統計数学特別講義第二 (4/19)

第0章 なぜ経験過程理論を学ぶか

標準正規分布 $N(0,1)$ から発生させた 10個の乱数を考える

$x_1, x_2, \ldots, x_n$ (一般性を失わず $x_1 < x_2 < \cdots < x_{10}$ とする)

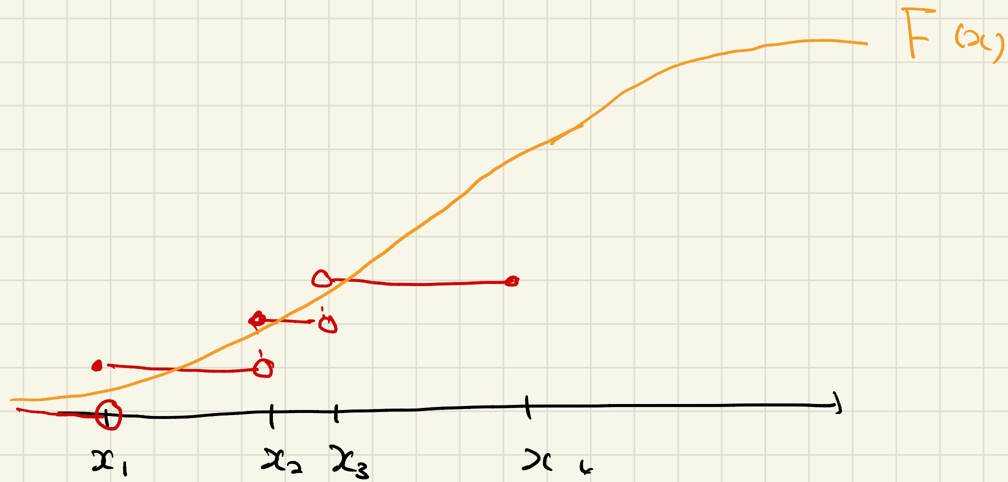
経験分布関数

$$\hat{F}_{10}(x) = \frac{1}{10} \# \{ x_j \leq x \,;\, j = 1, \ldots, 10 \} \quad (x \in \mathbb{R})$$

$N(0,1)$ の分布関数

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \quad (x \in \mathbb{R})$$

2



もう少し一般的に確率変数の言葉を用いる.

$$X_1, X_2, \ldots\ldots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu, 1), \quad \mu \in \mathbb{R}$$

すると $N(\mu, 1)$ の p.d.f. は

$$p_\mu(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2}\right\} \quad (x \in \mathbb{R}).$$

$X_1, \ldots, X_n$ に基づく経験分布関数は

$$\hat{F}_n(x) = \frac{1}{n} \#\{X_j \leq x ; j = 1, 2, \ldots, n\} \quad (x \in \mathbb{R})$$

ランダムな量

$\{X_j\}_{j=1}^n$ の定義されている確率空間を $(\Omega, \mathcal{F}, P_r)$ と表す.

$X_j : \Omega \ni \omega \longmapsto X_j(\omega) \in \mathbb{R}$ は可測であることに注意する.

$$\Omega \ni \omega \longmapsto \hat{F}_n(x)(\omega) = \frac{1}{n} \# \{X_j(\omega) \leq x ; j = 1, 2, \ldots, n\}$$

ランダムな関数.

具体的な関数 2

$\hat{F}_n(x)$ は統計量（推定量）

統計量 $\hat{F}_n(x)$ の性質

$$\overline{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$$

まず，標本平均 $\overline{X}_n$ をみてみる．

- $E[\overline{X}_n] = \mu$．
- $\forall \varepsilon > 0$ に対して

$$\lim_{n\to\infty} P_r(|\overline{X}_n - \mu| > \varepsilon) = 0.$$

となる．

同様に，固定した $x \in \mathbb{R}$ に対して

- $E[\hat{F}_n(x)] = F(x)$.

