確率数理要論 3

- 独立性と大数の法則

- 独立性

Λ を任意の集合（非可算集合も含む）とする.

<u>Def</u>（事象の独立性）

事象の族 $(A_\lambda)_{\lambda\in\Lambda}$ が独立（independent）とは,

$$\forall k\in\mathbb{N},\ \forall i_1,\dots,i_k\in\Lambda$$

$$P(A_{i_1}\cap A_{i_2}\cap\cdots\cap A_{i_k})=P(A_{i_1})\times\cdots\times P(A_{i_k})$$

が成り立つときをいう.

//

<u>Def</u>（σ-加法族の独立性）

σ-加法族の族 $(\mathcal{F}_\lambda)_{\lambda\in\Lambda}$ が独立とは,

$$\forall k\in\mathbb{N},\ \forall i_1,\dots,i_k\in\Lambda,\ \forall C_j\in\mathcal{F}_{i_j}\ (j=1,\dots,k),$$

$$P(C_1\cap\cdots\cap C_k)=P(C_1)\times\cdots\times P(C_k)$$

が成り立つときをいう.

//

例：積測度

2つの確率空間 $(\Omega_1,\mathcal{F}_1,P_1)$, $(\Omega_2,\mathcal{F}_2,P_2)$ に対し,

$\Omega_1\times\Omega_2$ 上の直積 σ-加法族 $\mathcal{F}_1\otimes\mathcal{F}_2$ 上に

$$P(A_1\times A_2)=P_1(A_1)\times P_2(A_2)\qquad(\forall A_1\in\mathcal{F}_1,\ \forall A_2\in\mathcal{F}_2)$$

を満たす測度 P が一意に存在する. これを直積測度と言う.

（Hopfの拡張定理を使う）

→ $\tilde{\mathcal{F}}_1=\{A_1\times\Omega_2\mid A_1\in\mathcal{F}_1\}$, $\tilde{\mathcal{F}}_2=\{\Omega_1\times A_2\mid A_2\in\mathcal{F}_2\}$ とすれば, $\tilde{\mathcal{F}}_1$ と $\tilde{\mathcal{F}}_2$ は独立.

Def (確率変数の独立性)

r.v. の族 $(X_\lambda)_{\lambda\in\Lambda}$ が独立とは、

X_λを可測にする最小のσ-加法族

$$\sigma(X_\lambda) := \{ X_\lambda^{-1}(A) \mid A \in \mathcal{B}(\mathbb{R}) \}$$

に対して、$(\sigma(X_\lambda)_{\lambda\in\Lambda})_{\lambda\in\Lambda}$ が独立となることである。 //

注: $\sigma(X_\lambda)$ はσ-加法族になっていることに注意 (前回講義)

Thm

$X_1, \dots, X_n$ が独立

$$\iff F_{X_1,\dots,X_n}(x_1,\dots,x_n) = F_{X_1}(x_1)\times\cdots\times F_{X_n}(x_n) \quad ({}^\forall x_i \in \mathbb{R})$$

$(X_1,\dots,X_n)$の同時分布関数　　X_iの分布関数

Proof

($\Rightarrow$) 分布関数の定義から自明

($\Leftarrow$)

π-λ定理を用いる。

- π-システム: 集合族$\mathcal{P}$がπ-システム$\iff$ $A, B \in \mathcal{P}$ なら $A\cap B \in \mathcal{P}$
- λ-システム: 集合族$\mathcal{L}$がλ-システム$\iff$
  (1) $\Omega \in \mathcal{L}$
  (2) $A, B \in \mathcal{L}$, $A \subset B \Rightarrow \underbrace{B\setminus A}_{= B\cap A^c} \in \mathcal{L}$
  (3) $A_1 \subset A_2 \subset A_3 \subset \cdots \in \mathcal{L}$
  $\Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{L}$

注: σ-加法族はπ-システムでもλ-システムでもある。

Thm (Dynkinのπ-λ定理)

あるλ-システム$\mathcal{L}$が、あるπ-システム$\mathcal{P}$を含む ($\mathcal{P} \subset \mathcal{L}$) なら、

$\sigma(\mathcal{P}) \subset \mathcal{L}$ である。 //

証明は第2回目補足資料を参照のこと。

2変数で示す。多変数の場合は同様に帰納法で示せる。

$\mathcal{A}_1 = \{(-\infty, x_1] \mid x_1 \in \mathbb{R}\}$, $\mathcal{A}_2 = \{(-\infty, x_2] \mid x_2 \in \mathbb{R}\}$ とする。

$\forall A \in \mathcal{A}_1$, $\forall B \in \mathcal{A}_2$ に対し、$P(X_1 \in A, X_2 \in B) = P(X_1 \in A)P(X_2 \in B)$ で、

また、$\sigma(\mathcal{A}_1) = \mathcal{B}(\mathbb{R})$, $\sigma(\mathcal{A}_2) = \mathcal{B}(\mathbb{R})$ であることにも注意する。

今、$A_2 \in \mathcal{A}_2$ を1つ固定して、

$$\mathcal{L} = \{A \in \mathcal{B}(\mathbb{R}) \mid P(X_1 \in A \cap X_2 \in A_2) = P(X_1 \in A)P(X_2 \in A_2)\}$$

とする。$\mathcal{L} \supset \mathcal{A}_1$ であることは独立性の定義から従う。

$\mathcal{L}$ がλ-システムであることが示せれば、π-λ定理より、$\sigma(\mathcal{A}_1) = \mathcal{B}(\mathbb{R}) \subset \mathcal{L}$ が言える。

