y, x_i) < r\}$ とできることである．K が**点列コンパクト**であるとは、K の中のどんな無限点列 $\{x_n\}_{n=1}^{\infty}$ からも、$\mathbb{D}$ の中のある点に収束する部分列 $\{x_{n_j}\}_{j=1}^{\infty}$ が選びだせることである[2)]. $(\mathbb{D}_1, d_1)$ と $(\mathbb{D}_2, d_2)$ を 2 つの距離空間とする．写像 $f : \mathbb{D}_1 \longrightarrow \mathbb{D}_2$ が点 $x_0 \in \mathbb{D}_1$ において**連続**であるとは、どんな正の数 ϵ に対しても適当に正の数 δ をとって、$d_1(x, x_0) < \delta$ ならば、$d_2(f(x), f(x_0)) < \epsilon$ となるようにできることである．

定義 4.2 ：線形空間 $\mathbb{X}$ で定義された写像 $\| \cdot \| : \mathbb{X} \longrightarrow [0, \infty)$ がつぎの性質 (i) ∼ (iii) を満たすとき、$\| \cdot \|_X$ を $\mathbb{X}$ の**ノルム** とよび、ノルムが定義されている線形空間 $\mathbb{X}$ を**ノルム空間**いう．

(i) ．任意の $x \in \mathbb{X}$ に対して、$\|x\|_X \Leftrightarrow x = 0$．
(ii) ．任意の $x, y \in \mathbb{X}$ に対して、$\|x + y\|_X \leq \|x\|_X + \|y\|_X$．
(iii) ．任意の係数 α と任意の $x \in \mathbb{X}$ に対して、$\|\alpha x\|_X = |\alpha| \|x\|_X$．

[1)] 点列 $\{x_n\}$ が Cauchy 列であるとは、どんな正の数 ϵ に対しても、それに応じて適当な自然数 N をとって、$j, k > N$ ば、$d(x_j, x_k) < \epsilon$ となるようにできることである．

[2)] これら 3 つの概念にはつぎのような同値関係がある．(i) ∼ (iii) は同値である．(i)．K は点列コンパクトである．(ii)．$\overline{K}$ はコンパクトである．(iii)．K は全有界かつ $\overline{K}$ は完備である．

·····4.2·····

分布の収束

$(\Omega, \mathscr{F}, P)$ と $(\Omega_n, \mathscr{F}_n, P_n) \quad (n = 1, 2, \ldots,)$ を確率空間とする．

定義 4.3 ：$\mathbb{D}$ 上の Borel σ– 集合体とは、$\mathbb{D}$ 上の開集合全体を含む最小の σ – 加法族のことである．これを $\mathscr{B}(\mathbb{M})$ と記す．確率空間 $(\Omega, \mathscr{A}, P)$ 上の写像 $X : \Omega \longrightarrow \mathbb{D}$ が Borel 可測[1] ならば、X を $\mathbb{D}$ – 値 random element とよぶ．

任意の写像 $U : \Omega \longrightarrow \overline{\mathbb{R}}$ に対して[2]、**外期待値**（**outer expectation** ）を

$$\mathbb{E}^* U = \inf\{\mathbb{E}V : V \geq U, V : \Omega \longrightarrow \overline{\mathbb{R}} \text{ は可測で、} \mathbb{E}V \text{ が存在 }\}$$

で定義する[3]．

定義 4.4 ： $(\mathbb{D}, d)$ を距離空間とする．
（分布収束）：任意の写像の列 $X_n : \Omega_n \longrightarrow \mathbb{D}$ がある random element X に**分布収束**するとは、任意の有界連続関数 f に対して、$n \to \infty$ のとき、

$$\mathbb{E}^* f(X_n) \to \mathbb{E} f(X)$$

が成立することである．ただし、X は Borel 可測とする．$X_n \rightsquigarrow X$ と記す．
（確率収束）：どんな正の数 ϵ に対しても、$n \to \infty$ のとき、

$$\mathbb{P}^*(d(X_n, X) > \epsilon) \to 0$$

が成立するとき、X_n は X に**確率収束**するという．$X_n \xrightarrow{P} X$ と記す．
（概収束）：ある可測な確率変数の列 $\{\Delta_n\}$ が存在して、$n \to \infty$ のとき、

$$d(X_n, X) \leq \Delta_n \qquad \Delta_n \xrightarrow{as} 0$$

を満足するとき、X_n は X に**概収束**するという．$X_n \xrightarrow{as^*} X$ と記す．

補題 4.1 （Portmanteau の補題）：つぎの条件は同値である．
(i)．すべての有界連続関数に対して、$\mathbb{E}^* f(X_n) \to \mathbb{E} f(X)$．
(ii)．すべての有界 Lipshitz 関数に対して、$\mathbb{E}^* f(X_n) \to \mathbb{E} f(X)$．
(iii)．すべての開集合 G に対して、$\liminf_{n\to\infty} \mathbb{P}_*(X_n \in G) \geq \mathbb{P}(X \in G)$
(iv)．すべての閉集合 F に対して、$\limsup_{n\to\infty} \mathbb{P}^*(X_n \in F) \leq \mathbb{P}(X \in F)$．
(v)．すべての $\mathbb{P}(\delta B = 0)$ なる Borel 可測集合 B に対して、$\mathbb{P}^*(X_n \in B) \to \mathbb{P}(X \in B)$．