$$\mathbb{I}_{(-\infty, x]}(t) = \begin{cases} 1 & (t \leq x) \\ 0 & (t > x) \end{cases}$$

- $\forall \varepsilon > 0$ に対して

$$\mathbb{I}_A(t) = \begin{cases} 1 & (t \in A) \\ 0 & (t \notin A) \end{cases}$$

$$\lim_{n \to \infty} \Pr(|\hat{F}_n(x) - F(x)| > \varepsilon) = 0$$

$\bar{X}_n \longleftrightarrow \hat{F}_n(x)$

$X_j \longleftrightarrow \mathbb{I}_{(-\infty, x]}(X_j)$

$E[X_j] = \mu \longleftrightarrow E[\mathbb{I}_{(-\infty, x]}(X_j)] = F(x)$

となる．実際

$$\hat{F}_n(x) = \frac{1}{n} \sum_{j=1}^{n} \mathbb{I}_{(-\infty, x]}(X_j)$$

と

$E[\mathbb{I}_{(-\infty, x]}(X_j)] = \Pr(X_j \leq x) = F(x)$ が成り立つ．

さらに、

$$\lim_{n\to\infty} \Pr\left(\sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F(x)| > \varepsilon \right) = 0$$

となることも知られている。 Glivenko-Cantelli

いま

$$P(A) := \int_A p_u(x)\,dx \qquad (A \in \mathcal{B}(\mathbb{R}))$$

$$\hat{P}_n(A) := \frac{1}{n} \# \{ x_j \in A \,;\, j = 1, \ldots, n \}$$

とおく。

$A = (-\infty, x]$ のとき

$$P(A) = F(x); \quad \hat{P}_n(A) = \hat{F}_n(x)$$

である。

固定した $A \in B(\mathbb{R})$ に対して

$$\hat{P}_n(A) = P(A); \quad \lim_{n\to\infty} \Pr(|\hat{P}_n(A) - P(A)| > \varepsilon) = 0$$

しかし

$$\sup_{A\in B(\mathbb{R})} |\hat{P}_n(A) - P(A)| = 1.$$

経験過程

関数 $g: \mathbb{R} \longrightarrow \mathbb{R}$ に対し

$$Pg = \int_{\mathbb{R}} g(x)\, dP(x), \quad \hat{P}_n g = \frac{1}{n} \sum_{j=1}^{n} g(X_j)$$

$$\nu_n(g) = \sqrt{n}\,(\hat{P}_n g - Pg) = \sqrt{n}\,(\hat{P}_n - P)(g)$$

とおく.

$$\mathcal{G}_1 := \{ \mathbb{I}_{(-\infty, x]}(t) \,;\, x \in \mathbb{R} \}$$

$$\mathcal{G}_2 := \{ \mathbb{I}_A(t) \,;\, A \in \mathcal{B}(\mathbb{R}) \}$$

とおくと

$$\{\nu_n(g) : g \in \mathcal{G}_1\} = \{\sqrt{n}(\hat{F}_n(x) - F(x)) : x \in \mathbb{R}\}$$

$$\{\Phi_n(g) : g \in \mathcal{G}_2\} = \{\sqrt{n}(\hat{P}_n(x) - P(x)) : A \in \mathcal{B}(\mathbb{R})\}$$

とする.

ある関数族 $\mathcal{G}$ に対して

$$\{\sqrt{n}(\hat{P}_n - P)g : g \in \mathcal{G}\}$$

を考えることにする.

統計的機械学習,
推測理論の注目點

- 族の確率の評価, 極限定理 (大数の法則,
中心極限定理)

第1章 準備：確率空間・事象・確率変数・期待値

1.1 記号

$$\mathbb{I}_A(x) = \begin{cases} 1 & (x \in A) \\ 0 & (x \notin A) \end{cases}$$

特に，$\mathbb{I}_{(-\infty, a]}(x)$ を $\mathbb{I}\{x \leq a\}$ と書くこともある

$$\mathbb{I}\{\cdots\} = \begin{cases} 1 & (\cdots \text{が真}) \\ 0 & (\cdots \text{が偽}) \end{cases}$$

1.2 確率と確率変数

定義 1.1 $\Omega(\neq \phi)$ を集合とし、$\mathcal{A}$ を Ω の部分集合族とする。部分集合族 $\mathcal{A}$ が次の3条件をみたすとき、$\mathcal{A}$ は σ 加法族とよばれる。

$$A^c := \{\omega \in \Omega ; \omega \notin A\}$$

(1) $\Omega \in \mathcal{A}$.