(a) $\Omega = \mathbb{R} \in \mathcal{L}$. なぜなら $P(X_1 \in \Omega, X_2 \in A) = P(X_2 \in A)$

$$= \underbrace{P(X_1 \in \Omega)}_{=1} P(X_2 \in A)$$

(b) $A, B \in \mathcal{L}$ かつ $A \subset B$ とする。このとき、

$$P(X_1 \in (B \setminus A), X_2 \in A_2) = P((\{X_1 \in B\} \cap \{X_2 \in A_2\}) \setminus (\{X_1 \in A\} \cap \{X_2 \in A_2\}))$$

$$= P(X_1 \in B, X_2 \in A_2) - P(X_1 \in A, X_2 \in A_2)$$

$A, B \in \mathcal{L}$ より

$$= P(X_1 \in B)P(X_2 \in A_2) - P(X_1 \in A) \cdot P(X_2 \in A_2)$$

$$= (P(X_1 \in B) - P(X_1 \in A))P(X_2 \in A_2)$$

$$= P(X_1 \in B \setminus A) \cdot P(X_2 \in A_2)$$

よって、$B \setminus A \in \mathcal{L}$

(c) $\hat{A}_1 \subset \hat{A}_2 \subset \cdots \in \mathcal{L}$ のとき、

$X^{-1}(\bigcup A_n) = \bigcup X^{-1}(A_n)$ より

$$P(X_1 \in \bigcup_{n=1}^{\infty} \hat{A}_n, X_2 \in A_2) = P\left[\left(\bigcup_{n=1}^{\infty} \{X_1 \in \hat{A}_1\}\right) \cap \{X_2 \in A_2\}\right]$$

$$= P\left(\bigcup_{n=1}^{\infty} (\{X_1 \in \hat{A}_n\} \cap \{X_2 \in A_2\})\right)$$

$$= \lim_{n \to \infty} P(\{X_1 \in \hat{A}_n\} \cap \{X_2 \in A_2\})$$ (∵確率の連続性 と $\hat{A}_n$ の単調性)

$$= \lim_{n \to \infty} P(X_1 \in \hat{A}_n) \cdot P(X_2 \in A_2)$$

$$= P(X_1 \in \bigcup_{n=1}^{\infty} \hat{A}_n) \cdot P(X_2 \in A_2)$$

こうも $\bigcup_{n=1}^{\infty} X^{-1}(A_n) = X^{-1}(\bigcup_{n=1}^{\infty} A_n)$ も使う

よって、$\bigcup_{n=1}^{\infty} \hat{A}_n \in \mathcal{L}$ である。

以上より、$\mathcal{L}$ はλ-システムなので、$\sigma(\mathcal{A}_1) \subset \mathcal{L}$ が言えた。

つまり、$\forall A \in \mathcal{B}(\mathbb{R})$, $\forall B \in \mathcal{A}_2$ に対し、$P(A \in X_1, B \in X_2) = P(A \in X_1) \cdot P(B \in X_2)$.

さらに、同じ議論を、$A \in \mathcal{B}(\mathbb{R})$ を固定して、$\mathcal{A}_2$ の方に適用し、

$\sigma(\mathcal{A}_1) = \mathcal{B}(\mathbb{R})$ と $\sigma(\mathcal{A}_2) = \mathcal{B}(\mathbb{R})$ まで拡張すれば良い。 //

(⇐) の別証明.

$Y_1, \dots, Y_n$ を $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上の r.v. で、

$P_i(Y_i \in A_i) = P(X_i \in A_i)$ であるものとする.

(そのような r.v. が存在することは. $Y_i(\omega) = \omega$ として,

$P_i((-\infty, x]) := P(X_i \in (-\infty, x])$ なる P_i が $\mathcal{B}(\mathbb{R})$ 上に構成できる

ことから従う.)

$(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ 上に P_i らの直積測度 $\tilde{P}$ を入れると.

$A_i \in \mathcal{B}(\mathbb{R})$ に対して.

$$\tilde{P}(Y_1 \in A_1 \cap Y_2 \in A_2 \cap \dots \cap Y_n \in A_n) = \prod_{i=1}^{n} P_i(Y_i \in A_i)$$

である. この $\tilde{P}$ を $(Y_1, \dots, Y_n)$ の同時確率とする.

すると. $(Y_1, \dots, Y_n)$ の同時分布と $(X_1, \dots, X_n)$ の同時分布は

等しくなる. なぜなら.

仮定

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)$$

$$= \prod_{i=1}^{n} F_{Y_i}(x_i) = F_{Y_1, \dots, Y_n}(x_1, \dots, x_n).$$

Y_i の構成より.

分布関数が等しければ.

対応する確率測度も一意に定まるので.

$$P((X_1, \dots, X_n) \in A) = \tilde{P}((Y_1, \dots, Y_n) \in A) \quad (\forall A \in \mathcal{B}(\mathbb{R}^n))$$

である. 特に. $A = A_1 \times \dots \times A_n$ と直積集合にすれば.

$$P(X_1 \in A_1 \cap X_2 \in A_2 \cap \dots \cap X_n \in A_n)$$

$$= P((X_1, \dots, X_n) \in A) = \tilde{P}((Y_1, \dots, Y_n) \in A)$$

$$= \prod_{i=1}^{n} P_i(Y_i \in A_i) = \prod_{i=1}^{n} P(X_i \in A_i)$$

が成り立つ. これは. $(\sigma(X_i))_{i=1}^{n}$ の独立性に他ならない.

