数理統計学

（～確率分布）

§1 確率の定義

Def.1.1（**Laplace** による素朴な確率の定義）
起こり得る場合の数が全体で n 個あり，それらの起こり方は全て同等であるとする．事象 A に対して A の起こる場合の数が全体の n 個のうち r 個であるとき，A の**確率** $P(A)$ を，

$$P(A) = \frac{r}{n}$$

と定義する．

e.g.1.1 サイコロを 2 回振った時に出た目の積が奇数になる確率は何か？
・起こり得る場合の数は全体で $6^2 = 36$ 通り
・出た目の積が奇数になる事象 A の起こる場合の数は 9 通り

$$\therefore P(A) = \frac{9}{36} = \frac{1}{4}$$

e.g.1.2 区間 (0,2] の中に 1 点 X を出鱈目に取る．どの点も取られる可能性は同じであるとする．点 X が区間 $[1/3, 3/2]$ に入っている確率は何か？
$\longrightarrow$ Laplace による定義では $\frac{\infty}{\infty}$ となってしまう．
$\longrightarrow$ 定義の見直しが必要

Def.1.2（σ-加法族，事象）
Ω を集合とする．また $\mathfrak{P}(\Omega) = \{A | A \subset \Omega\}$ とする．$\mathcal{F} \subset \mathfrak{P}(\Omega)$ が，

(1) $\forall \{A_i\}_{i=1}^{\infty} \subset \mathcal{F}, \displaystyle\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

(2) $\forall A \in \mathcal{F}, A^c \in \mathcal{F}$

(3) $\Omega \in \mathcal{F}$

を満たすとき，$\mathcal{F}$ を Ω の **σ-加法族**という．このとき $(\Omega, \mathcal{F})$ を**可測空間**といい，Ω を**標本空間**という．また $\mathcal{F}$ の元を**事象**という．Ω を**全事象**，$\emptyset$ を**空事象**という．$A \in \mathcal{F}$ に対し，$A^c := \Omega \setminus A$ を A の**余事象**という．

Prop.1.1 $\mathcal{F}$ を Ω の σ-加法族とする.

(1) $\emptyset \in \mathcal{F}$

(2) $\forall \{A_i\}_{i=1}^{\infty} \subset \mathcal{F}, \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$

〈proof〉

(1) $\Omega \in \mathcal{F}$ より $\emptyset = \Omega^c \in \mathcal{F}$

(2) $\{A_i\}_{i=1}^{\infty} \subset \mathcal{F}$ に対し, $A_i^c \in \mathcal{F}(\forall i \in \mathbf{N}), \bigcup_{i=1}^{\infty} A_i^c \in \mathcal{F}$ であるから,

$$\bigcap_{i=1}^{\infty} A_i = \left(\bigcup_{i=1}^{\infty} A_i^c \right)^c \in \mathcal{F} \ \blacksquare$$

Def.1.3.（Borel 集合族）

$\mathbf{R}^n$ において Euclid 距離が誘導する位相を $\mathcal{O}$ とする. $\mathcal{O}$ を含む最小の σ-加法族を $\mathbf{R}^n$ の **Borel 集合族**といい, $\mathcal{B}_n$ で表す.

Def.1.4（**Kolmogorov** による公理的な確率の定義）

Ω を集合とし, $\mathcal{F}$ を Ω の σ-加法族とする. 写像 $P : \mathcal{F} \to \mathbf{R}$ が,

(P1) $\forall A \in \mathcal{F}, 0 \leq P(A) \leq 1$

(P2) $P(\Omega) = 1$

(P3) $\forall \{A_i\}_{i=1}^{\infty} \subset \mathcal{F}, \left[(\forall i, j, i \neq j)[A_i \cap A_j = \emptyset] \Rightarrow P\left(\bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} P(A_i) \right]$

を満たすとき, P は $(\Omega, \mathcal{F})$ 上の**確率測度**であるといい, $(\Omega, \mathcal{F}, P)$ は**確率空間**であるという. また, $A \in \mathcal{F}$ に対し, $P(A)$ を事象 A の**確率**という.

Rem. (P3)は非交和の記号で次のようにかける.

$$P\left(\bigsqcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} P(A_i)$$

Rem.(P1)(P2)を満たすとき P は $(\Omega, \mathcal{F})$ 上の**測度**であるといい, $(\Omega, \mathcal{F}, P)$ は**測度空間**であるという.

e.g.1.3　e.g.1.2 において, $\Omega = \mathbf{R}, \mathcal{F} = \mathcal{B}_1 \subsetneq \mathfrak{P}(\Omega)$ とする.

$$P(A) = \frac{1}{2} \int_A \chi_{(0,2]}(t) dt \ (A \in \mathcal{F})$$

と定義すると P は $(\mathbf{R}, \mathcal{B}_1)$ 上の確率測度となる. 但し, χ_S は集合 S の**特性関数**である.

Prop.1.2 $(\Omega, \mathcal{F}, P)$ は確率空間とする．

(1) $P(\emptyset) = 0$

(2) $A \in \mathcal{F}$ に対し，$P(A^c) = 1 - P(A)$

(3) $A, B \in \mathcal{F}$ に対し，$A \subseteq B \Rightarrow P(A) \leq P(B)$（**単調性**）

(4) $A, B \in \mathcal{F}$ に対し，$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

(5) $\{A_i\}_{i=1}^N \subset \mathcal{F}$ に対し，

$$P\left(\bigcup_{i=1}^N A_i\right) \leq \sum_{i=1}^N P(A_i)$$

〈proof〉

(1) $\emptyset, \Omega \in \mathcal{F}$ で $\Omega \cap \emptyset = \emptyset$ より，　$P(\Omega) = P(\Omega \cap \emptyset) = P(\Omega) + P(\emptyset)$, $\therefore P(\emptyset) = 0$

(2) $A^c \in \mathcal{F}$ で $A \cap A^c = \emptyset$ より，$P(A) + P(A^c) = P(A \cup A^c) = P(\Omega) = 1$

(3) $B \setminus A \in \mathcal{F}$ で $A \cap (B \setminus A) = \emptyset$ より，

$$P(B) = P\big(A \cup (B \setminus A)\big) = P(A) + P(B \setminus A) \geq P(A)$$

(4) $P(A \cup B) = P(A) + P(B \setminus A), P(B) = P(A \cap B) + P(B \setminus A)$

$$\therefore P(A \cup B) - P(B) = P(A) - P(A \cap B)$$

$$\therefore P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

(5) $P(A_1 \cap A_2) \leq P(A_1) \cap P(A_2)$ より，

$$P\left(\bigcup_{i=1}^N A_i\right) \leq P\left(\bigcup_{i=1}^{N-1} A_i\right) + P(A_N) \leq \cdots \leq \sum_{i=1}^N P(A_i) \ \blacksquare$$

Prop.1.3 （**劣加法性**）

$(\Omega, \mathcal{F}, P)$ は確率空間とする．$\{A_i\}_{i=1}^N \subset \mathcal{F}$ に対し，

$$P\left(\bigcup_{i=1}^\infty A_i\right) \leq \sum_{i=1}^\infty P(A_i)$$

〈proof〉

$$B_1 = A_1, B_2 = A_2 \setminus A_1, B_3 = A_3 \setminus (A_1 \cap A_2), \cdots, B_j = A_j \setminus \bigcup_{i=1}^{j-1} A_i$$

とおくと，　$\displaystyle\bigcup_{j=1}^\infty B_j = \bigcup_{i=1}^\infty A_i$，　　$B_i \cap B_j = \emptyset \ (i \neq j)$，　　$\forall i \in \mathbf{N}, B_i \subseteq A_i$ より，

$$P\left(\bigcup_{i=1}^\infty A_i\right) = P\left(\bigcup_{i=1}^\infty B_i\right) = \sum_{i=1}^\infty P(B_i) \leq \sum_{i=1}^\infty P(A_i) \ \blacksquare$$

§2 条件付確率

Prop.2.1 $(\Omega,\mathcal{F},P)$ は確率空間とする．
$A\in\mathcal{F}$ $(P(A)>0)$ を固定し，写像 $P(\bullet\,|A):\mathcal{F}\to\mathbf{R}$ を，

$$P(B|A)=\frac{P(B\cap A)}{P(A)}\ (B\in\mathcal{F})$$

と定める．このとき $\big(\Omega,\mathcal{F},P(\bullet\,|A)\big)$ は確率空間となる．

Def.2.1（条件付確率）
$(\Omega,\mathcal{F},P)$ は確率空間とする．$A,B\in\mathcal{F},P(A)>0$ とする．

$$P(B|A)=\frac{P(A\cap B)}{P(A)}$$

を条件 A での B の**条件付確率**という．

e.g.2.1 サイコロを 1 回振った時,

$$\Omega=\{1{,}2{,}3{,}4{,}5{,}6\},\mathcal{F}=\mathfrak{P}(\Omega),P(A)=\frac{\#A}{\#\Omega}=\frac{\#A}{6}\ (A\in\mathcal{F})$$

$A=\{2{,}4{,}6\},B=\{1{,}2{,}3{,}4{,}5\}$ とおくと，$A,B\in\mathcal{F}$ である．

$$P(B|A)=\frac{\#(A\cap B)}{\#A}=\frac{2}{3}$$

一方，$\dfrac{\#(A\cap B)}{\#A}=\dfrac{\#\,(A\cap B)/6}{\#\,A/6}=\dfrac{P(A\cap B)}{P(A)}$

Thm.2.1（**Bayes の定理**）
$(\Omega,\mathcal{F},P)$ は確率空間とする．

$$\{A_j\}_{j=1}^{N}\subset\mathcal{F},\qquad \Omega=\bigsqcup_{j=1}^{N}A_j\,,\qquad \forall j\in\{1,\cdots,n\},P(A_j)>0$$

とする．このとき $P(B)>0$ を満たす全ての $B\in\mathcal{F}$ と $i\in\{1,\cdots,n\}$ に対して，

$$P(A_i|B)=\frac{P(A_i)P(B|A)}{\sum_{j=1}^{N}P(A_j)\,P(B|A_j)}$$

⟨proof⟩

$$P(B)=P(B\cap\Omega)=P\left(B\cap\bigsqcup_{j=1}^{N}A_j\right)=P\left(\bigsqcup_{j=1}^{N}(B\cap A_j)\right)=\sum_{j=1}^{N}P(B\cap A_j)$$

$$=\sum_{j=1}^{N}P(A_j)P(B|A_j)$$

（これを**全確率の定理**という.）また,

$$P(B)P(A_i|B) = P(A_i)P(B|A_i) = P(A_i \cap B)$$

から,

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{N} P(A_j)P(B|A_j)} \ \blacksquare$$