(2) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$.

(3) $A_n \in \mathcal{A} (n \in \mathbb{N}) \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

言葉 $(\Omega, \mathcal{A})$ を可測空間という

記号　$\mathcal{C}$ を Ω のある部分集合族（σ 加法族でなくてもよい）

とする．このとき

$$\sigma[\mathcal{C}] := \bigcap \{ \mathcal{A} ; \mathcal{A} \supset \mathcal{C},\ \mathcal{A} \text{ は } \sigma \text{ 加法族} \}$$

と定める．すると $\sigma[\mathcal{C}]$ は $\mathcal{C}$ を含む最小の σ 加法族と

なることがわかる．すなわち，$\mathcal{G}$ を $\mathcal{C}$ を含む任意の

σ 加法族とすると

$$\sigma[\mathcal{C}] \subset \mathcal{G}$$

となる．

定義 1.3 $\Omega = \mathbb{R}$ とする.

$$\mathcal{O} := \{ O \subset \mathbb{R} ; O \text{ は } \mathbb{R} \text{ の開集合} \}$$

このとき, $\sigma[\mathcal{O}]$ を Borel 集合族 という. 以後, $\mathcal{B}(\mathbb{R})$ と記す.

$\mathcal{B}(\mathbb{R}) \subsetneq 2^{\mathbb{R}}$ となることも知られている.

また

$$\mathcal{C} = \{ (-\infty, x] ; x \in \mathbb{R} \}$$

とすると

$$\sigma[\mathcal{C}] = \sigma[\mathcal{O}]$$

となることがわかる.

$\sigma[\mathcal{C}] \subset \sigma[\mathcal{O}]$ は自明です.
逆は Dynkin の定理と $\mathbb{R}$ の位相を用いて証明することができる.

定義 1.4 $(\Omega, \mathcal{A})$を可測空間とする. $\mathcal{A}$上の関数

$$\mu : \mathcal{A} \ni A \mapsto \mu(A) \in [0, \infty) \cup \{\infty\}$$

が次の2条件をみたすとき, 可測空間$(\Omega, \mathcal{A})$上の測度

とよばれる.

(1) $\mu(\phi) = 0$.

互いに排反

(2) $A_n \in \mathcal{A}\ (n \in \mathbb{N})$ かつ $A_n \cap A_m = \phi\ (n \neq m)$ に対して

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

用語　$(\Omega, \mathcal{A}, \mu)$ を測度空間という

特に，$\mu(\Omega) = 1$ のとき，$(\Omega, \mathcal{A}, \mu)$ を確率空間という

以後，確率空間の測度を $\Pr(\cdot)$ と記すことに

する．

$(\Omega, \mathcal{A}, \Pr)$：確率空間．

補題1.5 $(\Omega, \mathcal{A}, \Pr)$を確率空間とする.

このとき, 以下が成立.

(1) $\Pr(\phi) = 0$.

(2) $\Pr(A^c) = 1 - \Pr(A) \quad (\forall A \in \mathcal{A})$.

(3) $N \in \mathbb{N}$とする. $\{A_n\}_{n=1}^{N} \subset \mathcal{A}$が互いに排反ならば

$$\Pr\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{n=1}^{N} \Pr(A_n).$$ ← 有限加法性.

(4) $A, B \in \mathcal{A}, \ A \subset B \Rightarrow \Pr(B \setminus A) = \Pr(B) - \Pr(A)$.

よって. $A \subset B \Rightarrow \Pr(A) \leq \Pr(B)$.