//

Cor

r.v. $X_1, \dots, X_n$ の分布が絶対連続で
確率密度関数 (p.d.f.) $f_{X_i}(t)$ $(i=1, \dots, n)$
を持つとする. この時

$$\prod_{i=1}^{n} f_{X_i}(t_i) \text{ が } X_1, \dots, X_n \text{ の同時確率密度}$$

$$\Longleftrightarrow \quad X_1, \dots, X_n \text{ が独立.}$$

//

(証明は同様. p.d.f. と確率測度も一対一対応である)

Thm (乗法定理)

$X_1, \dots, X_n$: 可積分 かつ 独立

$$\Longrightarrow \prod_{k=1}^{n} X_k \text{ は可積分かつ } E\left[\prod_{k=1}^{n} X_k\right] = \prod_{k=1}^{n} E[X_k]$$

//

(証明は省略)

● 確率変数の収束

Def (概収束)

$(X_n)_{n=1}^{\infty}$: r.v. の列

X : r.v.

X_n が X に概収束 (convergence almost surely)

$$\overset{\text{def}}{\Longleftrightarrow} \quad P(\{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$$

$X_n \longrightarrow X$ (a.s.) や $X_n \xrightarrow{a.s.} X$ と書く.

Def (確率収束)

X_n が X に確率収束 (convergence in probability)

$$\overset{\text{def}}{\Longleftrightarrow} \quad \forall \varepsilon > 0, \ \lim_{n\to\infty} P(\{\omega \mid |X_n(\omega) - X(\omega)| > \varepsilon\}) = 0$$

$X_n \longrightarrow X$ (in prob) や $X_n \xrightarrow{p} X$ と書く.

ある正の数列 a_n に対し. $\frac{X_n}{a_n} \xrightarrow{p} 0$ なら $X_n = o_p(a_n)$ と書く.

特に $X_n \xrightarrow{p} 0$ のとき. $X_n = o_p(1)$ と書く.

(確率オーダー)

Def (tightness, 緊密性)

r.v. の列 $(X_n)_n$ がタイト(tight)であるとは、

$$\forall \varepsilon>0,\ \exists M>0 \quad \text{s.t.} \quad \sup_n P(|X_n|>M) \leqq \varepsilon$$

となる時をいう。

この時、$X_n = O_p(1)$ と書く。

Def (p次平均収束)

$p>0$ とする

$$L^p := \{ \text{r.v. } X \mid |X|^p \text{ が可積分} \}.$$

X_n が X に p次平均収束

$$\overset{\text{def}}{\Longleftrightarrow} \quad X_n, X \in L^p \text{ かつ } \lim_{n\to\infty} E[|X_n - X|^p] = 0$$

この時、$X_n \to X$ in L^p や $X_n \xrightarrow{L^p} X$ と書く。

Thm

X_n が X に概収束 $\overset{(1)}{\Longrightarrow}$ X_n が X に確率収束

X_n が X に p次平均収束 $\overset{(2)}{\Longrightarrow}$ X_n が X に確率収束

Proof

(1) $X_n \xrightarrow{a.s.} X$

$$\Longleftrightarrow P\left(\bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \left\{ |X_m(\omega) - X(\omega)| \leqq \frac{1}{k} \right\} \right) = 1 \qquad (*)$$

である。 $\left(x_n \to x \Longleftrightarrow \forall k, \exists n, \forall m \geq n,\ |x_m - x| \leq \frac{1}{k} \right)$

任意の $\varepsilon>0$ に対し、$\varepsilon > \frac{1}{k}$ となる k をとれば

$$(*)\text{左辺} \leq P\left(\bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \left\{ |X_m(\omega) - X(\omega)| \leqq \frac{1}{k} \right\} \right)$$

$$= \lim_{n\to\infty} P\left(\bigcap_{m=n}^{\infty} \left\{ |X_m(\omega) - X(\omega)| \leqq \frac{1}{k} \right\} \right)$$ (∵確率の連続性)

$$\leq \lim_{n\to\infty} P\left(|X_n - X| \leq \frac{1}{k}\right)$$

$$\leq \lim_{n\to\infty} P\left(|X_n - X| \leq \varepsilon\right) \quad \left(\because \frac{1}{k} < \varepsilon\right)$$

一方 (*)右辺=1より、$X_n \xrightarrow{P} X$ を得る。 //

(2) $\forall \varepsilon > 0$ に対し、

$$P(|X_n - X| \geq \varepsilon) = E[\mathbb{1}\{|X_n - X| \geq \varepsilon\}]$$

$$\underset{\text{チェビシェフの不等式}}{\leq} \frac{E[|X_n - X|^p]}{\varepsilon^p} \longrightarrow 0 \quad (n \to \infty)$$

$(\because X_n \xrightarrow{L^p} X)$ //

Ex1. (1) の逆は成り立たない。

$\Omega = [0,1]$, P は $[0,1]$ 上の一様分布:

$$P((a,b)) = b - a \qquad (0 \leq a < b \leq 1)$$

とする。

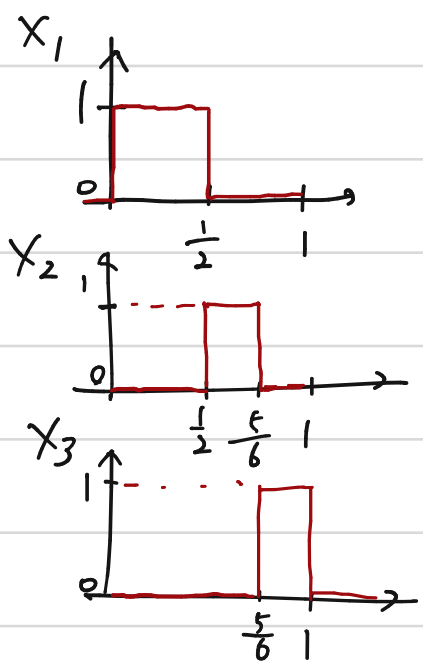
$a_n = \sum_{k=1}^{n} \frac{1}{k}$ として、

$X_n(\omega) = \mathbb{1}\{a_n - \lfloor a_n \rfloor < \omega \leq a_{n+1} - \lfloor a_{n+1} \rfloor\}$