Def.2.2 （事象の独立）
$(\Omega, \mathcal{F}, P)$ は確率空間とする．２つの事象 $A, B \in \mathcal{F}$ が,

$$P(A \cap B) = P(A)P(B)$$

を満たすとき，A, B は**独立**であるという．

Prop.2.2 $(\Omega, \mathcal{F}, P)$ は確率空間とする．２つの事象 $A, B \in \mathcal{F}$ に対し次は同値である．
(1) A, B は独立
(2) $P(A), P(A^c) > 0 \ \Rightarrow \ P(B|A) = P(B|A^c) = P(B)$

⟨proof⟩

$(1) \Rightarrow (2)$

(i) $P(A) = 0$ または $P(A^c) = 0$ のとき，(2) は恒真命題である．

(ii) $0 < P(A) < 1$ のとき,

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B)$$

$$P(B|A^c) = \frac{P(A^c \cap B)}{P(A^c)} = \frac{P(B) - P(A \cap B)}{P(A^c)} = \frac{P(B)\big(1 - P(A)\big)}{1 - P(A)} = P(B)$$

$(1) \Leftarrow (2)$

(i) $P(A) = 0$ のとき，$P(A \cap B) \leq P(A) = 0$ より $P(A \cap B) = 0$，また $P(A)P(B) = 0$
よって $P(A \cap B) = P(A)P(B)$ が成り立っている．

(ii) $P(A^c) = 0$ のとき，$P(A) = 1 - P(A^c) = 1$

$$\begin{aligned} P(A)P(B) &= 1 \cdot P(B \cap \Omega) = P\big(B \cap (A \cup A^c)\big) = P\big((B \cap A) \sqcup (B \cap A^c)\big) \\ &= P(B \cap A) + P(B \cap A^c) = P(B \cap A) \end{aligned}$$

(iii) $P(A), P(A^c) > 0$ のとき，$P(A)P(B) = P(A)P(B|A) = P(A \cap B)$ ■

e.g.2.2 サイコロを１回振った時,

$$\Omega = \{1,2,3,4,5,6\}, \mathcal{F} = \mathfrak{P}(\Omega), P(A) = \frac{\#A}{\#\Omega} = \frac{\#A}{6} \ (A \in \mathcal{F})$$

とする．$A = \{2,4,6\}, B = \{1,2,3,4,5\}$ とおくと，$A, B \in \mathcal{F}$ である．

$$P(A) = \frac{3}{6} = \frac{1}{2}, P(B) = \frac{5}{6}, P(A \cap B) = \frac{2}{6} = \frac{1}{3}$$

であるから，$P(A\cap B)\neq P(A)P(B)$ なので A,B は独立でない．

Def.2.3.（ N 個の事象の独立）

$(\Omega,\mathcal{F},P)$ は確率空間とする．N 個の事象 $\{A_i\}_{i=1}^N\subset\mathcal{F}$ が，

$$\forall\{i_1,\cdots,i_M\}\subseteq\{1,\cdots,N\},P\left(\bigcap_{k=1}^M A_{i_k}\right)=\prod_{k=1}^M P(A_{i_k})$$

を満たすとき，$A_1,\cdots,A_N$ は**独立**であるという

Rem. $N=3$ のとき

$A_1,A_2,A_3\in\mathcal{F}$ に対し，

$$A_1,A_2,A_3\text{ が独立 }:\overset{\text{def}}{\Leftrightarrow}\begin{cases}P(A_1\cap A_2)=P(A_1)P(A_2)\\P(A_1\cap A_2)=P(A_1)P(A_2)\\P(A_1\cap A_2)=P(A_1)P(A_2)\\P(A_1\cap A_2\cap A_3)=P(A_1)P(A_2)P(A_3)\end{cases}$$

Prop.2.3 $(\Omega,\mathcal{F},P)$ は確率空間とする．N 個の事象 $\{A_i\}_{i=1}^N\subset\mathcal{F}$ に対し次は同値である．

(1) $A_1,\cdots,A_N$ は独立

(2) 各 $j\in\{1,\cdots,N\}$ 各部分集合 $\{i_1,\cdots,i_M\}\subseteq\{1,\cdots,N\}\setminus\{j\}$ に対して，

$$P\left(\bigcap_{k=1}^M B_{i_k}\right)>0\Rightarrow P(A_j)=P\left(A_j\,\middle|\,\bigcap_{k=1}^M B_{i_k}\right)$$

（但し，$k\in\{1,\cdots,M\}$ に対して，$B_{i_k}=A_{i_k}$ or $A_{i_k}^c$ ）

§3 確率変数

Def.3.1（確率変数）
$(\Omega, \mathcal{F})$ を可測空間とする．関数 $X:\Omega\to\mathbf{R}$ が，
$$\forall x\in\mathbf{R}, X^{-1}\big((-\infty,x]\big)\in\mathcal{F}$$
を満たすとき，X を**確率変数**という．

Rem. このとき $X^{-1}\big((-\infty,x]\big)=\{\omega\in\Omega \mid X(\omega)\le x\}$ は **$\mathcal{F}$-可測**であるという．

Prop.3.1. $(\Omega,\mathcal{F},P)$ を確率空間とし，$X:\Omega\to\mathbf{R}$ を確率変数とする．
(1) $\forall x\in\mathbf{R}$,
$$X^{-1}((-\infty,x)), X^{-1}(\{x\}), X^{-1}\big([x,\infty)\big), X^{-1}\big((x,\infty)\big)\in\mathcal{F}$$
(2) $\forall a,b\in\mathbf{R}$,
$$X^{-1}\big((a,b)\big), X^{-1}\big([a,b)\big), X^{-1}\big((a,b]\big), X^{-1}([a,b])\in\mathcal{F}$$

〈proof〉
$\forall x\in\mathbf{R}, X^{-1}\big((-\infty,x]\big)\in\mathcal{F}$ とする．
$$X^{-1}\big((x,\infty)\big)=X^{-1}((-\infty,x]^c)=X^{-1}\big((-\infty,x]\big)^c\in\mathcal{F}$$
よって，
$$X^{-1}\big((a,b]\big)=X^{-1}\big((a,\infty)\cap(-\infty,b]\big)=X^{-1}\big((a,\infty)\big)\cap X^{-1}\big((-\infty,b]\big)\in\mathcal{F}$$
これと $\{x\}=\bigcap_{i\in\mathbf{N}}\left(x-\frac{1}{i},x\right]$ から，
$$X^{-1}(\{x\})=X^{-1}\left(\bigcap_{i\in\mathbf{N}}\left(x-\frac{1}{i},x\right]\right)=\bigcap_{i\in\mathbf{N}}X^{-1}\left(x-\frac{1}{i},x\right]\in\mathcal{F}$$
これとの和集合を取れば他も従う．

Prop.3.2 $X:\Omega\to\mathbf{R}, Y:\Omega\to\mathbf{R}$ をそれぞれ確率変数とすると，
(1) $\alpha X+\beta Y:\omega\mapsto\alpha X(\omega)+\beta Y(\omega)$ も確率変数 $(\forall\alpha,\beta\in\mathbf{R})$
(2) $XY:\omega\mapsto X(\omega)Y(\omega)$ も確率変数
(3) $(\forall\omega\in\Omega, Y(\omega)\neq 0)\Rightarrow\dfrac{X}{Y}:\omega\mapsto\dfrac{X(\omega)}{Y(\omega)}$ も確率変数

Rem. 次のように略記する．
$P(X\le x):=P\left(X^{-1}\big((-\infty,x]\big)\right)$, $\qquad P(X<x):=P\left(X^{-1}\big((-\infty,x)\big)\right)$
$P(X=x):=P\big(X^{-1}(\{x\})\big)$

Def.3.2（累積分布関数）
$(\Omega, \mathcal{F}, P)$ を確率空間とし，$X : \Omega \to \mathbf{R}$ を確率変数とする．関数 $F_X : \mathbf{R} \to [0,1]$ を

$$F_X(x) = P(X \leq x), \qquad x \in \mathbf{R}$$

と定めたとき，F_X を X の**累積分布関数**という．

e.g.3.1 硬貨を 1 回投げるとき

$$\Omega = \{H, T\}, \mathcal{F} = \mathfrak{P}(\Omega) = \{\emptyset, \{H\}, \{T\}, \Omega\}, P(A) = \frac{\#A}{\#\Omega} = \frac{\#A}{2} \ (A \in \mathcal{F})$$

$$X : \Omega \longrightarrow \mathbf{R}, \omega \longmapsto X(\omega) := \begin{cases} 1 \ (\text{if } \omega = H) \\ 0 \ (\text{if } \omega = T) \end{cases}$$

$$P(X = 1) = P(\{\omega \in \Omega | X(\omega) = 1\}) = P(\{H\}) = \frac{1}{2}$$

$$P(X \leq 1) = P(\{\omega \in \Omega | X(\omega) \leq 1\}) = P(\Omega) = 1$$

$$P(X < 1) = P(\{\omega \in \Omega | X(\omega) < 1\}) = P(\{\mathrm{T}\}) = \frac{1}{2}$$

累積分布関数は,

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & (\text{if } x < 0) \\ 1/2 & (\text{if } 0 \leq x < 1) \\ 1 & (\text{if } 1 \leq x) \end{cases}$$

e.g.3.2 サイコロを 2 回振った時,

$$\Omega = \{(i,j) |\ i,j \in \{1,2,3,4,5,6\}\}, \mathcal{F} = \mathfrak{P}(\Omega), P(A) = \frac{\#A}{\#\Omega} = \frac{\#A}{36} \ (A \in \mathcal{F})$$

$$X : \Omega \longrightarrow \mathbf{R}, (i,j) \mapsto i, \qquad Y : \Omega \longrightarrow \mathbf{R}, (i,j) \mapsto j$$