ただし、$B \setminus A = \{\omega \in \Omega ;\ \omega \in B$ かつ $\omega \in A^c\}$

(5) $A_n \in \mathcal{A}\ (n \in \mathbb{N})$ かつ $A_n \subset A_{n+1}$ のとき

$$\lim_{n\to\infty} \Pr(A_n) = \Pr\left(\bigcup_{n=1}^{\infty} A_n\right).$$

(6) $A_n \in \mathcal{A}\ (n \in \mathbb{N})$ かつ $A_n \supset A_{n+1}$ のとき

$$\lim_{n\to\infty} \Pr(A_n) = \Pr\left(\bigcap_{n=1}^{\infty} A_n\right)$$

登録することもある?

(7) Bounded Union $\quad \Pr\left(\bigcup_{n=1}^{\infty} A_n\right) \le \sum_{n=1}^{\infty} \Pr(A_n).$

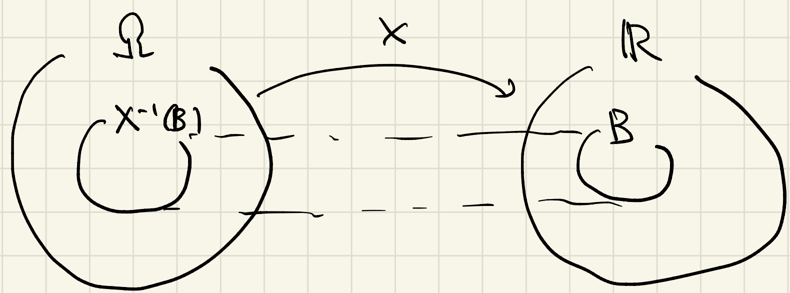
定義1.6 $(\Omega, \mathcal{A})$, $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ を可測空間とする.

写像 $X: \Omega \to \mathbb{R}$ は $(\Omega, \mathcal{A})$ から $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ への可測写像

であるとは

$$X^{-1}(B) := \{\omega \in \Omega ; X(\omega) \in B\} \in \mathcal{A} \quad ({}^{\forall} B \in \mathcal{B}(\mathbb{R}))$$

をみたすことをいう.



定義 1.7 (1) $(\Omega, \mathcal{A}, \Pr)$ を確率空間とする.

$X : \Omega \to \mathbb{R}$ が可測のとき, X を確率変数という

(2) $d \geq 2$, $d \in \mathbb{N}$ とする. $X : \Omega \to \mathbb{R}^d$ は可測のとき,

X を確率ベクトルという

定理1.8 $(\Omega, \mathcal{A})$を可測空間とする。$X : \Omega \to \mathbb{R}$を写像とし、

$\mathcal{C}$を$\mathbb{R}$の集合族（σ加法族でなくてもよい）とする。

${}^\forall C \in \mathcal{C}$に対して

$$\{\omega \in \Omega ; X(\omega) \in C\} \in \mathcal{A}$$

かつ

$$B(\mathbb{R}) = \sigma[\mathcal{C}]$$

のとき、Xは可測となる。

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

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

と記す. $\{B_n\}_{n=1}^{\infty} \subset \mathcal{B}(\mathbb{R})$ に対して

$$\{X \in \bigcup_{n=1}^{\infty} B_n\} = \bigcup_{n=1}^{\infty} \{X \in B_n\}$$

$$\{X \in B^c\} = \{X \in B\}^c$$

となる. したがって,

$$\mathcal{D} := \{B \in \mathcal{B}(\mathbb{R}) ; \{X \in B\} \in \mathcal{A}\}$$

は σ加法族 (必要ならば $\mathbb{R}$ も入っていること) となる.