とおく。($\lfloor a_n \rfloor$ は a_n を超えない最大の整数)

この時、$X_n \xrightarrow{P} 0$

だが、$X_n \xrightarrow{a.s.} 0$ ではない。 (チェックせよ)



//

(2) の逆も成り立たない。

$X_n = n^{\frac{1}{p}} \mathbb{1}_{[0,\frac{1}{n}]}$ とおくと、

$X_n \xrightarrow{P} 0$ だが、$E[|X_n|^p] = 1$ で、

X_n は 0 に p 次平均収束しない。

(3) $0 < p < q$ の時、

$$E[|X|^p] = E[|X|^{q \cdot \frac{p}{q}}]$$

$$\underset{\text{Jensenの不等式 } (\frac{p}{q} < 1)}{\leq} (E[|X|^q])^{\frac{p}{q}}$$

よって、$X \in L^q \Rightarrow X \in L^p$.

同様に $X \xrightarrow{L^q} 0$

$\Rightarrow X \xrightarrow{L^p} 0$ //

・大数の法則，一様大数の法則

Thm（大数の弱法則） weak laws of large numbers, w.LLN.

独立な r.v. の列 $(X_n)_{n=1}^{\infty}$ が $X_n \in L^2$ かつ

$$\mathrm{Var}[X_n] := E[(X_n - \underbrace{\mu_n}_{E[X_n]})^2] = \sigma_n^2 \text{ が } \sup_n \frac{\sum_{k=1}^{n} \sigma_k^2}{n} < \infty$$

をみたす.

$$\Longrightarrow \frac{X_1 + \cdots + X_n}{n} - \frac{\mu_1 + \cdots + \mu_n}{n} \xrightarrow{P} 0$$

//

Proof

独立性より. $i \neq j$ のとき

$$E[(X_i - \mu_i)(X_j - \mu_j)]$$
$$= E[X_i X_j] - \mu_i E[X_j] - \mu_j E[X_i] + \mu_i \mu_j$$
$$= E[X_i]E[X_j] - \mu_i \mu_j = 0.$$

よって.

$$E\left[\left|\frac{X_1 + \cdots + X_n}{n} - \frac{\mu_1 + \cdots + \mu_n}{n}\right|\right]$$
$$= \frac{1}{n^2} E\left[\left|\sum_{i=1}^{n}(X_i - \mu_i)\right|^2\right]$$
$$= \frac{1}{n^2} \sum_{i=1}^{n} E[(X_i - \mu_i)^2]$$
$$= \frac{1}{n} \cdot \underbrace{\left(\frac{\sum_{i=1}^{n} \sigma_i^2}{n}\right)}_{\sup < \infty} \longrightarrow 0 \quad (n \to \infty)$$

よって. $\frac{X_1 + \cdots + X_n}{n} - \frac{\mu_1 + \cdots + \mu_n}{n}$ は 0 に 2次平均収束するので、確率収束する

//

※ 確率収束なので"弱法則". 概収束なら"強法則"

※ 独立同一なら. $\frac{X_1 + \cdots + X_n}{n} \xrightarrow{P} \mu \quad (\mu = E[X_i])$

・一様大数の法則

・可測関数 $f: \mathbb{R} \to \mathbb{R}$

・$(X_n)_{n=1}^{\infty}$: i.i.d. (独立同一, independently identically distributed)

Xと同じ分布

s.t. $f(X_n) \in L^2$

$\Longrightarrow$ 大数の法則より,

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i) \xrightarrow{p} E[f(X)]$$

である.

しかし、機械学習や統計においては関数族 G に対し,

$$\sup_{g \in G} \left| \frac{1}{n}\sum_{i=1}^{n} g(X_i) - E[g(X)] \right| \xrightarrow{p} 0$$

を示す必要が出て来る. これを 一様大数の法則 と言う.

◎ Rademacher complexity

以下. $\|g\|_\infty \leqq M$ $(\forall g \in G)$ とする.

Def (Rademacher複雑度)

σ_i $(i=1,\dots,n)$: Rademacher変数 (i.i.d.)

$$P(\sigma_i = 1) = P(\sigma_i = -1) = \frac{1}{2}$$

とする.

絶対値付けないバージョンもある.

- $\hat{R}_n(G) := E_{(\sigma_i)_{i=1}^n}\left[\sup_{g \in G} \left| \frac{1}{n}\sum_{i=1}^{n} \sigma_i g(z_i) \right| \,\middle|\, z_1, \dots, z_n \right]$

- $R_n(G) := E_{(z_i)_{i=1}^n}\left[\hat{R}_n(G) \right]$

//

今後,

$$Pg := E[g(Z)]$$
$$P_n g := \frac{1}{n}\sum_{i=1}^{n} g(Z_i)$$

$((Z_i)_i$: i.i.d., Z と Z_i は同じ分布$)$

とする.