出た目の数の和が 7 である確率は,

$$\begin{aligned} P(X + Y = 7) &= P(\{\omega \in \Omega | X(\omega) + Y(\omega) = 7\}) \\ &= P(\{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}) = \frac{6}{36} = \frac{1}{6} \end{aligned}$$

e.g.3.3 サイコロを何回も降って $i+1$ 回目に初めて 1 が出たら $X(\omega) = 1$ と定めると,

$$P(X = i) = \frac{5^i}{6^{i+1}} = \left(1 - \frac{1}{6}\right)^i \cdot \frac{1}{6} \ (i \in \mathbf{N} \cup \{0\})$$

Def.3.3.（離散型確率変数）
$(\Omega, \mathcal{F}, P)$ を確率空間とし，$X : \Omega \to \mathbf{R}$ を確率変数とする．

$$|X(\Omega)| \leq \aleph_0$$

であるとき，X を**離散型確率変数**という．

Rem. 上の定義において，

$$X(\Omega) = \{a_1, \cdots, a_N\} \text{ or } \{a_1, a_2, \cdots\} \ \left(a_i \neq a_j (i \neq j)\right)$$

と表し，$p_i = P(X = a_i)$ とおく．このとき，

$$F_X(x) = \sum_{\{i : a_i \leq x\}} p_i$$

となる．各 i に対して $p_i \geq 0$ であって，$A_i = X^{-1}(\{a_i\}) = \{\omega \in \Omega | X(\omega) = a_i\}$ とおくと，

$$\sum_i p_i = \sum_i P(A_i) = P\left(\bigcup_i A_i\right) = P(\Omega) = 1$$

但し，２つ目の $=$ は $A_i = X^{-1}(\{a_i\})$ より $i \neq j$ のとき

$$A_i \cap A_j = X^{-1}(\{a_i\}) \cap X^{-1}(\{a_j\}) \ = X^{-1}(\{a_i\} \cap \{a_j\}) = X^{-1}(\emptyset) = \emptyset$$

であることによる．

Def.3.4.（連続型確率変数）
$(\Omega, \mathcal{F}, P)$ を確率空間とし， $X : \Omega \to \mathbf{R}$ を確率変数とする．
F_X が連続関数であるとき，X を**連続型確率変数**という．このとき，

$$\forall x \in \mathbf{R}, F_X(x) = \int_{-\infty}^{x} f(t)\, dt$$

を満たすある非負値関数 $f : \mathbf{R} \to [0, \infty)$ が存在すれば，f を X の**確率密度関数**(p.d.f.)という．

Rem. 確率密度関数は， X が取り得ない値 x では $f(x) = 0$ と定義して実数全体で定義できているものとする．

Rem. 確率密度関数は一意的に定まらない．例えば，累積分布関数が

$$F_X\colon \mathbf{R} \longrightarrow [0{,}1], x \longmapsto \begin{cases} 0 \ (x \leq 0) \\ x \ (0 < x \leq 1) \\ 1 \ (1 < x) \end{cases}$$

である連続型確率変数の確率密度関数として，

$$f_1(x) = \begin{cases} 1 \ (0 < x < 1) \\ 0 \ (\text{otherwise}) \end{cases}, \qquad f_2(x) = \begin{cases} 1 \ (0 \leq x \leq 1) \\ 0 \ (\text{otherwise}) \end{cases}$$

などがある．

Prop.3.3. $(\Omega,\mathcal{F},P)$ を確率空間，$X:\Omega\to\mathbf{R}$ を連続型確率変数，f を X の確率密度関数とする．$\forall a,b,c\in\mathbf{R}\ (a<b)$,

$$P(X=c)=0,$$

$$P(a\le X\le b)=P(a<X\le b)=P(a\le X<b)=P(a<X<b)=\int_a^b f(t)\,dt$$

e.g.3.4　e.g.1.3 において，

$X:\Omega\longrightarrow\mathbf{R},\omega\mapsto\omega$ とおくと，

$$F_X\colon\mathbf{R}\longrightarrow[0,1],x\mapsto\begin{cases}0\ (x\le 0)\\ \dfrac{x}{2}\ \ (0<x\le 2)\\ 1\ (2<x)\end{cases}$$

は連続関数であるから X は連続型確率変数である．また，

$$f(x):=\frac{1}{2}\chi_{(0,2]}(x)\ (x\in\mathbf{R})$$

と定めると，

$$\forall x\in\mathbf{R},P(X\le x)=\int_{-\infty}^{x}f(t)\,dt$$

であるから，f は X の確率密度関数である．

$$P\left(\left[\frac{1}{3},\frac{3}{2}\right]\right)=P\left(\frac{1}{3}\le X\le\frac{3}{2}\right)=\int_{\frac{1}{3}}^{\frac{3}{2}}f(t)\,dt=\frac{1}{2}\int_{\frac{1}{3}}^{\frac{3}{2}}dt=\frac{7}{12}$$

Rem. 離散型でも連続型でもない確率変数は存在する．

Rem. 一般に X が連続型確率変数で f を確率密度関数とするとき，,

$$\forall B\in\mathcal{B}_1,P(X\in B)=\int_B f(t)\,dt$$

§4 期待値と分散

Def.4.1（平均（期待値））

(1) $X:\Omega\to\mathbf{R}$ が離散型確率変数である場合,

$$X(\Omega)=\{a_1,\cdots,a_N\}\ \text{or}\ \{a_1,a_2,\cdots\}\ \big(a_i\neq a_j(i\neq j)\big)$$

と表し，$p_i=P(X=a_i)$ とおく．$\displaystyle\sum_i|a_i|p_i<\infty$ のとき,

$$\mu=E[X]:=\sum_i a_i p_X(a_i)$$

を X の**平均**（**期待値**）という.

(2) X が確率密度関数 f を持つ連続型確率変数である場合,

$\displaystyle\int_{-\infty}^{\infty}|x|f(x)\,dx<\infty$ のとき,

$$\mu=E[X]:=\int_{-\infty}^{\infty}xf(x)\,dx$$

を X の平均（期待値）という.

Def.4.2（平均・期待値）

(1) $X:\Omega\to\mathbf{R}$ が離散型確率変数である場合,

$g\circ X:\Omega\to\mathbf{R},\omega\mapsto g\big(X(\omega)\big)$ が離散型確率変数となるような関数 $g:\mathbf{R}\to\mathbf{R}$ に対し,

$\displaystyle\sum_i|g(a_i)|p_i<\infty$ のとき,

$$E[g(X)]:=\sum_i g(a_i)p_i$$

を $g\circ X$ の平均（期待値）という.

(2) $X:\Omega\to\mathbf{R}$ が確率密度関数 f を持つ連続型確率変数である場合,

$g\circ X:\Omega\to\mathbf{R},\omega\mapsto g\big(X(\omega)\big)$ が連続型確率変数となるような関数 $g:\mathbf{R}\to\mathbf{R}$ に対し,

$\displaystyle\int_{-\infty}^{\infty}|g(x)|f(x)\,dx<\infty$ のとき,

$$E[g(X)]:=\int_{-\infty}^{\infty}g(x)f(x)\,dx$$

を $g\circ X$ の平均（期待値）という.

Rem. g としては $g(x)=(x-\mu)^2,g(x)=e^{tx}$ などを使う.

Prop.4.1 （期待値の線型性）
X を確率変数とする．$E[g(X)], E[h(X)]$ が存在するような関数 $g, h : \mathbf{R} \to \mathbf{R}$ に対し，
$$E[\alpha g(X) + \beta h(X)] = \alpha E[g(X)] + \beta E[h(X)] \ (\forall \alpha, \beta \in \mathbf{R})$$
が成り立つ．

〈proof〉総和または積分の線型性から従う．■

Def.4.3 （分散）
確率変数 X に対し，$E[X^2]$ が存在するとき，
$$V[X] := E[(X-\mu)^2] \ (\mu := E[X])$$
を X の**分散**といい，$\sigma := \sqrt{V[X]}$ を**標準偏差**という．

Rem. $E[X^2]$ が存在するとき，$E[X]$ は存在する．（Cauchy-Shwarz の不等式による．）

Rem. X が離散型，連続型のとき，
$$V[X] = \sum_i (a_i - \mu)^2 p_i\,, \qquad V[X] := \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$$

Prop.4.2 X, Y を確率変数とし，$E[X^2]$ が存在するとする．このとき，
(1) $V[X] = E[X^2] - \mu^2$
(2) $V[\alpha X + \beta] = \alpha^2 V[X] \ (\forall \alpha, \beta \in \mathbf{R})$

〈proof〉
(1) $V[X] = E[(X-\mu)^2] = E[X^2 - 2\mu X + \mu^2] = E[X^2] - 2\mu E[X] + \mu^2 = E[X^2] - \mu^2$
(2) $V[\alpha X + \beta] = E[(\alpha X + \beta - E[\alpha X + \beta])^2] = E[(\alpha X + \beta - \alpha E[X] - \beta)^2]$
$= E[\alpha^2 (X - E[X])^2] = \alpha^2 V[X]$ ■

§5 確率ベクトル

Def.5.1（ n 次元確率ベクトル）
$(\Omega,\mathcal{F})$ を可測空間とする．関数 $\boldsymbol{X}=(X_1,\cdots,X_n):\Omega\to\mathbf{R}^n$ が，

$$\forall\boldsymbol{x}=(x_1,\cdots,x_n)\in\mathbf{R}^n, \boldsymbol{X}^{-1}\left(\prod_{i=1}^{n}(-\infty,x_i]\right)\in\mathcal{F}$$

を満たすとき，$\boldsymbol{X}$ を $\boldsymbol{n}$ **次元確率ベクトル**という．

Rem. このとき $\boldsymbol{X}^{-1}\left(\prod_{i=1}^{n}(-\infty,x_i]\right)$ は $\boldsymbol{\mathcal{F}}$**-可測**であるという．

Rem. Π は集合の直積の意味である．

$$\prod_{i=1}^{n}(-\infty,x_i]:=(-\infty,x_1]\times\cdots\times(-\infty,x_n]$$

$$\begin{aligned}\boldsymbol{X}^{-1}\left(\prod_{i=1}^{n}(-\infty,x_i]\right)&=\left\{\omega\in\Omega:\boldsymbol{X}(\omega)=\big(X_1(\omega),\cdots,X_n(\omega)\big)\in\prod_{i=1}^{n}(-\infty,x_i]\right\}\\&=\{\omega\in\Omega:\forall i\in\{1,\cdots,n\},X_i(\omega)\in(-\infty,x_i]\}\\&=\bigcap_{i=1}^{n}X_i^{-1}\big((-\infty,x_i]\big)\end{aligned}$$