実際,

(1) $\mathbb{R} \in \mathcal{D}$

(2) $B \in \mathcal{D} \Rightarrow \{X \in B\} \in \mathcal{A} \Rightarrow \{X \in B^c\} = \{X \in B\}^c \in \mathcal{A} \Rightarrow B^c \in \mathcal{D}$

(3) $B_n \in \mathcal{D}\ (n=1,2,\ldots) \Rightarrow \{X \in B_n\} \in \mathcal{A}\ (n=1,2,\ldots)$

$\Rightarrow \{X \in \bigcup_{n=1}^{\infty} B_n\} = \bigcup_{n=1}^{\infty} \{X \in B_n\} \in \mathcal{A} \Rightarrow \bigcup_{n=1}^{\infty} B_n \in \mathcal{D}$.

よって, $\mathcal{E} \subset \mathcal{D}$ となる. $\sigma[\mathcal{E}]$ の最小性より

$$\sigma[\mathcal{E}] \subset \mathcal{D}$$

となるので, X は可測. $\square$

注意 1.9 (1) $X: \Omega \to \mathbb{R}$ とする. $\mathcal{C} = \{(-\infty, r] \,;\, r \in \mathbb{R}\}$ としたとき, $\mathcal{B}(\mathbb{R}) = \sigma[\mathcal{C}]$ となることがわかる. ($\because$ Dynkinの補題)

よって, $\forall r \in \mathbb{R}$ において, $\{X \leq r\} \in \mathcal{A}$ ならば, X は可測.

$\therefore$ $\{X \leq r\} \in \mathcal{A}$ $(\forall r \in \mathbb{R})$ $\Leftrightarrow$ X は確率変数

注意 1.10 (1) $n \in \mathbb{N}$ とし、$X_1, X_2, \dots, X_n$ は確率変数とし、

$f: \mathbb{R}^n \to \mathbb{R}$ は可測のとき、$f(X_1, X_2, \dots, X_n)$ は確率変数となる。

(2) $\{X_n\}_{n=1}^{\infty}$ は確率変数列とする。

$$\inf_n X_n, \quad \sup_n X_n, \quad \limsup_n X_n, \quad \liminf X_n$$

も確率変数

記法

$$(X_1 + X_2)(\omega) := X_1(\omega) + X_2(\omega) \qquad ({}^\forall \omega \in \Omega)$$

によって、$X_1 + X_2$ と定めている。

同様に

$$\inf_n X_n(\omega) = \inf_n X_n(\omega) \qquad (\forall \omega \in \Omega)$$

によって、$\inf_n X_n$ を定めている。

$\{X_n(\omega)\}_{n=1}^{\infty}$ は下に有界 $\Leftrightarrow$ $\exists a \in \mathbb{R}$ s.t. $a \leq X_n(\omega)$ $(n=1,2,\ldots)$

このような a を $\{X_n(\omega)\}_{n=1}^{\infty}$ の下界という

a は $\{X_n(\omega)\}_{n=1}^{\infty}$ の下限 $\Leftrightarrow$ (1) a は $\{X_n(\omega)\}_{n=1}^{\infty}$ の下界

(2) $\forall \varepsilon > 0$ に対しても、$\exists n_0 \in \mathbb{N}$

s.t. $X_{n_0}(\omega) < a + \varepsilon$

a を $\inf X_n(\omega)$ とかく。

$$\liminf_{n\to\infty} x_n = \lim_{k\to\infty} \inf_{n\geq k} x_n$$

$$\limsup_{n\to\infty} x_n = \lim_{k\to\infty} \sup_{n\geq k} x_n$$

と定めている。

$$\liminf_{n\to\infty} x_n \leq \limsup_{n\to\infty} x_n$$

である。

$\liminf\limits_{n\to\infty} x_n = \limsup\limits_{n\to\infty} x_n$ のとき、これらを $\lim x_n$ と記す。

1.3 確率的独立性

定義 1.12 $N \geq 3$ とする. 事象列 $\{A_n\}_{n=1}^{N}$ が独立であるとは

$2 \leq {}^{\forall}\ell \leq N$ と任意の $1 \leq k_1 < k_2 < \cdots < k_\ell \leq N$ に対して

$$\Pr\left(\bigcap_{j=1}^{\ell} A_{k_j}\right) = \prod_{j=1}^{\ell} \Pr(A_{k_j}) \qquad (1.1)$$