定理 4.1 $X_n : \Omega_n \longrightarrow \mathbb{D}$ を任意の写像とし、X を $\mathbb{D}$ – 値 random element とする．
(i)．$X_n \xrightarrow{as^*} X$ ならば、$X_n \xrightarrow{P} X$．
(ii)．$X_n \xrightarrow{P} X$ ならば、$X_n \rightsquigarrow X$．
(iii)．c をある定数とする．このとき、$X_n \xrightarrow{P} c \Leftrightarrow X_n \rightsquigarrow c$．
(iv)．$X_n \rightsquigarrow X$ かつ $d(X_n, Y_n) \xrightarrow{P} 0$ ならば、$Y_n \rightsquigarrow X$．
(v)．c をある定数とする．$X_n \rightsquigarrow X$ かつ $Y_n \xrightarrow{P} c$ ならば、$(X_n, Y_n) \rightsquigarrow (X, c)$
(vi)．$X_n \xrightarrow{P} X$ かつ $Y_n \xrightarrow{P} Y$ ならば、$(X_n, Y_n) \xrightarrow{P} (X, Y)$．

---

[1] 任意の $C \in \mathscr{B}(\mathbb{D})$ に対しても $X^{-1}(C) \in \mathscr{A}$ が成立する．
[2] $\overline{\mathbb{R}}$ は、実数 $\mathbb{R}$ に $\{\infty\}$ と $\{-\infty\}$ を加えたもの．
[3] $\mathbb{E}\max(V, 0)$ と $\mathbb{E}\max(0, -V)$ の少なくと一方は有限のとき、$\mathbb{E}V$ が存在するという．

定理 4.2 （連続写像定理）: $D_n \subset \mathbb{D}$ を任意の部分集合とし、$g_n : \mathbb{D} \longrightarrow \mathbb{D}$ を任意の写像で、ある点 $x \in D_0$ に収束するどんなの点列 $x_n \in D_n$ に対しも、$g_n(x_n) \to g_0(x)$ を満足するとする．このとき、任意の写像 $X_n : \Omega_n \longrightarrow D_n$ と $\mathbb{D}$ – 値 random element X （したがって、$g_0(X)$ は $\mathbb{D}$ – 値 random element）に対し、
(i)．$X_n \rightsquigarrow X$ ならば、$g_n(X_n) \rightsquigarrow g_0(X)$
(ii)．$X_n \xrightarrow{P} X$ ならば、$g_n(X_n) \xrightarrow{P} g_0(X)$
(iii)．$X_n \xrightarrow{as^*} X$ ならば、$g_n(X_n) \xrightarrow{as^*} g_0(X)$
が成立する．

距離空間 $(\boldsymbol{M}, d)$ に値を取る Borel 可測な random element X が **tight** であるとは、どんな正の数 $\epsilon > 0$ に対しても、あるコンパクト集合 K が存在して、$P(X \notin K) < \epsilon$ を満足することである．また、任意の写像の列 $X_n : \Omega \longrightarrow \boldsymbol{M}$ が**漸近的に tight** であるとは、どんな正の数 $\epsilon > 0$ に対しても、あるコンパクトな集合 K が存在して、すべての $\delta > 0$ に対し、

$$\limsup_{n\to\infty} P^*(X \notin K^\delta) < \epsilon$$

を満足することである．ただし、$K^\delta = \{y : d(y, K) < \delta\}$ である．任意の写像の列 $X_n : \Omega \longrightarrow \boldsymbol{M}$ が**漸近可測**であるとは、すべての有界連続関数 $f : \boldsymbol{M} \longrightarrow \boldsymbol{R}$ に対し、$n \to \infty$ のとき、

$$\boldsymbol{E}^* f(X_n) - \boldsymbol{E}_* f(X_n) \to 0$$

を満足する．ただし、任意の写像 $U : \Omega \longrightarrow \boldsymbol{R}$ に対し、$\boldsymbol{E}_* U = \sup\{\boldsymbol{E} V : V : \Omega \longrightarrow \boldsymbol{R}$ は可測で $V \leq U$ を満足$\}$ とした．

定理 4.3 ：任意の写像 $X_n : \Omega_n \longrightarrow l^\infty(T)$ の列が tight な random element に弱収束するための必要十分条件はつぎのふたつである．
(i)　T のどんな有限個の点 $t_1, \ldots, t_k$ に対しても、列 $(X_{n,t_1}, \ldots, X_{n,t_k})$ は分布収束する．
(ii)　どんな $\epsilon > 0, \eta > 0$ に対しても、半径が $L_2(P)$ – norm に関して ϵ 以下の集合 $T_1, \ldots, T_k$ による T の分割 $T = \bigcup_{i=1}^k T_i$ が存在し、

$$\limsup_{n\to\infty} P^*\left(\sup_i \sup_{t,t' \in T_i} |X_{n,t} - X_{n,t'}| \geq \epsilon\right) \leq \eta$$

が成立する．

定理 4.4 （Prohorov の定理）: $X_n : \Omega \longrightarrow \boldsymbol{M}$ は任意の写像とする．
(i)．X をある tight な random element が存在し、$X_n \rightsquigarrow X$ ならば、$\{X_n\}_{n=1}^\infty$ は漸近 tight で漸近可測である．
(ii)．X_n が漸近 tight で漸近可測ならば、ある部分列 $\{X_{n_j}\}_{j=1}^\infty$ と tight な random element X が存在して、$j \to \infty$ のとき、$X_{n_j} \rightsquigarrow X$ が成立する．

·····4.3·····

Donsker の定理

・・・・・4.4・・・・・

平均偏差の和の極限分布

……4.5……

Z - 推定量の極限分布に関する Huber - Poland の定理

·····4.6·····

Donsker の定理の証明

参考文献