Prop.5.1 X_i $(i\in\{1,\cdots,n\})$ が確率変数であるとき，関数

$$\boldsymbol{X}=(X_1,\cdots,X_n):\Omega\to\mathbf{R}^n,\qquad \boldsymbol{X}(\omega):=\big(X_1(\omega),\cdots,X_n(\omega)\big)$$

は n 次元確率ベクトルになる．

〈proof〉 $\forall i\in\{1,\cdots,n\},\forall(x_1,\cdots,x_n)\in\mathbf{R}^n,X_i^{-1}\big((-\infty,x_i]\big)\in\mathcal{F}$ より

$$\boldsymbol{X}^{-1}\left(\prod_{i=1}^{n}(-\infty,x_i]\right)=\bigcap_{i=1}^{n}X_i^{-1}\big((-\infty,x_i]\big)\in\mathcal{F}\ \blacksquare$$

Def.5.2（同時分布関数）
$(\Omega,\mathcal{F},P)$ を確率空間とし，$\boldsymbol{X}:\Omega\to\mathbf{R}^n$ を n 次元確率ベクトルとする．
関数 $F_X:\mathbf{R}^n\to[0,1]$ を

$$F_{\boldsymbol{X}}(\boldsymbol{x})=P(X_1\le x_1,\cdots,X_n\le x_n),\qquad \boldsymbol{x}=(x_1,\cdots,x_n)\in\mathbf{R}^n$$

と定めたとき，$F_{\boldsymbol{X}}$ を $\boldsymbol{X}$ の**同時分布関数**という．

Rem.

$$P(X_1\le x_1,\cdots,X_n\le x_n):=P\left(\boldsymbol{X}^{-1}\left(\prod_{i=1}^{n}(-\infty,x_i]\right)\right)$$

Def.5.3（周辺分布）
$\boldsymbol{X} = (X_1, \cdots, X_n) : \Omega \to \mathbf{R}^n$ を n 次元確率ベクトル，$F_{\boldsymbol{X}}$ を $\boldsymbol{X}$ の同時分布関数とする．

$$F_{X_i} : \mathbf{R} \to [0,1] \ , \qquad F_{X_i}(x) = \lim_{x_j \to \infty (j \neq i)} F_{\boldsymbol{X}}\left(x_1, \cdots, \underset{i}{\underset{\smile}{x}}, \cdots, x_n\right)$$

を X_i の**周辺分布関数**という．

Prop.5.2 $\boldsymbol{X} = (X_1, \cdots, X_n)$を n 次元確率ベクトル， F_{X_i} を X_i の周辺分布関数とすると，

$$F_{X_i}(x) = P(X_i \leq x) \ (x \in \mathbf{R})$$

Def.5.4 （ n 次元離散型確率ベクトル）
$(\Omega, \mathcal{F}, P)$ を確率空間とし，$\boldsymbol{X} : \Omega \to \mathbf{R}^n$ を n 次元確率ベクトルとする．

$$|\boldsymbol{X}(\Omega)| \leq \aleph_0$$

であるとき，$\boldsymbol{X}$ を $\boldsymbol{n}$ **次元離散型確率ベクトル**という．

Rem. $\boldsymbol{X} = (X_1, \cdots, X_n) : \Omega \to \mathbf{R}^n$ を n 次元離散型確率ベクトルとする．

$$X_i(\Omega) = \{a_i(j) |\ j \in J_i\} \ \big(j \neq j' \Rightarrow a_i(j) \neq a_i(j')\big)$$

であるとき，

$$p_{j_1 \cdots j_n} := P\big(\boldsymbol{X} = (a_1(j_1), \cdots, a_n(j_n))\big)$$

が定まる．このとき，

$$p_{j_1 \cdots j_n} \geq 0, \qquad \sum_{j_n \in J_n} \cdots \sum_{j_1 \in J_1} p_{j_1 \cdots j_n} = 1$$

Rem. X_i $(i \in \{1, \cdots, n\})$ が離散型確率変数であるとき，

$$\boldsymbol{X} = (X_1, \cdots, X_n) : \Omega \to \mathbf{R}^n, \qquad \boldsymbol{X}(\omega) := \big(X_1(\omega), \cdots, X_n(\omega)\big)$$

は n 次元離散型確率ベクトルになる．

Def.5.5 （ n 次元連続型確率ベクトル）
$(\Omega, \mathcal{F}, P)$ を確率空間とし，$\boldsymbol{X} : \Omega \to \mathbf{R}^n$ を n 次元確率ベクトルとする．
$F_{\boldsymbol{X}}$ が連続関数であるとき，$\boldsymbol{X}$ を $\boldsymbol{n}$ **次元連続型確率ベクトル**という．このとき，

$$\forall \boldsymbol{x} = (x_1, \cdots, x_n) \in \mathbf{R}^n, F_{\boldsymbol{X}}(\boldsymbol{x}) = \int_{\prod_{i=1}^n (-\infty, x_i]} f(\boldsymbol{t}) \, d\boldsymbol{t} = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(t_1, \cdots, t_n) \, dt_1 \cdots dt_n$$

を満たすある非負値関数 $f : \mathbf{R}^n \to [0, \infty)$ が存在すれば，f を $\boldsymbol{X}$ の**同時確率密度関数**という．

Rem. 次が成り立つ．

$$\int_{\mathbf{R}^n} f(\boldsymbol{x}) \, d\boldsymbol{x} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(t_1, \cdots, t_n) \, dt_1 \cdots dt_n = 1$$

Rem. X_i $(i \in \{1, \cdots, n\})$ が $f_i : \mathbf{R} \to [0, \infty)$ を確率密度関数に持つ連続型確率変数であるとき,

$$\boldsymbol{X} = (X_1, \cdots, X_n) : \Omega \to \mathbf{R}^n, \qquad \boldsymbol{X}(\omega) := \big(X_1(\omega), \cdots, X_n(\omega)\big)$$

は $f : \mathbf{R}^n \to [0, \infty), f(x_1, \cdots, x_n) = \prod_{\mathrm{i}=1}^{\mathrm{n}} f_i(x_i)$ を同時確率密度関数に持つ n 次元離散型確率ベクトルになる.

Rem. $B \subset \mathcal{B}_n$ に対して,

$$\int_B f(\boldsymbol{x})\, d\boldsymbol{x} = P(\boldsymbol{X} \in B)$$

Def.5.6 連続型確率ベクトル $\boldsymbol{X} = (X_1, \cdots, X_n)$ が f を同時確率密度関数として持つとする.

$$f_i(x) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f\left(t_1, \cdots, \underset{i}{\underset{\smile}{x}}, \cdots, t_n\right) dt_1 \cdots dt_{i-1} dt_{i+1} \cdots dt_n$$

を X_i の**周辺確率密度関数**といい,確率密度関数が f_i で与えられる連続分布を X_i の**周辺分布**という.(連続分布は§7 にて定義される.)

期待値と分散が同様に定義される.

Def.5.7(期待値と分散)

$\boldsymbol{X} : \Omega \to \mathbf{R}^n$ を n 次元離散型確率ベクトルとする.Def.5.4 の Rem の記号を使う.

$\displaystyle\sum_{j_n \in J_n} \cdots \sum_{j_1 \in J_1} \|\big(a_1(j_1), \cdots, a_n(j_n)\big)\| p_{j_1 \cdots j_n} < \infty$ のとき,

$$\mu = E[\boldsymbol{X}] := \sum_{j_n \in J_n} \cdots \sum_{j_1 \in J_1} (a_1(j_1), \cdots, a_n(j_n)) p_{j_1 \cdots j_n}$$

を $\boldsymbol{X}$ の平均(期待値)という.

$\boldsymbol{X} : \Omega \to \mathbf{R}^n$ を n 次元連続型確率ベクトルとする.Def.5.5 の Rem の記号を使う.

$\displaystyle\int_{\mathbf{R}^n} \|\boldsymbol{x}\| f(\boldsymbol{x})\, d\boldsymbol{x} < \infty$ のとき,

$$\mu = E[\boldsymbol{X}] := \int_{\mathbf{R}^n} \boldsymbol{x} f(\boldsymbol{x})\, d\boldsymbol{x}$$

を $\boldsymbol{X}$ の平均(期待値)という.

§6 確率変数の独立

Def.6.1 （確率変数の独立）
$X:\Omega\to\mathbf{R}, Y:\Omega\to\mathbf{R}$ を確率変数とする．X,Y が**独立**である．

$$:\overset{\text{def}}{\Leftrightarrow} \forall a,b\in\mathbf{R}(a<b),\forall c,d\in\mathbf{R}(c<d),\left[X^{-1}\big((a,b]\big),Y^{-1}\big((c,d]\big)\text{が独立である}\right]$$

Rem. $X^{-1}\big((a,b]\big),Y^{-1}\big((c,d]\big)$が独立

$$\Leftrightarrow P\left(X^{-1}\big((a,b]\big)\cap Y^{-1}\big((c,d]\big)\right)=P(X^{-1}\big((a,b]\big))P(Y^{-1}\big((c,d]\big))$$
$$\Leftrightarrow P(a<X\le b,c<Y\le d)=P(a<X\le b)P(c<Y\le d)$$

e.g.6.1 サイコロを 2 回振るとき 1 回目，2 回目の目の数を X,Y とすると，X,Y は独立であるが，$X,Z:=Y-X$ は独立でない．実際,

$$0=P(X=5,Z=3)\neq P(X=5)P(Z=3)=\frac{1}{6}\cdot\frac{3}{36}=\frac{1}{72}$$

Prop.6.1（離散型確率変数の独立）
$X:\Omega\to\mathbf{R}, Y:\Omega\to\mathbf{R}$ が共に離散型確率変数である場合

$$X(\Omega)=\{a_i|\ i\in I\}\ \big(I=\{1,\cdots,M\}\text{ or }\mathbf{N},a_i\neq a_{i'}(i\neq i')\big)$$
$$Y(\Omega)=\{b_j|\ j\in J\}\ \big(J=\{1,\cdots,N\}\text{ or }\mathbf{N},b_j\neq b_{j'}(j\neq j')\big)$$