が成り立つこと

例 $\{A_n\}_{n=1}^{3}$ が独立 $\iff \Pr(A_1 \cap A_2 \cap A_3) = \Pr(A_1)\Pr(A_2)\Pr(A_3)$

かつ

$\Pr(A_i \cap A_j) = \Pr(A_i)P(A_j) \quad (i \neq j)$

定義 1.13 $\{A_\lambda\}_{\lambda \in \Lambda}$ $(A_\lambda \in \mathcal{A})$ は独立

$\Longleftrightarrow$ Λ の任意の有限部分集合 $\{\lambda_1, \dots, \lambda_\ell\} \subset \Lambda$ に対して

(1.1) が成立.

定義 1.14 $\mathcal{A}_1, \mathcal{A}_2$ を $\mathcal{A}$ の部分 σ 加法族とする.

$\mathcal{A}_1$ と $\mathcal{A}_2$ が独立 $\Longleftrightarrow$ $P_r(A_1 \cap A_2) = P_r(A_1) P_r(A_2)$

$(^\forall A_1 \in \mathcal{A}_1, {}^\forall A_2 \in \mathcal{A}_2)$.

定義 1.15 (1) $\mathcal{A}$ の部分σ加法族 $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_N$ は

独立 $\Leftrightarrow$

$$\Pr\left(\bigcap_{j=1}^{N} A_j\right) = \prod_{j=1}^{N} \Pr(A_j) \quad ({}^\forall A_j \in \mathcal{A}_j).$$

(2) $\mathcal{A}$ の部分σ加法族 $\{\mathcal{A}_\lambda\}_{\lambda\in\Lambda}$ は独立

$\Leftrightarrow$ 任意の有限部分集合 $\{\lambda_1, \ldots, \lambda_\ell\} \subset \Lambda$ に

対して，$\{\mathcal{A}_{\lambda_j}\}_{j=1}^{\ell}$ が独立

(3) $\{A_\lambda\}_{\lambda\in\Lambda}$ は対ごとに独立 $\Leftrightarrow$ ${}^\forall\{j, \ell\} \subset \Lambda$ ($j \neq \ell$) に対して

A_j と A_ℓ が独立.

定義1.16 (1) 確率変数列 $\{X_j\}_{j=1}^N$ は独立

$\Leftrightarrow$ ${}^\forall B_k \in \mathcal{B}(\mathbb{R})$ $(k=1, \dots, N)$ に対して

$$\Pr\left(\bigcap_{k=1}^{N} \{X_k \in B_k\}\right) = \prod_{k=1}^{N} \Pr(X_k \in B_k). \quad (a)$$

(2) 確率変数列 $\{X_\lambda\}_{\lambda \in \Lambda}$ は独立

$\Leftrightarrow$ Λ の任意の有限部分集合 $\{k_1, k_2, \dots, k_\ell\}$

に対して，$\{A_{k_j}\}_{j=1}^{\ell}$ は (a) をみたす．

1.4 期待値の定義

期待値は以下の①〜③の段階で定義とする

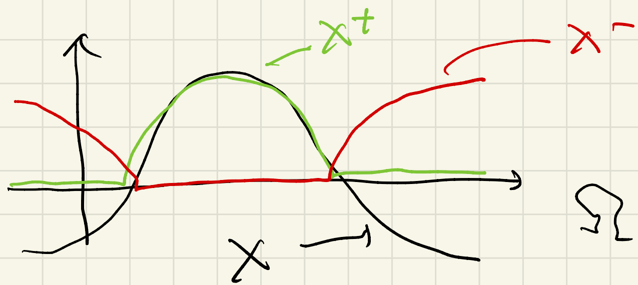
⇒ 標準機械 (Standard Machine)