また、以後 $\sup_{g\in G} P_n g$ や $\sup_{g\in G} \frac{1}{n}\sum_{i=1}^n \sigma_i g(z_i)$ は可測であるとする。

(可測ではない場合、Fubiniの定理が使えない、$E_x E_{\sigma_i}[\cdot] \neq E_{\sigma_i} E_x[\cdot]$)

c.f. Suslin image admissible, Dudley: Uniform central limit theorems

van der Vaart & Wellner: Weak convergence and empirical processes (1996). Sec 2.3.

<u>Thm</u> (対称化)

$\{z_i\}_{i=1}^n$ と書くのが大変なので z_i と書く。

$$E_{z_i}\left[\sup_{g\in G} |(P-P_n)g|\right] \leq 2R_n(G)$$

//

(直接 sup を評価するより、$R_n(G)$ の方が評価しやすい)

(Proof)

z_i と同じ分布に従う r.v.

$$\text{左辺} = E_{z_i}\left[\sup_{g\in G}\left| E_{z_i'}\left[\frac{1}{n}\sum_{i=1}^n g(z_i')\right] - \frac{1}{n}\sum_{i=1}^n g(z_i)\right|\right]$$

$$\leq E_{z_i, z_i'}\left[\sup_{g\in G}\left|\frac{1}{n}\sum_{i=1}^n (g(z_i') - g(z_i))\right|\right] \quad \text{(Jensen の不等式)}$$

$$= E_{\sigma_i} E_{z_i, z_i'}\left[\sup_{g\in G}\left|\frac{1}{n}\sum_{i=1}^n \sigma_i (g(z_i') - g(z_i))\right|\right]$$

(z_i' と z_i は同じ分布)

$$\leq E_{\sigma_i} E_{z_i'}\left[\sup_{g\in G}\left|\frac{1}{n}\sum_{i=1}^n \sigma_i g(z_i')\right|\right]$$

$$+ E_{\sigma_i} E_{z_i}\left[\sup_{g\in G}\left|\frac{1}{n}\sum_{i=1}^n \sigma_i g(z_i)\right|\right]$$

$$= 2R_n(G) \quad (\because \text{Fubiniの定理})$$

//

ここで可測性使う。

可積分性は有界性より O.K.

Thm

$$P\left(\sup_{g\in G}|(P-P_n)g| \geqq 2R_n(G) + M\sqrt{\frac{2\log(\frac{1}{\delta})}{n}}\right) \leqq \delta$$

$(\forall \delta > 0)$

Proof

Lem (McDiamidの不等式)

$X_1, \dots, X_n$ を $\mathcal{X}$ 上の独立な(同一とは限らない)確率変数とする.

$f: \mathcal{X}^n \longrightarrow \mathbb{R}$ を可測関数とし,

$$|f(x_1, \dots, x_i, \dots, x_n) - f(x_1, \dots, x_i', \dots, x_n)| \leqq c_i$$

が $\forall x_1, \dots, x_n, x_i' \in \mathcal{X}$ に対して成り立つ時,

$$P\left(f(X_1, \dots, X_n) - E[f(X_1, \dots, X_n)] \geqq \varepsilon\right) \leqq \exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^n c_i^2}\right) \quad (\forall \varepsilon > 0)$$

//

今, $\|g\|_\infty \leqq M$ $(\forall g \in G)$ より

$$f(z_1, \dots, z_n) = \sup_{g\in G}\left|\frac{1}{n}\sum_{i=1}^n g(z_i) - Pg\right|$$

とおくと.

$$f(z_1, \dots, z_n) - f(z_1, \dots, z_i', \dots, z_n)$$

$$= \sup_{g\in G}\left|\frac{1}{n}\sum_{j\neq i} g(z_j) + \frac{g(z_i)}{n} - \frac{g(z_i')}{n} + \frac{g(z_i')}{n} - Pg\right|$$

$$- \sup_{g\in G}\left|\frac{1}{n}\sum_{j\neq i} g(z_j) + \frac{g(z_i')}{n} - Pg\right|$$

$$\leqq \sup_{g\in G}\left|\underbrace{\frac{1}{n}\sum_{j\neq i} g(z_j) + \frac{g(z_i')}{n} - Pg}_{\text{//}}\right| + \sup_{g\in G}\left|\frac{g(z_i)}{n} - \frac{g(z_i')}{n}\right|$$

$$- \sup_{g\in G}\left|\quad\right|$$

$$\leqq \frac{2M}{n}$$

同様にして. $f(z_1, \dots, z_i', \dots, z_n) - f(z_1, \dots, z_n) \leqq \frac{2M}{n}$ より

絶対値 $\leqq \frac{2M}{n}$ である

あとは McDiamid と対称化より従う.

//

Prop (Rademacher複雑度の性質)

(1) $G \subset G'$ なら $R_n(G) \leq R_n(G')$

(2) $\phi_i : \mathbb{R} \to \mathbb{R}$ が $|\phi_i(x) - \phi_i(y)| \leq D|x-y|$ なら

$$E_{\sigma_i}\left[\sup_{g \in G} \left|\frac{1}{n}\sum_{i=1}^{n} \sigma_i \phi_i(g(z_i))\right|\right] \leq 2D\hat{R}_n(G)$$

(contraction inequality)

($R_n(G)$の定義で絶対値を使わない場合、左辺の$E[\cdots]$の中が絶対値をみすごせば $2D$ と D と 2通り.)