と表したとき，$p_{ij}=P\left((X,Y)=(a_i,b_j)\right),p_i=P(X=a_i),q_j=P(Y=b_j)$ とおくと，

$$X,Y\text{ が独立 }\Leftrightarrow\forall(i,j)\in I\times J,p_{ij}=p_ip_j$$

Prop.6.2（連続型確率変数の独立）
$X:\Omega\to\mathbf{R}, Y:\Omega\to\mathbf{R}$ が共に連続型確率変数である場合
(X,Y) の同時確率密度関数 $f:\mathbf{R}^2\to\mathbf{R}$ が存在するとし，f_1 を X の周辺確率密度関数, f_2 を Y の周辺確率密度関数とする．このとき，

$$f\in C(\mathbf{R}^2)\text{ または }\exists D\in\mathcal{B}_2\setminus\{\emptyset,\mathbf{R}^2\},\ f\in C(D),f|_{\mathbf{R}^2\setminus D}=0$$
$$\Rightarrow\left[X,Y\text{ が独立 }\Leftrightarrow\forall(x,y)\in\mathbf{R}^2,f(x,y)=f_1(x)f_2(y)\right]$$

Rem. $C(D):=\{(f:D\to\mathbf{R}):f\text{ is continuous}\}$

Prop.6.3 確率変数 X,Y が独立であれば，

$$E[XY]=E[X]E[Y]$$

〈proof〉 離散型のとき，

$$E[XY]=\sum_{(i,j)\in I\times J}a_ib_jp_{ij}=\sum_j\sum_i a_ib_jp_iq_j=\sum_i a_ip_i\sum_j b_jq_j=E[X]E[Y]$$

連続型のとき，

$$\begin{aligned}E[XY]&=\iint_{\mathbf{R}^2}xyf(x,y)d\boldsymbol{x}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf_1(x)f_2(y)\,dx\,dy\\&=\int_{-\infty}^{\infty}xf_1(x)\,dx\int_{-\infty}^{\infty}yf_2(y)\,dx=E[X]E[Y]\ \blacksquare\end{aligned}$$

Prop.6.4 独立な確率変数 X,Y と $\alpha,\beta\in\mathbf{R}$ に対して，

$$V[\alpha X+\beta Y]=\alpha^2V[X]+\beta^2V[Y]$$

〈proof〉 $E[X]=\mu_1,E[Y]=\mu_2$ とおく．

$$\begin{aligned}V[\alpha X+\beta Y]&=E\Big[\big(\alpha X+\beta Y-(\alpha\mu_1+\beta\mu_2)\big)^2\Big]=E\Big[\big(\alpha(X-\mu_1)+\beta(Y-\mu_2)\big)^2\Big]\\&=E[\alpha^2(X-\mu_1)^2+2\alpha\beta(X-\mu_1)(Y-\mu_2)+\beta^2(Y-\mu_2)^2]\end{aligned}$$

ここで $E[(X-\mu_1)(Y-\mu_2)]=E[X-\mu_1]E[Y-\mu_2]=0\cdot 0=0$ （独立性による.）より，

$$V[\alpha X+\beta Y]=\alpha^2V[X]+\beta^2V[Y]\ \blacksquare$$

Rem. X,Y が独立であるとき，実関数 ϕ,ψ に対して，$\phi(X),\psi(Y)$ も独立である．
特に $X-\mu_1$ と $Y-\mu_2$ は独立である．

Def.6.2（共分散）
X ,Y を確率変数とする．

$$\mathrm{Cov}(X,Y):=E[(X-E[X])(Y-E[Y])],\qquad \rho(X,Y):=\frac{\mathrm{Cov}(X,Y)}{\sqrt{V[X]V[Y]}}$$

と定義し，それぞれ X,Y の**共分散**，**相関係数**という．

Prop.6.5 確率変数 X,Y に対して，共分散の定義から次が従う．

$$\mathrm{Cov}(X,Y)=E[XY]-E[X]E[Y]$$

特に X,Y が独立であるとき，$\mathrm{Cov}(X,Y)=0$

〈proof〉 $E[(X-E[X])(Y-E[Y])]=E[XY-XE[Y]-E[X]Y+E[X]E[Y]]$

$$\begin{aligned}&=E[XY]-E[X]E[Y]-E[X]E[Y]+E[X]E[Y]\\&=E[XY]-E[X]E[Y]\ \blacksquare\end{aligned}$$

Prop.6.6 確率変数 X,Y に対して，

$$|\rho(X,Y)| \le 1$$

等号成立 ($\rho(X,Y)=\pm1$) は，$\exists a\in\mathbf{R}^\times, b\in\mathbf{R}, X=aY+b$ のときで，このとき

$$\frac{X-E[X]}{\sqrt{V[X]}} = \pm\frac{Y-E[Y]}{\sqrt{V[Y]}} \quad \Big(\rho(X,Y)=\pm1 \text{ と複合同順}\Big)$$

$\langle$proof$\rangle$ $t\in\mathbf{R}$ とする．

$$0 \le V[tX-Y] = E\Big[\big((tX-Y)-E[tX-Y]\big)^2\Big] = E\Big[\big(t(X-E[X])-(Y-E[Y])\big)^2\Big]$$
$$= V[X]t^2 - 2\,\mathrm{Cov}(X,Y)\,t + V[Y]$$

これが任意の $t\in\mathbf{R}$ に対して成り立つので，$V[X]t^2-2\,\mathrm{Cov}(X,Y)\,t+V[Y]=0$ の判別式を D とすると，

$$\frac{D}{4} = (\mathrm{Cov}(X,Y))^2 - V[X]\ V[Y] \le 0, \qquad \therefore \frac{(\mathrm{Cov}(X,Y))^2}{V[X]V[Y]} \le 1, \qquad \therefore |\rho(X,Y)| \le 1$$

$\rho(X,Y)=\pm1$ となるのは $D=0$ のときで，

$$\left(V[X]t^2 - 2\,\mathrm{Cov}(X,Y)\,t + V[Y]\right)\Big|_{t=\frac{\mathrm{Cov}(X,Y)}{V[X]}} = 0$$

となり，

$$E\left[\left(\frac{\mathrm{Cov}(X,Y)}{V[X]}(X-E[X])-(Y-E[Y])\right)^2\right]=0$$
$$\therefore \frac{\mathrm{Cov}(X,Y)}{V[X]}(X-E[X])-(Y-E[Y])=0$$

$\rho(X,Y)=\pm1$ より $\dfrac{\mathrm{Cov}(X,Y)}{\sqrt{V[X]V[Y]}}=\pm1$ なので，

$$\frac{X-E[X]}{\sqrt{V[X]}} = \pm\frac{Y-E[Y]}{\sqrt{V[Y]}}$$

を得る．■

§7 確率分布

確率変数を Ω 上の関数として定義したが，関心があるのは取り得る値とその確率または確率密度関数である．これらによってどこにどのような割合で散らばっているのかがわかる．この散らばりを確率分布または単に分布という．

Prop.7.1. $(\Omega, \mathcal{F}, P)$ を確率空間とし，$X : \Omega \to \mathbf{R}$ を確率変数とする．

$$P_X : \mathcal{B}_1 \to [0,1], \qquad P_X(B) := P\left(X^{-1}(B)\right)\ (B \in \mathcal{B}_1)$$

と定めると，P_X は可測空間 $(\mathbf{R}, \mathcal{B}_1)$ 上の確率測度となる．

〈proof〉まず $X^{-1}(B) \in \mathcal{F}$ （証明略）から $P\left(X^{-1}(B)\right)$ は定義できる．

P_X が(P1)(P2)(P3)を満たすことを確認する．

(P1)は定義から自明．(P2) $P_X(\mathbf{R}) = P\left(X^{-1}(\mathbf{R})\right) = P(\Omega) = 1$

(P3) $\{B_i\}_{i=1}^{\infty} \subset \mathcal{B}_1$ は $(\forall i, j, i \neq j)[B_i \cap B_j = \emptyset]$ を満たすとする．

$$P_X\left(\bigcup_{i=1}^{\infty} B_i\right) = P\left(X^{-1}\left(\bigcup_{i=1}^{\infty} B_i\right)\right) = P\left(\bigcup_{i=1}^{\infty} X^{-1}(B_i)\right) = \sum_{i=1}^{\infty} P\left(X^{-1}(B_i)\right) = \sum_{i=1}^{\infty} P_X(B_i) \ \blacksquare$$

Def.7.1.（確率分布）

Prop.7.1.で定義した P_X を X の**確率分布**といい，$(\mathbf{R}, \mathcal{B}_1, P_X)$ を X によって**誘導された確率空間**という．このとき X は確率分布 P_X に従うといい，$X \sim P_X$ で表す．

Rem. 離散型, 連続型確率変数に対応する確率分布はそれぞれ**離散型確率分布**（**離散分布**），**連続型確率分布**（**連続分布**）と呼ばれる．

Rem. 累積分布関数は

$$F_X(x) = P_X\left((-\infty, x]\right)\ ((-\infty, x] \in \mathcal{B}_1)$$

と表せる．

[離散分布]

Prop.7.2 $X:\Omega\to\mathbf{R}$ を離散型確率変数とする．

$$X(\Omega)=\{a_1,\cdots,a_N\}\text{ or }\{a_1,a_2,\cdots\}\ \bigl(a_i\neq a_j(i\neq j)\bigr)$$

と表し，$p_i=P(X=a_i)$ とおく．X に対応する離散分布は，

$$P_X(B)=\sum_{a_i\in B}p_i\,,\qquad (B\in\mathcal{B}_1)$$

で与えられる．

重要な Rem. 離散分布は $a_i\in X(\Omega)$ と $p_i=P(X=a_i)$ の対応と考えることができる．
そこでここからは $a_i\in X(\Omega)$ と $P(X=a_i)$ の対応を与えることで離散分布を紹介していく．

Def.7.2（二項分布）
$n\in\mathbf{N},0<p<1$ とする．離散型確率変数 $X:\Omega\to\mathbf{R}$ が $X(\Omega)=\{0,1,\cdots,n\}$ を満たし，

$$\forall i\in\{0,1,\cdots,n\},P(X=i)=\binom{n}{i}p^i(1-p)^{n-i}$$

を満たすとき， X は母数 p の**二項分布** $B(n,p)$ に従うといい，$X\sim B(n,p)$ とかく．
特に，$B(1,p)$ を比率 p の **Bernoulli 分布**といい，$B(p)$ とかく．

Rem.