① 確率変数 X の取りうる値の集合が有限集合の場合

② X が非負値確率変数

極限操作

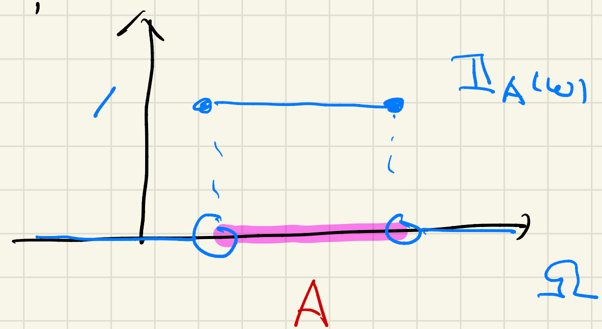
③ $X := X^+ - X^-$; $X^+ = \max\{X, 0\}$, $X^- = \max\{-X, 0\}$



Point

期待値の性質の証明 $\underbrace{}$ の方針 : $\mathbb{I}_A(\omega)$ $(A \in \mathcal{A})$ について示せばよい.

あとは Standard Machine !



① の段階

異なる $x_1, \ldots, x_n \in \mathbb{R}$ に対して

$$\{x_1, \ldots, x_n\} = \{X(\omega) ; \omega \in \Omega\}$$

とし

$$A_j = \{\omega \in \Omega ; X(\omega) = x_j\} \quad (j = 1, \ldots, n).$$

すると

$$X(\omega) = \sum_{j=1}^{n} x_j \mathbb{I}_{A_j}(\omega) \quad \leftarrow \text{単純な確率変数}$$

と書ける.

定義 1.18

$$E[X] = E\left[\sum_{j=1}^{n} x_j \mathbb{I}_{A_j}(\omega)\right] := \sum_{j=1}^{n} x_j \Pr(A_j)$$

と定める。

性質 (1) X と Y は単純な確率変数のとき

$$E[X+Y] = E[X] + E[Y]$$

が成立

$\Rightarrow$ XY も単純。

(2) さらに、X と Y が独立のとき、 $E[XY] = E[X]E[Y]$

② の段階

定義1.21 非負値確率変数 X に対して

$$E[X] := \sup\{E[Y];\ Y \text{は単純で} Y \leq X\}$$

ただし、$E[X] = \infty$ もゆるす。

③ の段階

$$X^+(\omega) := \max\{X(\omega), 0\};\quad X^-(\omega) := \max\{-X(\omega), 0\} \quad (\forall \omega \in \Omega)$$

と定める。すると、X^+ と X^- はともに非負値確率変数となる。

$E[X^+]$	$E[X^-]$	$E[X]$
有限	有限	$E[X^+]-E[X^-]$
∞	有限	∞
有限	∞	$-\infty$
∞	∞	定義しない

Point $(\Omega, \mathcal{A}, P_r)$ を確率空間, $X: \Omega \to \mathbb{R}$ としたところ,

X の期待値 $E[X]$ は $P_r(\cdot)$ を用いて定めている.

記法 $(\Omega, \mathcal{A}, \Pr)$ と確率変数 $X : \Omega \to \mathbb{R}$ に対して

$(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上の確率測度 P^X を

$$P^X(B) := \Pr(X \in B) \quad (\forall B \in \mathcal{B}(\mathbb{R}))$$

で定める.

$(\mathbb{R}, \mathcal{B}(\mathbb{R}), P^X)$ は X により誘導された確率空間!

$$(\Omega, \mathcal{A}, \Pr) \xrightarrow{X} (\mathbb{R}, \mathcal{B}(\mathbb{R}), P^X)$$

P^Xを「Xの分布」という

$(\Omega, \mathcal{A}, \Pr)$ ↑ background probability space ↑ Probabilists が気にしている確率空間

$(\mathbb{R}, \mathcal{B}(\mathbb{R}), P^X)$ ↑ 統計学者に気にしている確率空間.

$(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上の Lebesgue 測度

$a<b$ に対して, $\tilde{m}((a,b]) = b-a$ と定める.