(3) Gの凸包を

$$\mathrm{conv}(G) = \left\{\sum_{i=1}^{m} \lambda_i g_i \;\middle|\; \sum_{i=1}^{m} \lambda_i = 1, \lambda_i \geq 0, g_i \in G, m = 1, 2, \ldots\right\}$$

とすると

$$R_n(G) = R_n(\mathrm{conv}(G))$$

(4) $R_n(c \cdot G) = |c| R_n(G) \quad (c \in \mathbb{R})$

(2) $\mathcal{F}$が有限集合のとき

$\mathcal{F} = \{f_1, \ldots, f_N\}$, $\|f_j\|_\infty \leq R$ なら

$$R_n(\mathcal{F}) \leq R\sqrt{\frac{2\log(2N)}{n}}$$

← R_nに$|\cdot|$が入ってなければ$|\cdot|$をみすごせる.

(Massartの定理)

← Hoeffdingの不等式より示せる.

(証明は略)

※ 個数Nに対し、$O(\sqrt{\log(N)})$しか影響がない.

○ Rademacher 複雑度の上界：カバリングナンバー，エントロピーナンバー

Def (Covering Number, 被覆数)

$\mathcal{F}$: 関数の集合

d: $\mathcal{F}$ の距離

$$N(\mathcal{F}, \varepsilon, d) := \min\{n \geq 1 \mid \exists f_1, \dots, f_n \in \mathcal{F} \text{ s.t. } \forall f \in \mathcal{F} \text{ に対して } \exists f_j \text{ s.t. } d(f, f_j) \leq \varepsilon\}$$

//

Def (エントロピー数)

$$e_i(\mathcal{F}, d) := \inf\{\varepsilon > 0 \mid N(\mathcal{F}, \varepsilon, d) \leq 2^{i-1}\}$$

データ点 $x_1, \dots, x_n \in \mathcal{X}$ に対して，

$$\|f\|_{n,p} := \left(\frac{1}{n}\sum_{i=1}^{n} |f(x_i)|^p\right)^{\frac{1}{p}} \quad (1 \leq p < \infty)$$

とする．

thm $\forall f \in \mathcal{F}$ が $\|f\|_\infty \leq R$ をみたすなら，あるデータ $x_1, \dots, x_n \in \mathbb{R}^d$ が与えられたもとで

$$\hat{R}_n(\mathcal{F}) \leq \inf_{\alpha \geq 0}\left\{\alpha + R\sqrt{\frac{2\log(2N(\alpha, \mathcal{F}, \|\cdot\|_{n,1}))}{n}}\right\}$$

Proof $f \in \mathcal{F}$ に対して，$v[f]$ を f に一番近い α-被覆の点とする．

$$\hat{R}_n(\mathcal{F}) = E_\sigma\left[\sup_{f \in \mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i f(x_i)\right| \,\middle|\, x_1, \dots, x_n\right]$$

以下略

$$= E_\sigma\left[\sup_{f \in \mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i\left(f(x_i) - v[f](x_i)\right) + \frac{1}{n}\sum_{i=1}^{n}\sigma_i v[f](x_i)\right|\right]$$

$$\leq E_\sigma\left[\sup_{f \in \mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i\left(f(x_i) - v[f](x_i)\right)\right|\right] + E_\sigma\left[\sup_{f \in \mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i v[f](x_i)\right|\right]$$

$$\leq \sup_{f \in \mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}|f(x_i) - v[f](x_i)| + E_\sigma\left[\max_{1 \leq j \leq N(\mathcal{F}, \alpha, \|\cdot\|_{n,1})}\left|\frac{1}{n}\sum_{i=1}^{n}\sigma_i f_j(x_i)\right|\right]$$

α-被覆の j 番目の元

$$\leq \alpha + R\sqrt{\frac{2\log(2N)}{n}} \quad (\because \text{Massart の定理})$$

//

・$p \leq q$ なら $\|\cdot\|_{n,p} \leq \|\cdot\|_{n,q}$ であることから
$N(\alpha, \mathcal{F}, \|\cdot\|_{n,p}) \leq N(\alpha, \mathcal{F}, \|\cdot\|_{n,q})$ であるので、左辺の被覆数は
$N(\alpha, \mathcal{F}, \|\cdot\|_{n,p})$ $(p>1)$ でおさえて良い。 ($p=2$ は上で用いられる)

Thm (Dudley のチェイニング)
固定 $x_1, \dots, x_n \in \mathbb{R}^d$ が与えられたとき、$\hat{\delta} := \sup_{f \in \mathcal{F}} \|f\|_{n,2}$ とおくと

$$\hat{R}_n(\mathcal{F}) \leq \inf_{\alpha > 0} \left\{ 4\alpha + \frac{12}{\sqrt{n}} \int_\alpha^{\hat{\delta}} \sqrt{\log 2N(\mathcal{F}, \varepsilon, \|\cdot\|_{n,2})}\, d\varepsilon \right\}$$

$$\hat{R}_n(\mathcal{F}) \leq \frac{2}{\sqrt{n}} \left(\sum_{i=1}^{\infty} 2^{i/2} e_{2^i}(\mathcal{F} \cup \{0\}, \|\cdot\|_{n,2}) + \sup_{f \in \mathcal{F}} \|f\|_{n,2} \right)$$

※ $\mathcal{F}$ を有限個の代表点で近似し、その近似精度を粗いものから密なものまでつなげてゆくという操作に対応。

Proof (エントロピー数の方だけ与えておく)

$$T = \{ (f(x_1), \dots, f(x_n)) \mid f \in \mathcal{F} \} \cup \{(0, \dots, 0)\} \subseteq \mathbb{R}^n$$