$$P(X\in\{0,1,\cdots,n\})=\sum_{i=0}^{n}P(X=i)=\sum_{i=0}^{n}\binom{n}{i}p^i(1-p)^{n-i}=\bigl(p+(1-p)\bigr)^n=1$$

e.g.7.1 サイコロを 5 回振って 1 が出る回数を X とするとき，

$$\Omega=\{1,2,3,4,5,6\}^5,X(\Omega)=\{0,1,2,3,4,5\}$$

$$\forall i\in\{0,1,2,3,4,5\},P(X=i)=\binom{5}{i}\left(\frac{1}{6}\right)^i\left(1-\frac{1}{6}\right)^{5-i}$$

が成り立つので，$X\sim B\left(5,\frac{1}{6}\right)$ ．

Def.7.3（Poisson 分布）
$\lambda>0$ とする．離散型確率変数 $X:\Omega\to\mathbf{R}$ が $X(\Omega)=\mathbf{N}\cup\{0\}$ を満たし，

$$\forall i\in\mathbf{N}\cup\{0\},P(X=i)=e^{-\lambda}\frac{\lambda^i}{i!}$$

を満たすとき，X は平均 λ の **Poisson 分布** $Po(\lambda)$ に従うといい，$X\sim Po(\lambda)$ とかく．

Rem. e^{λ} の Maclaurin 展開を用いて，

$$P(X\in\mathbf{N}\cup\{0\})=\sum_{i=0}^{n}P(X=i)=e^{-\lambda}\sum_{i=0}^{n}\frac{\lambda^i}{i!}=e^{-\lambda}e^{\lambda}=1$$

Prop.7.3（**小数の法則**）

$\lim_{n\to\infty} np_n = \lambda > 0$ であり $X_n \sim B(n, p_n), X \sim Po(\lambda)$ とすると，

$$\lim_{n\to\infty} P(X_n = i) = P(X = i)$$

が成り立つ．

$\langle$proof$\rangle$ $i \in \{0\} \cup \mathbf{N}$ とする．$i \leq n$ なる n を取る．

$$P(X_n = i) = \binom{n}{i} p^i (1-p)^{n-i} = \frac{n!}{i!\,(n-i)!} \cdot \frac{(np_n)^i}{n^i} \left(1 - \frac{1}{n}(np_n)\right)^{n-i}$$

$$= \frac{(np_n)^i}{i!} \cdot 1 \cdot \left(1 - \frac{1}{n}\right) \cdot \cdots \cdot \left(1 - \frac{i-1}{n}\right) \left\{\left(1 - \frac{1}{n}(np_n)\right)^{\frac{n}{np_n}}\right\}^{np_n} \left(1 - \frac{1}{n}(np_n)\right)^{-i}$$

$$\xrightarrow{n\to\infty} e^{-\lambda} \frac{\lambda^i}{i!}$$

Rem. Poisson 分布は単位時間あたり平均 λ 回ランダムに起こる（滅多に起こらない）イベントが単位時間に起こる回数の分布と考えられる．

e.g.7.2　1 時間に平均 4 回電話がランダムにかかってくるコールセンターにおいて 1 時間に 6 回電話がかかってくる確率は，

$$P(X = 6) = e^{-4} \frac{4^6}{6!} = \frac{256}{25e^4} \approx 0.18755$$

Def.7.4（幾何分布）

$0 < p < 1$ とする．離散型確率変数 $X : \Omega \to \mathbf{R}$ が $X(\Omega) = \mathbf{N} \cup \{0\}$ を満たし，

$$\forall i \in \mathbf{N} \cup \{0\}, P(X = i) = (1-p)^i p$$

を満たすとき，X は**幾何分布** $Geo(p)$ に従うといい，$X \sim Geo(p)$ とかく．

Rem.

$$P(X \in \mathbf{N} \cup \{0\}) = \sum_{i=0}^{\infty} P(X = i) = p \sum_{i=0}^{\infty} (1-p)^i = p \cdot \frac{1}{1-(1-p)} = 1$$

e.g.7.3 サイコロを何度も振って $i+1$ 回目に初めて 1 が出たら $X(\omega) = 1$ とするとき，

$$X(\omega) = \mathbf{N} \cup \{0\}, \qquad \forall i \in \mathbf{N} \cup \{0\}, P(X = i) = \left(1 - \frac{1}{6}\right)^i \cdot \frac{1}{6}$$

が成り立つので，$X \sim Geo\left(\frac{1}{6}\right)$ ．

[連続分布]

Prop.7.4 確率変数 X が確率密度関数 f を持つとすると, X に対応する連続分布は,

$$P_X(B) = \int_B f(t)dt, \quad (B \in \mathcal{B}_1)$$

で与えられる.

Def.7.5（一様分布）

$a, b \in \mathbf{R}$ $(a < b)$ を用いて確率密度関数 f が,

$$f(x) = \frac{1}{b-a}\chi_{[a,b]}(x) = \begin{cases} \dfrac{1}{b-a} & (a \le x \le b) \\ 0 & (\text{otherwise}) \end{cases}$$

で与えられる連続分布を区間 (a, b) 上の**一様分布**といい, $X \sim U(a, b)$ とかく.

e.g.7.4 区間 [0,2] の中に 1 点 X をランダムに取る.

$$\Omega = \mathbf{R}, \mathcal{F} = \mathcal{B}_1 \subsetneq \mathfrak{P}(\Omega), P(A) = \frac{1}{2}\int_A \chi_{[0,2]}(t)dt \quad (A \in \mathcal{F}), \qquad X : \Omega \to \mathbf{R}, \omega \longmapsto \omega,$$

とすると,

$$f(x) \coloneqq \begin{cases} 1/2 \, (0 \le x \le 2) \\ 0 \quad (\text{otherwise}) \end{cases}, \qquad \forall x \in \mathbf{R}, F_X(x) = \int_{-\infty}^{x} f(t)\, dt$$

が成り立つので, $X \sim U(0,2)$.

Prop.7.5

X, Y は独立で $X, Y \sim U(a, b)$ であるとき, $Z = X + Y$ の確率密度関数は,

$$f_Z(z) = \begin{cases} \dfrac{z-2a}{(b-a)^2} & (2a < z < a+b) \\ \dfrac{2b-z}{(b-a)^2} & (a+b < z < 2b) \\ 0 & (\text{otherwise}) \end{cases}$$

となる. 特に, 一様分布は再生性を持たない. (再生性は§9 参照)

Rem. f_Z のような確率密度関数を持つ確率分布を**三角分布**という.

Def.7.6 （無記憶性）
X を確率変数とする．

$$\forall t, s > 0, P(X > t + s | X > t) = P(X > s)$$

を満たすとき，X が従う確率分布は**無記憶性**を持つという．

Def.7.7（指数分布）
$\lambda > 0$ とする．確率密度関数 f が,

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

で与えられる連続分布をパラメタ λ の**指数分布**といい，$Exp(\lambda)$ とかく．

Rem.

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{0}^{\infty} \lambda e^{-\lambda x}\,dx = 1$$

e.g.7.5 1 時間に平均 4 回電話がランダムにかかってくるコールセンターにおいて次に電話がかかってるくるまでの時間が 30 分以内である確率は,

$$P\left(X \leq \frac{30}{60}\right) = \int_{0}^{\frac{30}{60}} 4e^{-4t}\,dt = 1 - \frac{1}{e^2} \approx$$

Prop.7.6
指数分布は無記憶性を持つ．また，$\mathbf{R}_{>0}$ 上で定義された連続型確率分布で無記憶性を有するものは指数分布に限る．

Def.7.8（正規分布）

$\mu \in \mathbf{R}, \sigma > 0$ とする．確率密度関数 f が

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \ (x \in \mathbf{R})$$

で与えられる連続分布を平均 μ , 分散 σ^2 の**正規分布**といい, $N(\mu, \sigma^2)$ で表す. 特に $N(0,1)$ を**標準正規分布**という.

Rem.　Gauss 積分によって次がわかる.

$$\int_{-\infty}^{\infty} f(x)\,dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx = \frac{1}{\sqrt{2\pi}\sigma} \cdot \sqrt{2\pi}\sigma = 1$$

Rem. 標準正規分布 $N(0,1)$ の確率密度関数を ϕ ，累積分布関数を Φ と表す.

$$\phi(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \qquad \Phi(x) := \int_{-\infty}^{x} \phi(t)\,dt$$

このとき，$N(\mu, \sigma^2)$ の確率密度関数を f とすると，$f(x) = \dfrac{1}{\sigma} \phi\left(\dfrac{x-\mu}{\sigma}\right)$ となる.

Prop.7.7 $X \sim N(\mu, \sigma^2)$ とすると,

(1) $Y := a + bX \sim N(a + b\mu, b^2\sigma^2)$ $(a \in \mathbf{R}, b \neq 0)$

(2) $Z := \dfrac{X - \mu}{\sigma} \sim N(0,1)$

Rem. (2)における Z を X の**基準化**という.

§8 積率母関数

Def.8.1（積率）
$a \in \mathbf{R}$ に対し，$E[(X-a)^n]$ が存在すれば，これを X の a 周りの $\boldsymbol{n}$ **次積率**（**モーメント**）という．特に $E[X^n]$ が存在すれば，これを原点周りの n 次積率という．

Def.8.2（積率母関数）
X を確率変数とする．ある $\delta > 0$ が存在して，各 $\theta \in (-\delta, \delta)$ に対して $E[e^{\theta X}]$ が存在するとき，

$$M_X(\theta) := E[e^{\theta X}]$$

を X の**積率母関数**という．

Rem. X が離散型，連続型のとき，それぞれ

$$M_X(\theta) = \sum_i p_i e^{\theta a_i} \ \big(p_i := P(X = a_i)\big), \qquad M_X(\theta) = \int_{-\infty}^{\infty} f(x) e^{\theta x}\, dx$$

この定義は確率ベクトルにも一般化される．

Def.8.3（積率母関数）
$\boldsymbol{X}$ を確率ベクトルとする．ある $\delta > 0$ が存在して，各 $\boldsymbol{\theta} \in U_\delta(\mathbf{0})$ に対して $E\left[e^{\boldsymbol{\theta}\boldsymbol{X}^{\mathrm{T}}}\right]$ が存在するとき，

$$M_{\boldsymbol{X}}(\boldsymbol{\theta}) := E\left[e^{\boldsymbol{\theta}\boldsymbol{X}^{\mathrm{T}}}\right]$$

を $\boldsymbol{X}$ の積率母関数という．但し，$\boldsymbol{X}^{\mathrm{T}}$ は $\boldsymbol{X}$ の転置である．