すると, $\mathcal{B}(\mathbb{R})$ 上の測度 m が一つだけ存在して

$$m((a,b]) = \tilde{m}((a,b])$$

となることが知られている. この測度 m を $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 上の Lebesgue 測度という (完備化していない!).

非負値可測関数 p^X で、$\int_{\mathbb{R}} p^X(x)\,dm(x) = 1$ で

$$P^X(B) = \int_{\mathbb{R}} p^X(x)\, \mathbb{I}_B(x)\, dm(x) = \int_B p^X(x)\, dm(x) \quad (\forall B \in \mathcal{B}(\mathbb{R}))$$

をみたすとき、p^X を X の p.d.f. という。

命題1.24 確率変数 X は p.d.f. $p^X(\cdot)$ を持つとする。

可測関数 $h: \mathbb{R} \to \mathbb{R}$ は $E[|h(X)|] < \infty$ をみたすとする。

このとき

統計学者の無意識の公式

$$E[h(X)] = \int_{\mathbb{R}} h(x)\, p^X(x)\, dm(x)$$

が成り立つ。

証明：$\forall B \in \mathcal{B}(\mathbb{R})$ に対し，$h(x) = \mathbb{1}_B(x)$ とする．

このとき

$$E[h(X)] = E[\mathbb{1}_B(X)] = P_r(X \in B)$$

一方

$$\int_{\mathbb{R}} h(x)\, p^X(x)\, dm(x) = \int_{\mathbb{R}} \mathbb{1}_B(x)\, p^X(x)\, dm(x) = P^X(B) = P_r(X \in B)$$

よって，$h = \mathbb{1}_B$ のとき

$E[h(X)] = \int_{\mathbb{R}} h(x)\, p^X(x)\, dm(x)$ が成立 $\Rightarrow$ Standard Machine.

定義 1.25 $E[X^2] < \infty$ のとき

$$\mathrm{Var}[X] := E\left[(X - E[X])^2\right]$$

で定める

注意 1.26 $E[X^2] < \infty \Rightarrow E[|X|] < \infty \Rightarrow -\infty < E[X] < \infty$.

重要な性質

- 非負値確率変数 X に対して

$$E[X]=0 \Rightarrow \Pr(X=0)=1$$

$$E[X]=\int_0^\infty \Pr(X>x)\,dm(x)$$

- $$\Pr(X\geq t)\leq \frac{E[X]}{t} \quad (\forall t>0)$$

- $E[X^2]<\infty$ かつ $E[Y^2]<\infty$ のとき

$$E[|XY|]\leq\sqrt{E[X^2]\cdot E[Y^2]}$$

Cauchy-Schwarz

- $g: \mathbb{R} \to \mathbb{R}$ は凸関数. $E[|X|] < \infty$ かつ $E[|g(X)|] < \infty$ のとき

$$E[g(X)] \geq g(E[X])$$

Jensen

-

収束定理

Fatouの補題 $\exists C > -\infty$ に対し、$X_n \geq C \ (n \in \mathbb{N})$ とする。

このとき

$$E[\liminf_{n\to\infty} X_n] \leq \liminf_{n\to\infty} E[X_n]$$

・**優収束定理** $X_n \xrightarrow{a.s.} X$ かつ $\exists Y$ s.t. $E[Y] < \infty$ かつ

$|X_n| \leq Y$ のとき

$$\lim_{n\to\infty} E[X_n] = E[X].$$

1.5 条件付き期待値

X, Y を確率変数とする. $E[|X|] < \infty$ のとき

$$E[X \mid Y]$$

をどう定義するか?

p.d.f. を用いて定義

・統計学者 … ベイズ公式の条件付きバージョンにて.

$$E[X \mid Y = y] =: h(y)$$

を定めて, $E[X \mid Y] := h(Y)$ とする.

p.d.f. を用いず定義

・確率論者 … Radon-Nikodym の定理から直接 $E[X \mid Y]$ を定める.

結局，towering property

$$E[X] = E[E[X|\mathcal{Y}]]$$

が必要なだけ．