$d(t, t') := \sqrt{\frac{1}{n} \sum_{i=1}^{n} (t_i - t_i')^2}$ とおくと、(T, d) は可分な距離空間。

以後、$\forall i$ で $e_i(T, d)$ は有限であるとして一般性を失わない。
特に、T は有界で、$i \to \infty$ で $e_i(T, d) \to 0$ である。

任意の $\varepsilon > 0$ に対し

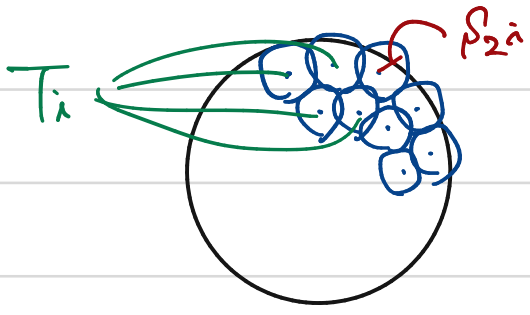
$$\rho_1 := \sup_{t \in T} d(t, 0)$$

$$\rho_{2^i} := (1+\varepsilon)\, e_{2^i}(T, d) \qquad (i = 1, 2, \dots)$$

とする。また、$\forall i \geq 1$ に対し、ρ_{2^i}-ネット T_i を $|T_i| \leq 2^{2^i - 1}$ なるものとして取ってくる

$\boxed{\bigcup_{t \in T_i} B_d(t, \rho_{2^i}) \supseteq T \text{ なるもの}}$

(エントロピー数の定義より可能)



$T_0=\{0\}$ とする。すると、$e_j(T,d)\to 0$ $(j\to\infty)$ より、$\mathcal{U}:=\bigcup_{j=0}^{\infty} T_j$ に対し、

$$\sup_{t\in\mathcal{U}} \frac{1}{n}\sum_{i=1}^{n}\sigma_i t_i = \sup_{t\in T} \frac{1}{n}\sum_{i=1}^{n}\sigma_i t_i$$

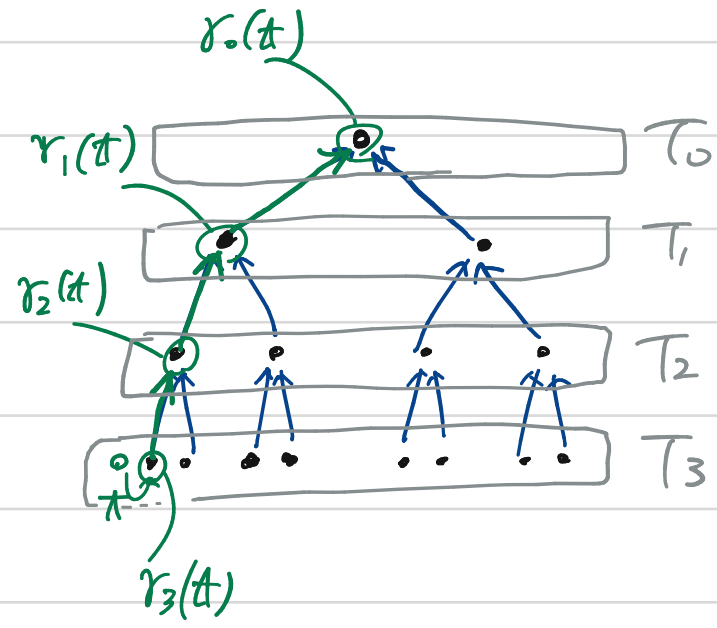
が成り立つ。（f は T 内で稠密）よって、T の代わりに $\mathcal{U}$ を考えれば良い。

$j\geq 1$、$t\in T$ に対し、$\pi_j(t)\in T_j$ を、$d(t,\pi_j(t))\leq \delta_{2^j}$ なるものとして取ってくる
特に、$t\in T_j$ なら、$\pi_j(t)=t$ としておく。

$j\geq 1$ を1つ固定して、γ_i $(i\leq j)$ を

$$\gamma_j=\pi_j$$
$$\gamma_{i-1}=\pi_{i-1}\circ\gamma_i \quad (i=1,\dots,j)$$

とする。



当然 $d(\gamma_i(t),\gamma_{i-1}(t))\leq \delta_{2^{i-1}}$ である。

今、$h(t):=\frac{1}{n}\sum_{i=1}^{n}\sigma_i t_i$ $(t\in T)$

とおくと、$t\in T_j$ について

$$|h(t)|=|h(t)-0|$$
$$=\left|\sum_{i=1}^{j}\left(h(\gamma_i(t))-h(\gamma_{i-1}(t))\right)\right|\leq \sum_{i=1}^{j}\max_{\tilde{t}\in T_i}\left|h(\tilde{t})-h(\pi_{i-1}(\tilde{t}))\right|$$

となる。よって、$\mathcal{U}_k:=T_0\cup\cdots\cup T_k$ に対しても

$$\max_{t\in\mathcal{U}_k}|h(t)|\leq \sum_{j=1}^{k}\max_{t\in T_j}\left|h(t)-h(\pi_{j-1}(t))\right|$$

が成り立つ。すると、有限集合の Rademacher 複雑度より

$$E_{\sigma_i}\left[\max_{t\in\mathcal{U}_k}|h(t)|\right]\leq \sum_{j=1}^{k}E_{\sigma_i}\left[\max_{t\in T_j}\left|h(t)-h(\pi_{j-1}(t))\right|\right]$$
$$\leq \sum_{j=1}^{k}\sqrt{\frac{2\log(2\cdot 2^{2^j}-1)}{n}}\cdot\delta_{2^{j-1}}$$