Thm.8.1 X を確率変数とする．ある $\delta > 0$ が存在して，各 $\theta \in (-\delta, \delta)$ に対して積率母関数 $M_X(\theta)$ が定義できるならば，$M_X(\theta)$ は C^∞ 級であり，

$$\frac{d^n M_X}{d\theta^n}(0) = E[X^n] \ (n \in \mathbf{N} \cup \{0\}) \qquad \cdots (*)$$

特に，

$$E[X] = M_X'(0), \qquad V[X] = M_X''(0) - \big(M_X'(0)\big)^2$$

〈proof〉 X が連続型のとき，

$$M_X'(\theta) = \frac{d}{d\theta} \sum_i p_i e^{\theta a_i} = \sum_i a_i p_i e^{\theta a_i}\,, \qquad M_X'(0) = \sum_i a_i p_i = E[X]$$

$$M_X''(\theta) = \frac{d^2}{d\theta^2} \sum_i p_i e^{\theta a_i} = \sum_i a_i^2 p_i e^{\theta a_i}\,, \qquad M_X''(0) = \sum_i a_i^2 p_i = E[X^2]$$

これを繰り返して $(*)$ を得る．

X が連続型のとき，

$$M'_X(\theta)=\frac{d}{d\theta}\int_{-\infty}^{\infty}f(x)e^{\theta x}\,dx=\int_{-\infty}^{\infty}xf(x)e^{\theta x}\,dx,\qquad M'_X(0)=\int_{-\infty}^{\infty}xf(x)\,dx=E[X]$$

$$M''_X(\theta)=\frac{d^2}{d\theta^2}\int_{-\infty}^{\infty}f(x)e^{\theta x}\,dx=\int_{-\infty}^{\infty}x^2f(x)e^{\theta x}\,dx,\qquad M''_X(0)=\int_{-\infty}^{\infty}x^2f(x)\,dx=E[X^2]$$

これを繰り返して $(*)$ を得る．また，$V[X]=E[X^2]-(E[X])^2$ より次を得る．

$$V[X]=M''_X(0)-\left(M'_X(0)\right)^2\ \blacksquare$$

Prop.8.1 $\boldsymbol{X}=(X_1,\cdots,X_n)$ を確率ベクトルとする．ある $\delta>0$ が存在して，各 $\boldsymbol{\theta}=(\theta_1,\cdots,\theta_2)\in U_\delta(\mathbf{0})$ に対して積率母関数 $M_{\boldsymbol{X}}(\boldsymbol{\theta})$ が定義できるならば，$M_{\boldsymbol{X}}(\boldsymbol{\theta})$ は C^∞ 級であり，

$$\frac{\partial^{k_1+\cdots+k_n}M_{\boldsymbol{X}}}{\partial\theta_1^{k_1}\cdots\partial\theta_1^{k_n}}(\mathbf{0})=E\left[X_1^{k_1}\cdots X_n^{k_n}\right]\ (k_i\in\mathbf{N})$$

$$\mathrm{Cov}(X_i,Y_j)=\frac{\partial^2M_{\boldsymbol{X}}}{\partial\theta_i\partial\theta_j}(\mathbf{0})-\frac{\partial M_{\boldsymbol{X}}}{\partial\theta_i}(\mathbf{0})\frac{\partial M_{\boldsymbol{X}}}{\partial\theta_j}(\mathbf{0})$$

$$M_{X_i}(\theta_i)=M_{\boldsymbol{X}}(0,\cdots,0,\theta_i,0,\cdots,0)$$

Prop.8.2 確率変数 X,Y は独立とする．このとき，

$$M_{X+Y}(\theta)=M_X(\theta)M_Y(\theta)$$

$\langle$proof$\rangle$

$$M_{X+Y}(\theta)=E\left[e^{\theta(X+Y)}\right]=E[e^{\theta X}e^{\theta Y}]=E[e^{\theta X}]E[e^{\theta Y}]=M_X(\theta)M_Y(\theta)\ \blacksquare$$

Prop.8.3 X,Y を確率変数とする．

$$\exists\delta>0,\forall\theta\in(-\delta,\delta),M_X(\theta)=M_Y(\theta)$$

であるとき，X が従う確率分布と Y が従う確率分布は一致する．$(P_X=P_Y)$

確率変数の従う分布は積率母関数によって特徴づけられる．よって 2 つの確率変数の従う分布が等しいことを示す代わりにそれらの積率母関数が一致することを示せばよい．（積率母関数の一致の方が簡単に示せることがある．）

§9 各分布の積率母関数と平均, 分散

Prop.9.1 $X \sim B(n,p)$ のとき,

$$M_X(\theta) = \left(pe^\theta - (1-p)\right)^n, \qquad E[X] = np, \qquad V[X] = np(1-p)$$

⟨proof⟩

$$M_X(\theta) = \sum_{i=0}^{n} e^{\theta i} \binom{n}{i} p^i (1-p)^{n-i} = \sum_{i=0}^{n} \binom{n}{i} (pe^\theta)^i (1-p)^{n-i} = \left(pe^\theta - (1-p)\right)^n$$

$$M_X'(\theta) = n\left(pe^\theta - (1-p)\right)^{n-1} pe^\theta$$

$$M_X''(\theta) = n(n-1)\left(pe^\theta - (1-p)\right)^{n-2} p^2 e^{2\theta} + n\left(pe^\theta - (1-p)\right)^{n-1} pe^\theta$$

$$E[X] = M_X'(0) = np, \qquad V[X] = M_X''(0) - (np)^2 = np(1-p) \ \blacksquare$$

Prop.9.2 $X \sim Po(\lambda)$ のとき,

$$M_X(\theta) = e^{\lambda(e^\theta - 1)}, \qquad E[X] = \lambda, \qquad V[X] = \lambda$$

⟨proof⟩

$$M_X(\theta) = \sum_{i=0}^{n} e^{\theta i} e^{-\lambda} \frac{\lambda^i}{i!} = e^{-\lambda} \sum_{i=0}^{n} \frac{(\lambda e^\theta)^i}{i!} = e^{-\lambda} e^{\lambda e^\theta} = e^{\lambda(e^\theta - 1)}$$

$$M_X'(\theta) = \lambda e^\theta e^{\lambda(e^\theta - 1)}, \qquad M_X''(\theta) = \lambda e^\theta e^{\lambda(e^\theta - 1)} + \lambda e^\theta M_X'(\theta)$$

$$E[X] = M_X'(0) = \lambda, \qquad V[X] = M_X''(0) - \lambda^2 = (\lambda + \lambda^2) - \lambda^2 = \lambda \ \blacksquare$$

Prop.9.3 $X \sim Geo(p)$ のとき,

$$M_X(\theta) = p\sum_{i=0}^{\infty} \left((1-p)e^\theta\right)^i \ \left(\theta < \log\frac{1}{1-p}\right), \qquad E[X] = \frac{1-p}{p}, \qquad V[X] = \frac{1-p}{p^2}$$

⟨proof⟩

$$M_X(\theta) = \sum_{i=0}^{\infty} e^{\theta i}(1-p)^i p = p\sum_{i=0}^{\infty}\left((1-p)e^\theta\right)^i$$

$(1-p)e^\theta < 1$ つまり $\theta < \log\dfrac{1}{1-p}$ のとき, $M_X(\theta) = \dfrac{p}{1-(1-p)e^\theta}$,

$$M_X'(\theta) = \frac{p(1-p)e^\theta}{(1-(1-p)e^\theta)^2}, \qquad M_X''(\theta) = \frac{p(1-p)e^\theta(1+(1-p)e^\theta)}{(1-(1-p)e^\theta)^3}$$

$$E[X] = M_X'(0) = \frac{1-p}{p},$$

$$V[X] = M_X''(0) - \left(\frac{1-p}{p}\right)^2 = \frac{p(1-p)(2-p)}{p^3} - \left(\frac{1-p}{p}\right)^2 = \frac{1-p}{p^2} \ \blacksquare$$

Prop.9.4　$X \sim U(a,b)$ のとき，

$$M_X(\theta) = \begin{cases} \dfrac{1}{b-a} \cdot \dfrac{e^{b\theta} - e^{a\theta}}{\theta} & (\theta \neq 0) \\ 1 & (\theta = 0) \end{cases}, \qquad E[X] = \frac{b-a}{2}, \qquad V[X] = \frac{(b-a)^2}{12}$$

⟨proof⟩

$$M_X(\theta) = \int_a^b e^{\theta x} \frac{1}{b-a} dx = \frac{1}{b-a} \int_a^b e^{\theta x}\, dx = \begin{cases} \dfrac{1}{b-a} \cdot \dfrac{e^{b\theta} - e^{a\theta}}{\theta} & (\theta \neq 0) \\ 1 & (\theta = 0) \end{cases}$$

$$M'_X(\theta) = \frac{1}{b-a} \cdot \frac{d}{d\theta} \left(\frac{e^{b\theta} - e^{a\theta}}{\theta} \right) = \frac{(be^{b\theta} - ae^{a\theta})\theta - (e^{b\theta} - e^{a\theta})}{\theta^2}$$

$$M''_X(\theta) = \frac{1}{b-a} \cdot \frac{(b^2 e^{b\theta} - a^2 e^{a\theta})\theta^2 - 2(be^{b\theta} - ae^{a\theta})\theta + 2(e^{b\theta} - e^{a\theta})}{\theta^3}$$

$$E[X] = M'_X(0) = \lim_{\theta \to 0} M'_X(\theta) =_{(*)} \frac{1}{b-a} \lim_{\theta \to 0} \frac{(b^2 e^{b\theta} - a^2 e^{a\theta})\theta}{2\theta} = \frac{1}{b-a} \cdot \frac{(b-a)^2}{2} = \frac{a+b}{2}$$

$$M''_X(0) = \lim_{\theta \to 0} M''_X(\theta) =_{(*)} \frac{1}{b-a} \cdot \lim_{\theta \to 0} \frac{(b^3 e^{b\theta} - a^3 e^{a\theta})\theta^2}{3\theta^2} = \frac{1}{b-a} \cdot \frac{(b-a)^3}{3} = \frac{b^2 + ba + a^2}{3}$$

$$V[X] = M''_X(0) - \left(\frac{a+b}{2} \right)^2 = \frac{(b-a)^2}{12} \ \blacksquare$$

Rem. $(*)$ では l'Hospital の定理を用いた.

Prop.9.5　$X \sim Exp(\lambda)$ のとき，

$$M_X(\theta) = \frac{\lambda}{\lambda - \theta}\ (\theta < \lambda), \qquad E[X] = \frac{1}{\lambda}, \qquad V[X] = \frac{1}{\lambda^2}$$

⟨proof⟩

$$M_X(\theta) = \int_0^\infty e^{\theta x} \cdot \lambda e^{-\lambda x}\, dx = \lambda \int_0^\infty e^{(\theta - \lambda)x}\, dx = \frac{\lambda}{\lambda - \theta} \lim_{R \to \infty} (1 - e^{(\theta - \lambda)R})$$

$\theta < \lambda$ のとき， $M_X(\theta) = \dfrac{\lambda}{\lambda - \theta}$, $\qquad M'_X(\theta) = \dfrac{\lambda}{(\lambda - \theta)^2}$, $\qquad M''_X(\theta) = \dfrac{2\lambda}{(\lambda - \theta)^3}$

$$E[X] = M'_X(0) = \frac{1}{\lambda}, V[X] = M''_X(0) - \left(\frac{1}{\lambda} \right)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \ \blacksquare$$

Prop.9.6 $X \sim N(\mu, \sigma^2)$ のとき，

$$M_X(\theta) = e^{\mu\theta + \frac{\sigma^2\theta^2}{2}}, \qquad E[X] = \mu, \qquad V[X] = \sigma^2$$

〈proof〉

$$M_X(\theta) = \int_{-\infty}^{\infty} e^{\theta x} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2 - 2\theta x \quad {}^2}{2\sigma^2}} dx$$

ここで指数の分子を平方完成すると，

$$(x-\mu)^2 - 2\theta x\sigma^2 = x^2 - 2(\mu + \theta\sigma^2)x + \mu^2 = \{x - (\mu + \theta\sigma^2)\}^2 - 2\mu\theta\sigma^2 - \theta^2\sigma^4$$

$$\therefore M_X(\theta) = e^{\mu\theta + \frac{\sigma^2\theta^2}{2}} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{\{x-(\mu+\theta\sigma^2)\}^2}{2\sigma^2}} dx = e^{\mu\theta + \frac{\sigma^2\theta^2}{2}}$$

$M_X'(\theta) = (\mu + \sigma^2\theta)e^{\mu\theta + \frac{\sigma^2\theta^2}{2}} = (\mu + \sigma^2\theta)M_X(\theta)$

$M_X''(\theta) = \sigma^2 M_X(\theta) + (\mu + \sigma^2\theta)M_X'(\theta) = \sigma^2 M_X(\theta) + (\mu + \sigma^2\theta)^2 M_X(\theta)$

$E[X] = M_X'(0) = \mu, V[X] = M_X''(0) - \mu^2 = (\sigma^2 + \mu^2) - \mu^2 = \sigma^2$ ■

パラメタを持つ確率分布には再生性という性質を持つものがある．

Def.9.1（再生性）

Λ を集合とし，$\lambda \in \Lambda$ に対して $f(x, \lambda)$ をある確率変数の確率密度関数とする．

(1) $\mathfrak{F} = \{f(x, \lambda) : \lambda \in \Lambda\}$ を**確率分布族**という．

(2) X, Y は独立，$X \sim f(x, \lambda_1), Y \sim f(x, \lambda_2)$

$$\Rightarrow \exists \lambda_3 \in \Lambda, X + Y \sim f(x, \lambda_3)$$

が成り立つとき，確率分布族 $\mathfrak{F}$ は λ に関して**再生性**を持つという．

Rem. $X \sim f(x)$ とは，X が確率密度関数 f を持つ確率分布に従うこと．

Rem. 離散分布でも同様に定義される．

Prop.9.7（二項分布の再生性）

確率変数 X, Y は独立，$X \sim B(n_1, p), Y \sim B(n_2, p)$ のとき，

$$X + Y \sim B(n_1 + n_2, p)$$

つまり二項分布は n に関する再生性を持つ．

〈proof〉 $X \sim B(n_1, p), Y \sim B(n_2, p)$ とすると，

$$M_{X+Y}(\theta) = M_X(\theta)M_Y(\theta) = \left(pe^\theta - (1-p)\right)^{n_1} \left(pe^\theta - (1-p)\right)^{n_2} = \left(pe^\theta - (1-p)\right)^{n_1+n_2}$$

$\therefore$ Prop.8.3 により，$X + Y \sim B(n_1 + n_2, p)$ ■

Prop.9.8（Poisson 分布の再生性）

確率変数 X, Y は独立，$X \sim Po(\lambda_1), Y \sim Po(\lambda_2)$ のとき，

$$X + Y \sim Po(\lambda_1 + \lambda_2)$$

つまり Poisson 分布は λ に関する再生性を持つ.

〈proof〉$X \sim Po(\lambda_1), Y \sim Po(\lambda_2)$ とすると，

$$M_{X+Y}(\theta) = M_X(\theta)M_Y(\theta) = e^{\lambda_1(e^\theta - 1)}e^{\lambda_2(e^\theta - 1)} = e^{(\lambda_1+\lambda_2)(e^\theta - 1)}$$

∴ Prop.8.3 により，$X + Y \sim Po(\lambda_1 + \lambda_2)$ ■

Prop.9.9 （正規分布の再生性）

確率変数 X, Y は独立，$X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$ のとき，

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

つまり正規分布は (μ, σ^2) に関して再生性を持つ.

〈proof〉$X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$ とすると，

$$M_{X+Y}(\theta) = M_X(\theta)M_Y(\theta) = e^{\mu_1\theta + \frac{\sigma_1^2\theta^2}{2}}e^{\mu_2\theta + \frac{\sigma_2^2\theta^2}{2}} = e^{(\mu_1+\mu_2)\theta + \frac{(\sigma_1^2+\sigma_2^2)\theta^2}{2}}$$

∴ Prop.8.3 により，$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ ■

§10 多元分布

Prop.10.1 $(\Omega, \mathcal{F}, P)$ を確率空間とし，$\boldsymbol{X}: \Omega \to \mathbf{R}^n$ を n 次元確率ベクトルとする．

$$P_{\boldsymbol{X}}: \mathcal{B}_n \to [0,1], \qquad P_{\boldsymbol{X}}(B) := P\Big(\boldsymbol{X}^{-1}(B)\Big)\ (B \in \mathcal{B}_n)$$

と定めると，$P_{\boldsymbol{X}}$ は可測空間 $(\mathbf{R}^n, \mathcal{B}_n)$ 上の確率測度となる．

Def.10.1（確率分布）
Prop.10.1 で定義した $P_{\boldsymbol{X}}$ を $\boldsymbol{X}$ の**同時確率分布**といい, $(\mathbf{R}^n, \mathcal{B}_n, P_{\boldsymbol{X}})$ を $\boldsymbol{X}$ によって**誘導された確率空間**という．このとき $\boldsymbol{X}$ は確率分布 $P_{\boldsymbol{X}}$ に従うといい，$\boldsymbol{X} \sim P_{\boldsymbol{X}}$ で表す.

Rem. 離散型, 連続型確率ベクトルに対応する同時確率分布はそれぞれ**離散型同時確率分布**, **連続型同時確率分布**と呼ばれる．

Rem. 1次元の離散分布と同様に離散型同時確率分布は,
$(a_1(j_1), \cdots, a_n(j_n)) \in X(\Omega)$ と $p_{j_1 \cdots j_n} := P\Big(\boldsymbol{X} = \big(a_1(j_1), \cdots, a_n(j_n)\big)\Big)$ の対応と考えられる．

Def.10.2 離散型確率ベクトル $\boldsymbol{X}$ に対して，上の Rem.と同じ表記を用いて,

$$p_{j_i} := \sum_{j_n \in J_n} \cdots \sum_{j_{i+1} \in J_{i+1}} \sum_{j_{i-1} \in J_{i-1}} \cdots \sum_{j_1 \in J_1} p_{j_1 \cdots j_n}$$

と表す．$a_i(j)$ と p_{j_i} の対応で与えられる離散分布を $\boldsymbol{X}$ の**周辺分布**という．

具体的な多元分布として正規分布の多元 ver.を取り上げる．

Def.10.3（2 変量正規分布）
$\mu_i \in \mathbf{R}, \sigma_j > 0, \rho \in (-1,1)$ とする．同時確率密度関数が

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right\}} \Big((x,y) \in \mathbf{R}^2\Big)$$

で与えられる確率分布を**2変量正規分布**といい，$N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ と表す.

Prop.10.2
$\boldsymbol{X} = (X_1, X_2) \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ のとき，X_1, X_2 の周辺確率密度関数はそれぞれ,

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}, \qquad f_2(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}$$

で与えられる．

さらに一般化して $\mathbf{R}^p$ の正規分布が定義される．

Def.10.4（ p 変量正規分布）

$p \in \mathbf{N}, \boldsymbol{\mu} \in \mathbf{R}^p, \Sigma$ を p 次正定値対称行列とする．同時確率密度関数が

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^p \det \Sigma}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}}$$

で与えられる確率分布を平均ベクトル $\boldsymbol{\mu}$ ，分散共分散行列 Σ の $\boldsymbol{p}$ **変量正規分布**といい，$N_p(\boldsymbol{\mu}, \Sigma)$ と表す．

Rem. $p = 2, \boldsymbol{\mu} = (\mu_1, \mu_2), \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \left(\mu_i \in \mathbf{R}, \sigma_j > 0, \rho \in (-1,1)\right)$

とすると $N_p(\boldsymbol{\mu}, \Sigma)$ は二変量正規分布 $N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ となる，つまり二変量正規分布の一般化になっていることを確認する．

$\boldsymbol{x} = (x, y)$ とする．まず，$\det \Sigma = \sigma_1^2\sigma_2^2(1-\rho^2)$ ．

$$(\boldsymbol{x}-\boldsymbol{\mu})\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}} = \begin{pmatrix} x-\mu_1 \\ y-\mu_2 \end{pmatrix} \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} \begin{pmatrix} x-\mu_1 & y-\mu_2 \end{pmatrix}$$

$$= \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \{\sigma_2^2(x-\mu_1)^2 - 2\rho\sigma_1\sigma_2(x-\mu_1)(y-\mu_2) + \sigma_1^2(y-\mu_2)^2\}$$

$$= \frac{1}{1-\rho^2} \left\{ \frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right\}$$

$$\therefore f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right\}}$$