$\mathcal{U}=\bigcup_{j=0}^{\infty}T_j$ については、$k\to\infty$ とすることで、単調増大性より

$$E_{\sigma_i}\left[\sup_{t\in\mathcal{U}}h(t)\right]\leq \sum_{j=0}^{\infty}\sqrt{\frac{4\cdot 2^j}{n}}\,\delta_{2^j}.$$

あとは $\delta_1=\sup_t d(t,0)$ に注意すれば、証明完了。

● 大数の強法則

$A_1, A_2, \cdots$ を事象の列とする.

$$\{A_n \text{ i.o.}\} := \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n = \{\omega \in \Omega \mid {}^{\forall}k \geqq 1, {}^{\exists}n \geqq k,\ \omega \in A_n\}$$

infinitely often の意味

どんな大きな k に対しても, ${}^{\exists}n \geqq k$ で $\omega \in A_n$ なので, $\omega \in A_n$ なる n は無限個ある つまり i.o. である.

$\{A_n \text{ i.o.}\}$ は「A_n が無限回起きる」という事象.

Thm (Borel-Cantelliの補題)

1. $\sum_{n=1}^{\infty} P(A_n) < \infty \Rightarrow P(A_n \text{ i.o.}) = 0$
2. $(A_n)_{n=1}^{\infty}$ が独立なら
$$\sum_{n=1}^{\infty} P(A_n) = \infty \Rightarrow P(A_n \text{ i.o.}) = 1$$

//

※ 2 において, 独立性は必要.

今, コイン投げを考え,

$$P(\omega_1 = \omega_2 = \cdots = 1) = P(\omega_1 = \omega_2 = \cdots = 0) = \frac{1}{2}$$

とする. $A_n = \{\omega_n = 1\}$ とすると, $P(A_n) = \frac{1}{2}$ である.

よって, $\sum_{n=1}^{\infty} P(A_n) = \infty$ であるが $P(A_n \text{ i.o.}) = \frac{1}{2}$ である.

Proof

1. $$P(A_n \text{ i.o.}) = P\left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n\right) \leqq P\left(\bigcup_{n=k'}^{\infty} A_n\right) \quad (k' \text{は任意})$$
$$\leqq \sum_{n=k'}^{\infty} P(A_n) \longrightarrow 0 \quad (k' \to \infty)$$

2. $$P(\{A_n \text{ i.o.}\}^c) = P\left(\bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n^c\right) \leqq \sum_{k=1}^{\infty} P\left(\bigcap_{n=k}^{\infty} A_n^c\right)$$

(やや informal)

$$= \sum_{k=1}^{\infty} \prod_{n=k}^{\infty} P(A_n^c) \quad (\text{独立性と確率の連続性})$$
$$= \sum_{k=1}^{\infty} \prod_{n=k}^{\infty} (1 - P(A_n)) \quad (1 - x \leqq e^{-x})$$
$$\leqq \sum_{k=1}^{\infty} \exp\Big(-\underbrace{\sum_{n=k}^{\infty} P(A_n)}_{\to \infty}\Big) = 0$$

//

Thm (L^4を仮定した大数の強法則)

$(X_n)_{n=1}^{\infty}$: i.i.d. , $X_1 \in L^4$

$$\Rightarrow \frac{X_1 + \cdots + X_n}{n} \xrightarrow{a.s.} E[X_1]$$

//

Proof

$Y_n = X_n - E[X_n]$ とすると. $E[Y_n] = 0$ とできるので. $E[X_n] = 0$ として一般性を失わない.

$$E\left[\left(\sum_{i=1}^{n} X_i\right)^4\right] = n E[X_1^4] + \binom{n}{2}\binom{4}{2} E[X_1^2]^2$$

(異なる i, j に対し. $E[X_i X_j X_k^2] = E[X_i] E[X_j] E[X_k^2] = 0$ 等を用いる)

よって. $E\left[\left(\sum_{i=1}^{n} X_i\right)^4\right] \leqq c\, n^2$ (c: 定数)

チェビシェフの不等式より

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n}\right| \geqq \varepsilon\right) \leqq \frac{E\left[\left|\frac{X_1 + \cdots + X_n}{n}\right|^4\right]}{\varepsilon^4}$$

$$\leqq \frac{c\, n^2}{n^4 \varepsilon^4} = \frac{c}{\varepsilon^4} \cdot \frac{1}{n^2}$$

よって.

$$\sum_{n=1}^{\infty} P\Bigg(\overbrace{\left|\frac{X_1 + \cdots + X_n}{n}\right| \geqq \varepsilon}^{A_n}\Bigg) < \infty$$

となるので. Borel-Cantelliの補題より

$$P\left(\limsup_{n \to \infty} \left|\frac{X_1 + \cdots + X_n}{n}\right| \geqq \varepsilon\right) = 0$$

$\varepsilon \searrow 0$ とすることで. 確率の連続性より

$$P\left(\limsup \left|\frac{X_1 + \cdots + X_n}{n}\right| > 0\right) = 0$$

//

Thm (大数の強法則)

$(X_n)_{n=1}^{\infty}$: i.i.d. , $X_i \in L^1$

$$\Rightarrow \frac{X_1 + \cdots + X_n}{n} \xrightarrow{a.s.} E[X_1]$$

/

※ 分散の存在も L^4 も不要. 証明は補足資料 or Thm 2.4.1 in Durret 5th ed.